Statistical Learning Approaches to Sentiment Analysis in the Nigerian Banking Context

A thesis submitted in partial fulfilment of the requirements of Sheffield Hallam University for the degree of Doctor of Philosophy

By

Bayode Oluwatoba Ogunleye

October 2021

I hereby declare that:

1. I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.
2. None of the material contained in the thesis has been used in any other submission for an academic award.
3. I am aware of and understand the University's policy on plagiarism and certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged.
4. The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy.
5. The word count of the thesis is 44,849.

Name	Bayode Oluwatoba Ogunleye
Date	October 2021
Award	PhD
College	Business, Technology and Engineering
Director(s) of Studies	Dr Jotham Gaudoin

Abstract

Banking is an essential component of our day-to-day activity, and the sector contributes to the development of every Nation's economy. To ensure bank stability, the banks have gone through several deregulation and reformation processes. Unfortunately, these processes and technological advancement have led to increased competitiveness, saturated market, and low profitability. Thus, the need for banks to build customer centric service to gain profitability and stability is vital. This opens the need to investigate customers attitude towards banking.

With the growing usage of social network sites (SNS), the user generated content (UGC) has opened opportunities for banks to mine customers opinion. This can help the banks to generate insight into their product and services, create marketing strategies and manage their reputation. To the customers, this serves as source to information that can help in supporting decision making on product or service purchase. In this study, sentiment analysis (SA) techniques were employed to investigate customers attitude towards banking using Twitter data. However, the unstructured nature of the data and word ambiguity made sentiment analysis complicated. In the context of this study, SA is more difficult because it involves the natural language processing of Pidgin English and English words in the bank domain. Unfortunately, there are limited or no resources for this purpose.

This study is of two main tasks namely, sentiment classification and aspect extraction for sentiment analysis. For the sentiment classification task, this study utilised both lexical based approach and the machine learning approach. The lexical based approach relies on opinion words, it is easy to understand the sentiment classification result and ways to improve. Machine learning algorithms are black box models and often are not interpretable by a human. However, they produce models with good performance. The study proposes SentiLeye, a novel lexicon algorithm and compared with existing lexicons. Results showed SentiLeye outperformed others due to domain terms, opinionated-objective words, negation, and language which was put into consideration during the development. Alternatively, the machine learning models were compared. The performance result of the classification models validated Support Vector Machine (SVM) with accuracy of 82% as the most performed classification model in the banking context. The second task involved the use of statistical topic modelling techniques for aspect extraction. The topic models were compared, and Latent Dirichlet Allocation (LDA) showed the best performance in terms of topic coherence and interpretable terms. Thus, this study proposes a Topic-Sentiment Banking System (TSBS) framework which was used to demonstrate the aspect of banking which customers

were happy or unhappy with. Our result showed the significance of customer service experience, transaction problem, and bank charges. Thus, recommends the banks to pay attention to these topics as our findings showed significant proportion of customers are unhappy with these aspects of banking.

Acknowledgement

I would like to thank my supervisors Teresa Brunsdon, Tonderai Maswera, Jotham Gaudoin and Laurence Hirsch. I am deeply grateful for your guidance, support, and kindness.

My special thanks go to my family and friends for their continuous support and motivation.

Lastly, indispensable to this project have been the contributions of the annotators, without whose diligent work, completion of this thesis would have been impossible.

List of Figures

List of Tables

List of Abbreviations

ABSA - Aspect Based Sentiment Analysis
API: Application Programming Interface
ATM - Automated Teller Machine
BERT - Bi-Directional Encoder Representations from Transformers
BI-LSTM - Bi-Directional Long Short-Term Memory
CBN - Central Bank of Nigeria
CNB - Complement Naive Bayes
CNN - Convolutional Neural Network
GB - Gradient Boosting
GNB - Gaussian Naive Bayes
HDP - Hierarchical Dirichlet Process
HMM - Hidden Markov model
JJ - Adjective
KNN - K Nearest Neighbour
LDA - Latent Dirichlet Allocation
LIWC - Linguistic Inquiry and Word Count
LM - Loughran & McDonald
LR - Logistic Regression
LSI - Latent Semantic Indexing
LSTM - Long Short-Term Memory
ME - Maximum Entropy
ML - Machine Learning
MNB - Multinomial Naive Bayes
NB - Naive Bayes
NLP - Natural Language Processing
NLTK - Natural Language ToolKit
NN - Singular Noun
NNP - Proper Noun Singular
NNPS - Proper Noun Plural
NNS - Noun Plural
NRC – National Research Council Canada
OVR - One versus Rest
POL - Polynomial Kernel
POS - Part of Speech
POS - Point of Sales
RB - Adverb
RBF - Radial Basic Kernel function
RE: Regular Expression
RF - Random Forest
RNN - Recurrent Neural Network
SA - Sentiment Analysis
SMA - Social Media Analytics
SNS - Social Network Sites
SVM - Support Vector Machine
TF-IDF - Term Frequency Inverse Document Frequency
USSD - Unstructured Supplementary Service Data
WKWSCI - Wee Kim Wee School of Communication & Information

TABLE OF CONTENTS

CHAPTER 1

1.0 Introduction

Banking is considered a daily activity in our society and importantly contributes to the economic stability and growth of any country (Alsmadi et al. 2019; Durusu-Ciftci et al. 2017). To the governments, the banks are of great importance in the conduct of their economic policies. However, digitalisation and globalisation have changed banking. The deregulation and reformation of banks are aimed to provide stability in the banking system (Altintas, 2020; CBN, 2012) but these have significantly increased competitiveness (Olokoyo, 2018) and reduced banks' profitability (Adaramola & Kolapo, 2019). In the last decade, traditional banks have been challenged with the emergence of several Fintech (financial technology) companies such as Starling, Stripe and Paypal (Thakor, 2020; Romānova & Kudinska, 2016). These companies offer financial services with endless innovative technology. Other notable ones are the cryptocurrency companies like bitcoin that also renders financial service in form of digital tokens, coins, or credit card. These companies have been able to reduce operation cost by innovative technological service delivery below those incurred by traditional banks. Thus, enhances profitability and gain competitive advantage by using some of the saved operation cost as interest to customers. These banking options provide competition to the traditional banks but advantageous to customers in terms of increased service provision and reduced service charge (Cecchetti, 2011; Ajisafe & Ajide, 2014).

Technological advancement has changed a lot in banking. For example, the traditional use of the bank hall capacity was to allocate a greater part for administrative work. However, in recent time, a large capacity has been allocated to customers for sales marketing and enquiries. More recently, there are various online (app-based) banks such as Monzo that renders general banking services online without any physical branch presence. There are also various e-commerce companies such as Tencent and Alibaba that also renders bank services without any physical branch presence. Their lower operation cost helped them to rise rapidly in their banking business (Baesens et al. 2016). The banking technology has also been applied to social networks for delivering bank services. These banking options have led to high customer churn rate (CIUCI, 2016). The bank customers are now in control of their banking need which has resulted to high expectation due to the strained competitiveness. The traditional banks are therefore under intense pressure and scrutiny to find ways to understand and maintain strong connection with their customers since they rely on their customers for profitability and stability. Thus, the banks are at the end to design their service model from customers' perspective for survival. Kaufman (2014) stressed the need for organisations to build a good customer service culture to enjoy a sustainable competitive advantage. Vyas & Raitani (2014) added since banks offer similar services, improved service experience is needed for competitive advantage.

1.1 Background of Nigeria Banking Industry

The reformation in Nigeria banking sector has been a regular feature over the years. This reform is done by Central Bank of Nigeria (CBN) to improve banks' performance and stability (CBN, 2009). The CBN with the goal of stabilizing and creating more robust banking system had also created an intense and competitive banking market where banks need to fight for survival.

The earliest bank reform was in 1952 which was done to provide framework for regulation and supervision of banks. In the mid 1980’s deregulation policies were implemented and universal banking guidelines in 2001. In 2005, commercial banks were reduced and merged from 89 to 24 banks through the requirement of minimum capitalization raised from 2 billion to 25 billion naira. In 2009, five commercial banks were investigated by CBN to have capital inadequacy, liquidity, and solvency problem. Thus, were merged with other existing ones considered strong enough to acquire them (CBN annual report, 2009). The monetary policy made by CBN is considered harsh and thus, led to low loan income rate (Nwanne, 2018). This implies banks that rely on loan income experience low profitability. CBN (2017) showed the performance outlook of banks (shown in table 1.0 below) from 2005 to 2017 using return on asset (ROA). The report shows a high instability in the banking system over the period. In conclusion, the reformations have contributed to high instability, increased competitiveness, and low profitability (Adaramola & Kolapo, 2019).

Period	ROA
2005	2.096
2006	1.754
2007	2.669
2008	2.105
2009	2.748
2010	3.8
2011	0.419
2012	3.269
2013	2.051
2014	1.753
2015	1.639
2016	2.838
2017	3.123

Table 1.0: CBN statistical report 2017

Nigeria still operates a cash-based economy especially the banking sector (Ayo et al. 2016). In the society, most buying, and selling are done with physical cash. The banks still have large number of customers that operate physical branch banking where the customers wait in a queue, to deposit cash, transfer funds & pay bills despite the availability of electronic banking services (Ayo et al. 2016). This results in a large amount of customer traffic in the banking hall. The banks are trying to meet up with their customers' expectation in terms of service delivery using banking technologies. Unfortunately, adoption of these technologies in the bank customer society is still limited (Oyelami et al. 2020; Tarhini et al. 2015). Automated teller machine (ATM) is considered the most adopted means of electronic banking in Nigeria (Oyelami et al. 2020; Jegede, 2014). The machine was praised for improvement in operations in terms of reduction in customer service delivery time, improvement of service quality, access to bank services at any time, and reduction in operation cost (Abdullai & Nyaoga, 2017). However, the complaint on the ATM now seems to have overshadowed the benefits. More recently, the Unstructured Supplementary Service Data (USSD) and mobile banking technologies have witnessed significant growth (KPMG, 2018). As this is just developing, the impact is yet unknown.

1.2 Research Motivation

The proactive approach for banks to gain profitability and stability is to improve their customer experience and manage their expectation. Thus, the need for banks to build a better service culture which is customer focused is imperative. This can be achieved through analytics of customers' attitude towards bank services. However, to understand customers' banking attitude is challenging due to customers' constantly changing behaviour. This opens the need for an extensive investigation on customers' attitude. Customers' attitude is simply defined as the perception or learned experience, feelings (positive or negative) and intended behaviour towards a product or service (Szmigin & Piacentini, 2018). A common way to understand customers' attitude or perception is by retrieving information from them which can be done through survey. Baumann et al. (2012) stated customer survey is an effective and economical way of retrieving and measuring customers' perception. The survey method is one of the best ways to understand customer' perception thereby finding out how to meet their needs and expectation. There are many studies (discussed further in chapter 2) found on bank customer experience, service quality and satisfaction performance index. However, most of these studies conducted their study using dataset retrieved from customers by means of interview, focus group and/or questionnaire (survey data). This means the studies are limited to pre-defined variables and more biased. This is because surveys are designed based on a particular investigation and thus, aligned to the author's purpose. More importantly, survey data collection is labour intensive and time consuming.

Lately, the use of internet and social network sites is increasing. Social network sites (SNS) are networked communication platforms where users connect and interact with other users or express their view (Clark et al. 2018). This study recognises there is a slight difference in the term "social network" and "social media". However, these terms will be used interchangeably because they mean same thing in this context. Mogaji et al. (2016) defined social media as the web platform that allows users to express their opinion or feelings freely towards a topic or subject. The communication media is simply the web forum where people comment, review, and give feedback on various issues. In the context of customer-to-organisation relationship, this is on products purchased or service experienced. Social media opens channel of information for analytics through customer interaction on service experience, feedback and complaints thereby made textual data available for organisations to earn knowledge. In this digital world, people spend time on social media and thus, makes the interaction of people on a target spread quickly across the globe. People spend more time on social media to communicate their experience, issues, and complaints. For example, potential customers tend to read product reviews before their purchase. In the banking context, customers tweet to confirm the ATM service performance in a particular area or bank. This impacts the reputation of the target, attributes of the target and the company (Szmigin & Piacentini, 2018). More recently, the need to deploy banking technologies to render service on social media became more popular due to Covid-19 pandemic where most countries were on lockdown. People could not go to bank branches physically and thus banks were limited to deliver services online. The popular ways that customers communicated with their banks were through telephone, emails, web chats, and social media. Therefore, the online banking is rapidly becoming an efficient tool to satisfy customers. Lin et al. (2018) stated social media has overpowered traditional customer care services owing to convenience, swiftness, and quick responsiveness. Agnihotri et al. (2021) described social media as a key communication platform for banks and their customers during Covid-19 pandemic. Their study highlighted the significance of social media during the Covid 19 pandemic and how the banks have utilised the platform to address customers' issues and complaints. They stated during the pandemic, service industries such as banks have utilised social media as a key platform to interact with their customers and deliver services. On social media customers interact, share their experiences, issues, and grievances. The experiences, issues, and grievances shared thus have direct influence on the reputation of the subject.

The banks have deployed technologies to bring banking services to customers at hand by launching banking technologies on social media. For example, In March 2015, Barclays bank (UK) implemented payment on Pingit app where customers can transact using their Twitter handle (Barclays, 2015). Royal bank of Canada implemented P2P payment between Facebook friends (RBC, 2013). ICICI bank (India) also launched their Pocket (Facebook) app which allows social customers to transfer payment (ICICI, 2013). In Nigeria, few banks have launched their banking services on social media platforms like "Whatsapp" to render banking services (Udenze et al. 2020). For example, United Bank for Africa (UBA), First Bank of Nigeria (FBN), Guaranty Trust Bank (GTBank) and Access Bank. GTbank was the first to launch social banking on Facebook where you can perform the basic bank transactions like checking of account balance, opening account, payment of bills and transferring money locally. Other banks are now trying to strategically deploy such technology to engage their customers as the social trend grows. In summary, the banks use social media platforms to deliver service faster and more conveniently to customers while customers also use social media to express their opinion freely and the content expressed can affect the reputation of the banks and their services.

Social network coverage is increasing every year. The bank processes have moved from paper to electronic so are customers. Rather than the traditional way of retrieving customer data, which is survey, social media data is more appropriate based on increased usage & popularity. Social media data is beneficial because it is unlimited to attribute or aspect, provides real time information and avoids time and cost of setting up and analysing a questionnaire or focus group. In addition, it saves time, provides updated customer information, big volume, and velocity of data which improves generalisation of research outcome. The adoption and usage of social media is increasing yearly. Statista (2020) conducted a study on social network usage. They reported a continuous yearly increment in usage worldwide. From 0.97 billion users in 2010, to 3.6 billion users in 2020 and forecast suggest there will be 4.41 billion users in 2025. This indicates social network will continue to grow and become more popular in the society.

1.3 Social Network Usage in Nigeria

Nigeria has her official language as English. However due to over 520 primary languages (Orife, 2020; Blench, 2014) spoken in the country, Pidgin English was developed to bridge the communication gap among citizens. The most spoken and used language in Nigeria is English and Pidgin English which also applies to social media (Chiluwa, 2016; Udofot & Mbarachi, 2016). Thus, bank customers interact with their respective bank in English and/or Pidgin English.

Social network sites (SNS) in Nigeria have witnessed significant growth in usage over the years. Poushter et al. (2018) reported 28% of Nigerian adults use SNS in 2013, 33% in 2015 and increased to 35% in 2017. Statista (2021) conducted a study on SNS usage (as shown in figure 1.1 below) in Nigeria and thus forecast usage till 2026. Their study forecasted an increment from 43 million users in 2021 to 103 million in 2026.

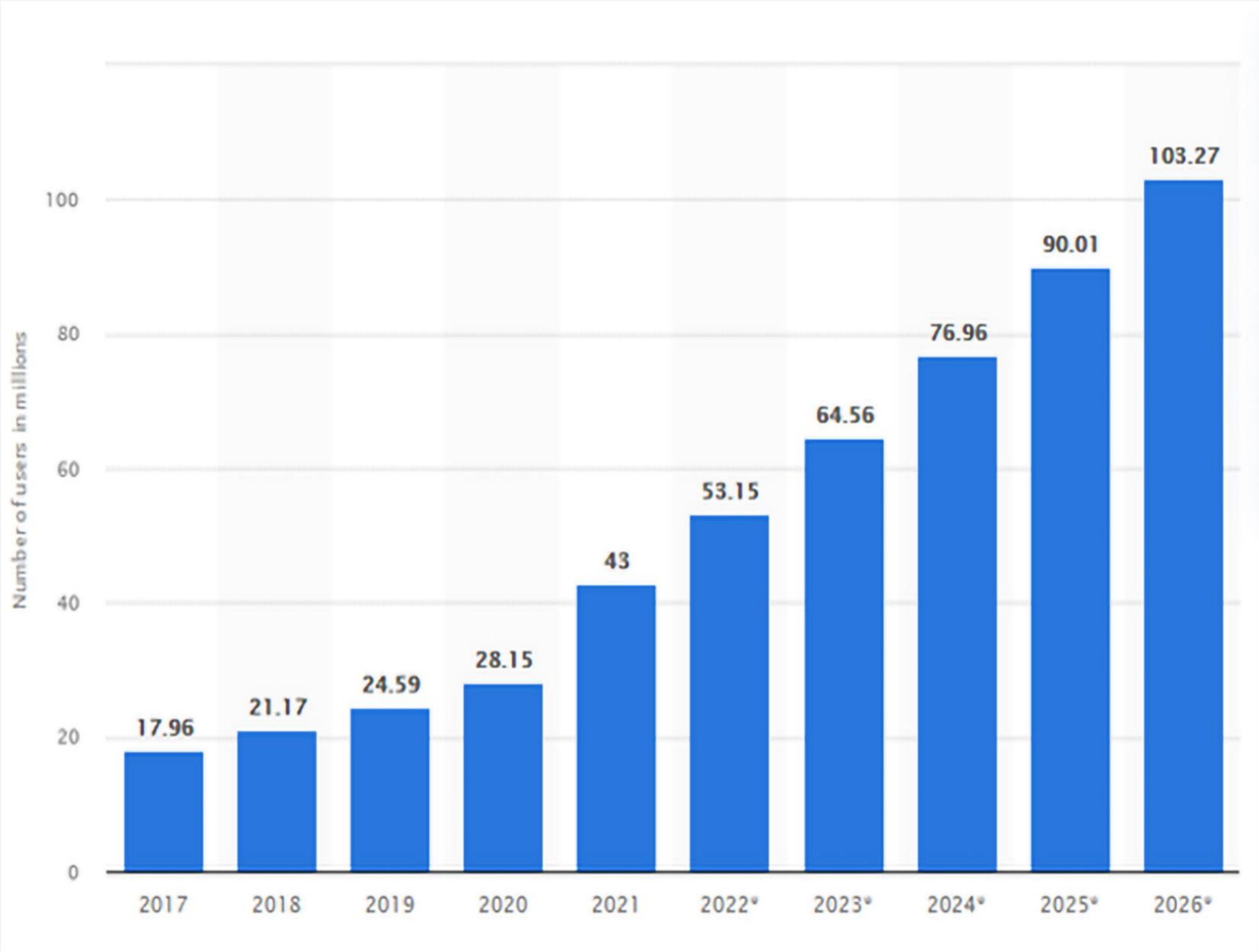


Figure 1.1: Statistics of social network users in Nigeria

The growing usage of social media in our daily life indicates how rich and important the media is in the society. It is also worth acknowledging that the banks have established their social media presence through their bank handles and thus engage with customers for different reasons. This study uses social network site in Nigeria banking context as source to customers' social data. Charoensukmongkol & Sasatanun (2017) found out entrepreneurs who use social media intensively for customer relationship management are more satisfied with their business performance. Organisations that analyse and manage customer feedback/complaints are 5% more productive and 6% more profitable on average than their competitors (McAfee & Brynjolfsson, 2012). A unified complete view of customers helps the banks to build a customer focused service model (SAS, 2018). To achieve this, the social voice of customers cannot be ignored.

1.4 Social Media Analysis

Due to the huge volume of information deposited on social media daily, the web platform provides opportunities for organisation to access textual data for in-depth analytics (Stieglitz et al. 2018). An important type of social media analytics that analyses peoples' opinion, attitudes, reviews, comments, and emotions toward a topic is sentiment analysis (SA). Sentiment analysis can be used to support business decision-making processes, monitor customers' attitude, understand what customers are saying and how they feel about a product or service. Liu (2012) stated SA techniques can be used to understand the opinion or intention of customers toward subjects such as product or service. In this study, SA is considered appropriate to investigate what bank customers are saying and their attitude towards bank's product or service. For example, tweets like *"ATMs in Bank X are working fine"* can be classed as positive sentiment towards ATM of Bank X. Thus, the aspect-target sentiment analysis is considered suitable. However, the unstructured nature of SNS data and word ambiguity made sentiment analysis complicated. In the context of this study, SA is more difficult because it involves the natural language processing of Pidgin English, and English terms. For example, tweets like *"na sure bank even for weekend no wahala customer care dey there"* is a positive tweet due to the pidgin term included. In the example *"sure"* is an English term used as a pidgin word to express positive experience. Unfortunately, there are limited resources for this purpose. Specifically, there is no available lexical resource for sentiment analysis in this context. In addition, there is no available annotated set for supervised machine learning systems in the banking context.

1.5 Aim & Objectives

To address these problems, this study aims to investigate customers' attitude towards banking using customers' social voice and be guided by answering the following research questions (RQ):

- RQ1. To what extent can automated sentiment analysis (SA) systems help generate actionable insights on customers' attitude towards bank product and service?
- RQ2. What are the effects (in terms of performance) of local & context-based terms in lexicon-based sentiment analysis?
- RQ3. Can machine learning models perform as an off-the-shelf sentiment classification method in the banking domain?
- RQ4. What are the major causes of misclassification?

To achieve the aim of this study and answer the research questions above, objectives formulated are:

- To review literature on bank customers' attitude, service experience, sentiment classification and aspect extraction for sentiment analysis. This will provide background knowledge to this study especially on the methodologies applicable in this context.
- To propose Topic-Sentiment Banking System (TSBS), a context-informed framework to understand customers' attitude towards up-to-date undefined aspects of banking. This will help detail approaches towards answering the research questions.
- To develop a novel lexicon algorithm that captures the tone and semantic orientation of banking and Pidgin English terms. This will help establish one of the components needed for the TSBS and thus answer RQ2.
- To conduct a benchmark comparison of sentiment classification and topic modelling techniques at tweet-level. This will help establish other components needed for the TSBS and thus answer RQ1, RQ3 and RQ4.

1.6 Contributions of the Thesis

This study contributes to existing knowledge in many ways. These can be summarised as follows:

I. Most of the SA studies found focused on dataset in official languages such as English, Spanish, and Chinese (Pontiki et al. 2016). This study contributes to the development of sentiment lexical resources for Pidgin English.

II. Loughran & McDonald (2011) warned future researchers in the finance sector not to use any sentiment lexicon created and evaluated outside the domain for analysis in the financial context. This is because financial terms differ with their semantic orientation in a different domain. Thus, makes lexicon evaluated in a different domain prone to misclassification when applied to financial domain. Krishnamoorthy (2018) added sentiment analysis tools accurately classify financial text better when performance measures of financial terms are considered. Yet, there is no available sentiment lexicon validated in the banking domain. Based on these assertions, this study contributes to fill this gap with the development of "SentiLeye" a novel domain-specific lexicon.

III. There is no study found to have investigated customers' attitude using their social voice. In this modern age, social media usage is increasing exponentially. Therefore, this study contributes by providing an up-to-date insight into customers' opinion towards banking using social media data.

IV. Construction and evaluation of standard annotated dataset (in Pidgin and/or English) in the banking context. The dataset can be accessed at https://www.kaggle.com/batoog/datasets

V. Model performance in financial domain is less accurate compared to other domains such as IMDB review (Xing et al. 2020). This is because the review domain has several studies that have compared models and proposed different techniques for improvement. There is no study found on banking in this field. This study thus contributes by conducting a benchmark comparison of the sentiment classification models in the banking context. Thus, propose a classification model with robust performance.

VI. Comparison of the unsupervised statistical topic models in the banking context.

VII. The development of TSBS framework to automatically uncover hidden topics and their sentiment.

1.7 Thesis Overview

This thesis is divided into nine chapters as follows:

Chapter 1: the research motivation was introduced. A brief background to the study was discussed. The research problem, aim and objectives were stated and then the research questions to be answered were outlined. This chapter provides details about subsequent chapters.

Chapter 2: presents literature review on the concept of customer attitude. The chapter provides background knowledge on customer experience, behaviour and attitude in the Nigeria banking context. Thus, identified the gap that previous studies relied on survey data for their investigation.

Chapter 3: presents comprehensive review of literature. The chapter provides background knowledge on research methodologies of sentiment classification techniques and aspect extraction for sentiment analysis. Thus, outlines the potential challenges and gap in knowledge.

Chapter 4: discusses the research methodologies of this study. This chapters presents the proposed TSBS framework which is helpful in classifying tweets into sentiment polarities and topics. In addition, this chapter provides extensive explanation into the algorithmic problems and solutions encountered in the field of sentiment analysis such as class imbalance, data sparsity and numeric vector representation.

Chapter 5: introduce the approaches to sentiment lexicon generation. This chapter presents SentiLeye, a novel lexicon algorithm developed to capture Nigeria Pidgin English and English words in Nigeria bank domain. Thus, the result presents the answer to RQ2.

Chapter 6: presents the comparative result of machine learning models for tweet-level sentiment analysis. This chapter also shows the evaluation of the classification models using appropriate evaluation metrics such as precision, recall, F1-score, and accuracy. Thus, the result presents the answer to RQ3 & RQ4 and discusses the findings against literature.

Chapter 7: presents the comparative result of statistical topic models for aspect extraction. Thus, presents the result of the proposed TSBS framework to answer RQ1 and discusses the findings against literature.

Chapter 8: evidences the evaluation process of the TSBS framework. The chapter presents the potential benefit, relevance, reliability, and generalizability of the TSBS research outcome. The discussion presented will help understand if the research findings and recommendations are reliable and relevant.

Chapter 9: concludes on the research work, discusses the challenges and limitation of the study. Thus, provides recommendation for future research work.

CHAPTER 2: Research Context

2.0 Introduction

As discussed in the previous chapter, there is need to understand customers' attitude towards banking. This can help develop proactive approach towards a better customer service culture, and thus enhance profitability and stability of the banks. To achieve this, this study will review literature on customer experience, behaviour, and attitude towards banking. Therefore, this chapter presents the research context of customer attitude in subsequent section.

2.1 Customer Experience

The instability and competitiveness in Nigeria banking sector have forced commercial banks to fully rely on customers to drive their business for survival. The banks have faced different types of reform by the Central Bank of Nigeria (CBN) to have a more stabilized banking system. However, still far from perfect (Olokoyo, 2018; Olokoyo, 2013). Therefore, the banks are faced with the challenge to improve customer experience for competitive edge. Customer experience has been defined differently by different studies. For example, Klaus & Maklan, (2013) defined customer experience as the customers' cognitive and affective assessment of all service encountered either directly or indirectly. Meyer & Schwanger, (2007) described customer experience as the internal and subjective response from service encountered either direct or indirect.

Customer experience is an understudied area (Pudaruth, 2021; Johnston & Kong, 2011). However, in this modern era of banking where customers are needed for competitive edge, there is need to understand and enhance customer experience. Johnston & Kong (2011) stated the buying of products and service comes with experience. For example, in Nigeria banking context, customers buy products like fixed deposit account, cards (credit, prepaid or debit), savings account (individual, children, student or domiciliary), current account (individual, Business or cooperate), ATM (automated teller machine) and loans (personal, automobile, salary, mortgage). In addition, customers encounter experience while trying to get bank services like international transfers (Form A or M, Western union, MoneyGram, Sendwave and Transfast), PTA (personal travel allowance), BTA (business travel allowance), USSD (Unstructured Supplementary Service Data), and online banking. This experience encountered provides emotional and sentiment bound decisions which influence repurchase or re-use. Thus, enhancing such experience is a crucial differentiator in today's hyper competitive and global market. The banks can benefit from this, by providing a re-use or repurchase emotional sentiment bound experience. Customer experience has been studied to impact satisfaction (Kumar et al. 2021; Garzaro et al. 2020; Loureiro & Ferreira, 2017; Chahal & Dutta, 2015; Liljander & Strandvik, 1997), and loyalty (Mokha & Kumar, 2022; Mascarenhas et al. 2006). Andaleeb et al. (2016) ascertained banks can benefit profitably when they enhance customer experience. Thus, this study identifies the investigation of bank customer experience as an urgent need to furnish the banks with insight on how to satisfy customers and enhance profitability.

2.3 Factors Influencing Bank Customer Experience

Few studies have investigated the influencing factors of bank customer experience (CE). For example, In India, Garg et al. (2014) found convenience as the most significant factor influencing customer experience. They also highlighted customer interaction, employees, speed, core service, online functional element as other contributing factors to CE. Chauhan et al. (2022) reported that CE is determined by functional clues (like functional quality, trust, and convenience), mechanical clues (like website attributes and design, perceived usability), and human clues (like customer complaint handling). Ramaswamy et al. (2021) found security concern, network connectivity, convenience as major issues of CE. As regards electronic banking, studies like Kida et al. (2018) investigated the challenges facing cashless banking in Nigeria and showed machine malfunction, technical issues, lack of investment in infrastructural development, religious beliefs and infrastructural deficit were the major problems encountered. They further stated the problem of cashless system in terms of availability and reliability significantly mar the prospect and implementation. Soetan et al. (2021) reported service maintenance, service technology and service dynamics as significant factors of customer experience. In Iran, Ashrafpour et al. (2021) reported affective and functional experience as the significant components of customer experience. Thus, service quality & system quality are the significant factors as they were shown to have positive impact on online bank customers.

Mokhlis et al. (2018) stated ATM (automated teller machine) is one of the most used banking services. Their study ascertained ATM service and availability as the major drivers or determinants of customers' bank choice. Kida et al. (2018) added banks should prioritize ATM service because it is the most used banking channel in Nigeria. They further stated banks need to create more ATM, provide education on ATM usage, and tackle the problem of poor network connectivity as these are the major challenges faced by customers. Kumar

(2018) stated the use of banking technologies increases the bank's likelihood to provide improved service delivered within shorter time. Thus, reduces cost and enhances customer experience, satisfaction, and retention. Their study investigated the challenges of improving customer experience through RPA (Robotic Process Automation) adoption in Nigeria. Their study found out reliability and availability have the most significant influence on customer experience. Seconded are usefulness, security, and privacy. Abdullahi et al. (2018) recommended educating customers on the use of electronic banking, mobile banking and point of sales (POS). Their study identified the need for judicious use of electronic banking in the form of internet banking and mobile banking to remain competitive since they were identified to have low usage. Wasan (2017) examined the dimensions of customer experience on discretionary behaviour and found out customization and convenience are the most significant factors. However, Wasan, found image creditability of bank and service environment not significant. Kim & Kang (2012) identified perceived ease of use & usefulness as major factors influencing the intentions to use mobile banking in Korea. Shaikh & Karjaluoto (2015) reviewed literature on mobile banking in developing countries and identified perceived usefulness and attitude has significant influences to adopt mobile banking. In comparison to developed countries like Finland, Komulainen & Saraniemi (2019) identified ease of use, real time operation, visuality, trust, social status, and sense of control as influencing factors of CE in the mobile banking context. Shaikh et al. (2020) found customer awareness, perceived usefulness, and perceived ease of use as significant factors of CE in the mobile banking. In Spain, Moliner-Tena et al. (2019) confirmed bank customer engagement as the factor influencing customer experience and behaviour. In the United Kingdom, Mbama & Ezepue (2018) reported service quality, functional quality, perceived value, employee-customer engagement, perceived usability, and perceived risk are the main factors that influence bank CE. In summary, ease of use, usefulness, convenience, network connectivity, electronic banking availability, and accessibility, system functional quality, trust, and customer complaint handling are the major themes of CE. In the next section, this study will review literature to understand and measure customers' attitude because attitude is learned from experience.

2.3 Customer Attitude & Behaviour

Szmigin & Piacentini (2018) stated attitude comprises of thoughts (what you think), feelings (what you feel) and action (what you do). They outlined three main components of attitude namely, affect or feelings, behaviour, and cognitions. The affective component involves the emotional connection with the attitude formed, behaviour component is the action associated and the cognitive component involves the belief and thought individual have. Customer behaviours are made based on their emotion or sentiment to service received. The behavioural decision to re-purchase or re-use a service, to switch or refer others are all important to bank performance. Customer attitude has been identified in psychology studies to have great potential to guide voluntary behaviours (Madhavaram & Appan, 2010; Fazio & Towles-Schwen, 1999). Wasan (2017) stated such behaviours are greatly dependent on perception which is made through emotions or sentiment to service. Oskamp & Schultz (2005) described attitude as what is learned (through experience), with measurable attributes (good or bad), that have intensity (strong or weak), consistency and stability. Their study stated a way to conceptualise attitude is that there is an association between the attitude object (aspect) and its evaluation dimensions (good or bad, positive, or negative). For example, customer tweets like *"bank X na sure bank"*. The tweet attitude can be known by identifying the attitude object, which in this case, is bank X, and the evaluation dimension is positive. This implies the customer has positive attitude towards bank X. Another example is, *"bank X na sure bank their ATM no dey charge for withdrawals"*. In this case, the attitude object is ATM of bank X, aspect is price, and the evaluation dimension is positive. This implies positive attitude towards ATM of bank X due to the sentiment valence of the aspect "price" which is positive.

Studies have investigated factors leading to the customer attitude and constantly changing banking behaviour. For example, Srivastava & Kani (2014), stated service convenience is an integral driver of customer behaviour due to significant changes in customers' socio-economic profile, income, and a hyper competitive market. Haruna et al. (2018) examined the factors that determine customers bank choice behaviour and found out low interest (charges) rate, speed of service, easy access to loan, higher interest on deposit all contribute significantly. In Malaysia, Aziz et al (2020) found out perceived security is the most important factor of customer attitude to mobile banking. This study reviewed literature on customers' attitude to branch and electronic banking. For example, Bravo (2019), investigated the effect of customer perception on branch and online banking. They found out branch banking experience is more important to customers than online in terms of trust and commitment. To support this, Kingshott et al. (2018) stated branch banking customer experience influences the adoption of banking technologies and thus electronic banking experience influences the feelings of bank trust and commitment. This disagrees with the findings of Mokhlis et al. (2018) that showed customers consider the use of ATM (automated teller machine) when choosing a bank due to the problem of long queues in the bank branches. The electronic banking was introduced to supress the long queue in the banking hall and improve service delivery. Unfortunately, customers still encounter this problem in their daily branch banking. Dauda & Lee (2015) stated customers are yet to fully adopt web technologies. The banks are trying to redirect more of branch banking customers to electronic banking. However, this has been challenging in some developing countries. For example, in Tanzania, Mwanajimba (2019) showed bank customers have negative attitude towards bank technologies which is due to lack of knowledge to use and some cultural beliefs. Sakhaei et al. (2014) ascertained reliability is the most influential factor of internet banking in Iran. Shrestha et al. (2020) investigated customer's attitudes on internet banking. They showed perceived usefulness, perceived risk, and perceived ease of use influences Internet banking adoption. Their findings showed Nepal bank customers had positive experience using Internet banking. This result is consistent with the study of Inegbedion (2018). The latter found perceived ease of use, perceived risk, and nature of transaction as factors that influence customers' attitude toward electronic banking in Nigeria.

2.4 Chapter Summary

In summary, past studies found to have investigated bank customers' experience, attitude, or behaviour have reported the drivers and challenges faced in banking. It is worth noting that these studies were conducted using data from survey questionnaire or interview. Survey data is limited to pre-defined attribute and more biased. This is because the customer attitude measure applied relies on authors' self-report and occur with respondent's conscious reflection. Oskamp & Schultz (2005) provided comprehensive report on customer attitude They stated explicit measures of attitude relies on self-report like the interview, questionnaire and telephone call which are self-driven by the author and respondent are conscious with their response, however, the implicit measure assess attitude indirectly, as it does not require awareness or honest responses. Their study further discussed the need of both measures. Specifically, they emphasized on the need for research to consider implicit measures of attitude. Lately, the use of internet and social media is increasing and thus, makes the interaction of people on a target spread quickly across the globe. For example, potential customers tend to read product reviews before their purchase. In the banking context, customers tweet to confirm the ATM service performance in a particular area or bank. This impacts the reputation of the target, attributes of the target and the company (Szmigin & Piacentini, 2018). Thus, this serves as a source of research data for this study. This is because this study can assess evaluation of an object that occur without conscious reflection. Attitude is the combination of affect or feelings, behaviour, and cognitions (Szmigin & Piacentini, 2018). Since customers' feeling or behaviour towards a service is based on emotional or sentiment connection made. Sentiment analysis is deemed appropriate to measure the attitude and understand customers' emotional or sentiment connection towards bank service. This study takes a different approach to investigate customers' attitude by measuring the sentiment of bank customers towards product and service received using social media data. Social media data is beneficial because it is unlimited to attribute or aspect, provides real time information and avoids time and cost of setting up and analysing a questionnaire or focus group. This approach enables us to understand customers explicit and implicit attitudes. This will benefit the bank management or policy makers with insights on undefined dimensions of customer experience and attitude in terms of usage. Therefore, an illustration on what was observed from literature and the idea of emotional sentiment bound with bank product and service can be seen below.

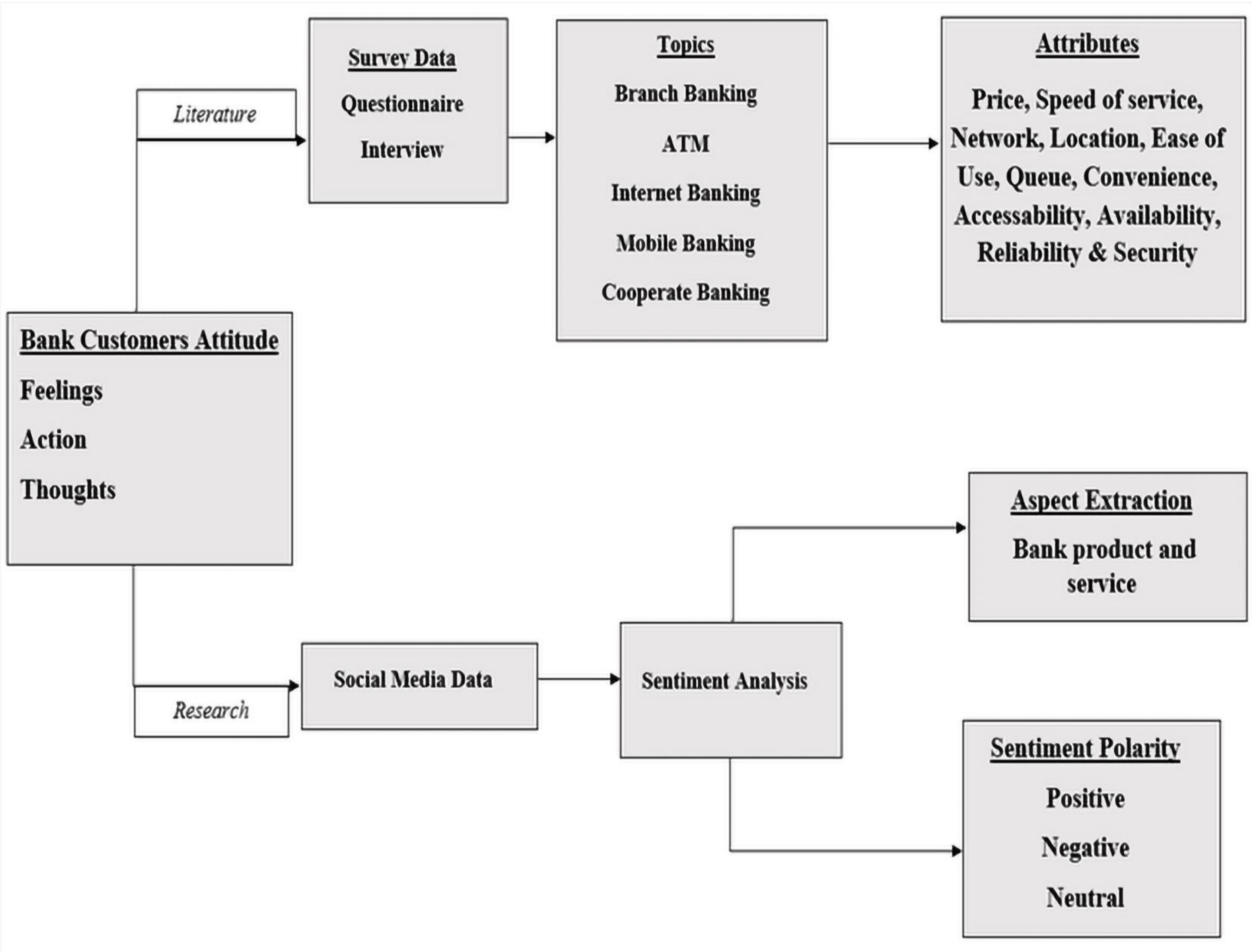


Figure 2.1: A representation of conceptualising customer attitude

To conclude, the background knowledge in this research context, can be summarized as shown above (figure 2.1). This study was unable to adopt existing consumer behaviour theory because the study is focused on measuring the sentiment polarity of customer attitude towards bank product and service. Sentiment analysis as identified to suitably classify the positive or negative polarity of bank customers' attitude, will thus be reviewed in the next chapter.

CHAPTER 3: Literature Review

3.0 Introduction

In the previous chapter, it was established from review of literature that studies on bank customers' attitude are limited to survey data. This implies their findings are limited to attributes predefined by the researcher. In this study, social media is considered as research data to mine the sentiment of bank customers due to unlimited attributes of social media content. Thus, this chapter presents literature review on social media, sentiment classification and aspect extraction for sentiment analysis.

3.1 Social Media

Social media is regarded as an online community where people connect with each other to share their experience or communicate issues on a subject (Mogaji et al. 2016). The wide usage of social media as tool of interaction and source of information has enriched and lighten analytics for generating insight and improving business value. Industries ranging from airline (Liau & Tan, 2014), retail (He et al. 2015), transportation (Salehan & Kim, 2016), and hospitality (Duan et al. 2013) use social media to drive their business decisions. In developing countries like Nigeria, social media data has been used to predict election (Amusa et al. 2016) because social media serves as access to public opinion, experience, desires, and expectation. Liau & Tan (2014) stated in this modern age social media is the most powerful tool for identifying churn drivers, improving customer acquisition and retention. Thus, serves as a marketing tool, source to competitive edge and brand awareness. Charoensukmongkol & Sasatanun (2017) found out entrepreneurs who use social media intensively for customer relationship management are more satisfied with their business performance. Organisations that analyse and manage customer feedback/complaints are 5% more productive and 6% more profitable on average than their competitors (McAfee & Brynjolfsson, 2012). Therefore, it is advantageous to monitor and analyse customers' feedback, complaints, opinion, emotions, and attitude to improve profitability.

3.1.1 Social Media Analytics

Social media analytics (SMA) is a broad area of research and an innovation developing interest across industries after years of rapid adoption (Chen et al. 2014). SMA is the computational study of users' association, thinking and feelings by analysing their data spread across online sources (Chen et al. 2012). Most organisations have social media accounts where they engage with customers and monitor their attitudes. Organisations use social media data to make decisions, develop strategies, attend proactively to customers' complaints, and understand users' feeling regarding their product, service, and reputation.

All these have not only increased the attention of banks to engage customers through social media but also increased the need for continuous research on how to profitably use social media. Social media platforms like Facebook, Instagram and Twitter have emerged as the most popular platform where people communicate their opinion on a subject. Facebook was launched in 2004 and the network stood out in terms of popularity. However, Facebook follows a strict oriented policy to protect users' privacy which appears to be a limitation (in the research community) because data available depends on the level of authorization granted by the user. While Twitter is a microblog formally launched on 13th July 2006. It is free to extract tweets that are publicly available (through Twitter API) from Twitter, but this is limited to the last seven days. In summary, to get a 360-degree customer view, there is need to listen and understand customers' social voice. In this study, customers' opinion will be mined using sentiment analysis techniques to understand, customers' sentiment polarity, patterns, and trend, what people are saying, potential triggers of churn and customers' banking experience.

3.2 Sentiment Analysis

Sentiment analysis (SA) is an important type of SMA which uses computational linguistics, statistical techniques, and natural language processing to generate insight from text data. SA is an ongoing research area that helps determine customers' opinion or the situation of market on a particular entity (Bhargava & Rao, 2018). Liu (2012) gave a simple definition of sentiment analysis as a tool that analyses peoples' opinion, evaluation, attitude, and emotions towards tangible or intangible issues like product, services, or topics. Liau & Tan (2014) described SA as a powerful tool that helps develop a comprehensive understanding of customers' intention and behaviour thereby helping companies to identify gaps in their service. Fan & Gordon, (2014) defined SA as a tool used for identifying groups of customers, determine brand image, monitor stock market, discover trend, and manage crises. Organisations like the banks are interested in using sentiment analysis to monitor and understand customers' attitude towards their product or service and gain insights on meeting customers' needs. In this context, sentiment analysis can be defined as the computational linguistics and statistical tool useful in gaining insights towards bank customers’ attitude and emotions towards bank products and services. The knowledge added through this analysis is valuable because large numbers of people have expressed their opinion about the topic publicly and freely, which reduces response and observer biasedness.

3.3 Concept of Sentiment Analysis

Social media is an important source of customer perception because it provides up-to-date customer information expressed freely online. Social media provides an extensive source of information due to the large volume of potential data available. For example, in the case of banking, social networks such as Facebook and Twitter are good data resources because bank customers express their opinion, thoughts, feelings, and complaints on these social networks. Thus, monitoring their attitude using SA approach is therefore tapping into a useful resource.

This study agrees with Liu (2015) that identified aspect extraction and sentiment classification as two different tasks in sentiment analysis system. The aspect extraction involves the process of extracting topics in the context of Nigeria bank product and services. While the latter involves the classification of opinion/sentiment words. SA techniques are useful in classifying words or phrases into labels such as positive, negative, or neutral. From the banks' perspective, sentiment analysis is useful to understand customer preference, acceptance and performance of products or services. However, the natural language, ambiguity and colloquial nature of social media text made sentiment classification task complicated. As will be illustrated below, an extensive review of literature shows there is lack of empirical studies on the application of sentiment classification to the Nigeria banking industry. Therefore, this study contributes several new strands of research. Firstly, this study adds to existing knowledge in the application of sentiment analysis to the banking domain and to the issue of customer experience and attitude. Secondly, this study contributes on the application of sentiment analysis to the Nigeria Pidgin English. For guidance, subsequent sections (3.4 – 3.7) will provide knowledge on sentiment analysis while section 3.8 will specifically review literature on aspect extraction.

3.4 Level of Sentiment Analysis

Sentiment analysis has been studied at three (3) different levels of granularity namely, document level, sentence level and aspect level.

3.4.1 Document Level

Document level sentiment analysis aims at determining the overall sentiment polarity of an entire document towards a single entity. For example, the sentiment classification of an entire document into positive, negative, or neutral polarity towards a product like "ATM". This level of analysis assumes the document contains same opinion/sentiment towards a particular entity. The assumption is effective when investigating sentiment of a product review as either positive or negative and studies like Balage & Pardo (2013), Moraes et al. (2013) and Duo (2017) have classified sentiment at document level to good effect. However, the technique had been criticized by Liu (2012) as unrealistic and lacking in-depth because the method is not applicable in the case of mixed opinionated documents.

3.4.2 Sentence Level

Sentence level sentiment analysis involves two steps. Firstly, is to determine if a sentence is subjective or objective in a document. Subjective sentences are statements that express opinion while objective statement expresses fact. The subjective sentences are extracted for the second step. The second step is to determine the sentiment orientation of the subjective sentences as positive, negative, or neutral. This level of analysis assumes a subjective sentence holds a single opinion however, the assumption does not hold for all sentences in a document. Sentence level sentiment analysis is preferred over document level when there are different opinionated sentences in a document (Shirsat et al. 2019). However, this level of granularity is considered insufficient when classifying comparative subjective sentences towards a product (Song et al. 2019). For example, sentences like *“ATM of Bank A is very good, but their credit card is rubbish”*. The sentence does not express a positive or negative sentiment when classified on sentence level. The sentence expresses more than one sentiment and thus, a finer sentiment classification level is required to classify such comparative sentence accurately.

3.4.3 Aspect Level

Aspect based sentiment analysis (ABSA) is the sentiment classification of an entity by recognising the sentiment polarity of the aspect. This level of analysis classifies sentiment based on the sentiment orientation of words expressed towards the aspect of an entity. For example, *"the network of bank X is great, but their mobile app is rubbish"*. The sentence shows a positive sentiment towards aspect "network" and negative sentiment towards aspect "mobile app" of an entity "bank X". Therefore, the steps involved are to first extract the relevant aspects (aspect extraction) of the product reviewed and secondly to determine the sentiment polarity (sentiment classification) of the aspect towards an entity.

3.5 Sentiment Analaysis Approaches

There are two common approaches to sentiment analysis namely, machine learning (ML) and lexicon-based method. However, recent research tends to integrate both methods and/or two ML methods to propose a hybrid approach to sentiment analysis (Song et al. 2019). For example, Minaee et al. (2019) & Zhou et al. (2015) in their studies, combined convolutional neural network and LSTM to solve SA problem. These approaches will be discussed further in subsequent sections.

3.5.1 Machine Learning Methods

Machine learning (ML) is a field of study that provides computers that opportunity to learn without being explicitly programmed (Samuel, 1967). Machine learning is a computation algorithm which is built to emulate human intelligence by learning through experience (Greener et al. 2022; Carleo et al. 2019; EL-Naqa & Murphy, 2015; Jordan & Mitchell, 2015). In sentiment analysis context, machine learning approach uses algorithms such as support vector machines (Al-Smadi et al. 2018), maximum entropy (Xie et al. 2019), Naïve Bayes (Goel et al. 2016), LSTM (Huang et al. 2021), BI-LSTM (Lin et al. 2021), XLNet (Sweidan et al. 2021), BERT (Sun et al. 2019), RoBERTa (Liao et al. 2021), ULMFiT (Kulkarni et al. 2021), ELMo (Nurifan et al. 2019), and Albert (Ding et al. 2021) to classify text polarity. The methods perform well and are also reliable (Bhargava & Rao, 2018; Singh & Goel, 2019). However, the performance of these classifiers varies in different domains (Pand & Lee, 2008). There are three types of machine learning models namely, supervised, semi-supervised and unsupervised machine learning models. This will be discussed further in subsequent section.

3.5.1.1 Supervised Machine Learning

Supervised machine learning methods require a labelled dataset to train the learning algorithm. The process involves using labelled set to train the model and thus, the model classifies new instances based on what is learnt during the training. Supervised models require a large amount of training dataset to achieve accurate sentiment classification. Supervised learning algorithms such as support vector machine have been widely used by previous studies for sentiment classification. This statistical learning classifier has produced results with high accuracy in the application of sentiment analysis (Devi et al. 2016; Al-Smadi et al. 2018). Another supervised model is the Naive Bayes classifier, a probabilistic classifier based on Bayes' theorem, has also produced results with high accuracy (Goel et al.

2016). Bhargava & Rao (2018) utilised support vector machine and Naïve Bayes on different cryptocurrency datasets and both supervised methods performed well with high accuracy.

3.5.1.2 Semi-Supervised Machine Learning

This approach is used when training data are limited for supervised models. The method involves using unlabelled data to complement limited labelled data available at the training phase of sentiment classification (Zhou et al. 2014). Lately, sentiment analysis studies have benefitted from this approach. For example, Variational auto-encoder was used as semi-supervised model for aspect-based sentiment analysis (Chen et al. 2018; Fu et al. 2019) and dimensional sentiment analysis (Wu et al. 2019). Deep belief networks have also been used as semi-supervised model for ABSA (Zhou et al. 2014). Recently, the pre-trained language models are becoming popular for SA task. It is worth noting that the language models were developed to overcome the problem of vector space models which are frequency based and does not recognise the context of text. Another case is the word embeddings that learn the embeddings only with a context-based objective and as a result limits the efficiency to reflect the sentiment of the text. Thus, semi-supervised ML approach are useful to complement such cases. For example, Park et al. (2019) employed a semi-supervised approach by using partial sentiment information to complement the distributed representation learning model to solve SA problem.

3.5.1.3 Unsupervised Machine Learning

Unsupervised machine learning methods are not popular in sentiment classification problems due to not being easily applicable in non-review dataset. The learning algorithm uncovers pattern from unlabelled textual data. Examples are the clustering models such as k-means clustering (Kumari & Babu, 2016), Density-Based Spatial Clustering of Applications with Noise (Khanaferov et al. 2014) and fuzzy clustering (Suresh, 2016). The pattern or structure aids in classifying the sentiment of a topic or subject. This approach is considered better than supervised models because it does not require training data, linguistics knowledge and saves time (Suresh, 2016). However, in aspect extraction, this is a very popular and efficient technique. This is because the algorithms clusters latent terms from the distribution of words across the document which forms topics. Examples of such models are Latent Dirichlet Allocation (Blei et al. 2003), Latent Semantic Indexing (Deerwester et al. 1990) and Probabilistic Latent Semantic Indexing (Hofmann, 1999). These models have been successfully implemented in previous studies (Wahid et al. 2022; Ozyurt et al. 2021; Lv et al.

2021; Yiran & Srivastava, 2019; Poria et al. 2016; Li et al. 2010). This will be discussed further in section 3.9

3.5.2 Lexicon Based Methods

Lexicon based approach uses a sentiment lexicon algorithm to classify text. The approach weighs words with their associated semantic orientation to detect the overall polarity and strength of document. Dictionaries for lexicon-based method can be generated based on syntactic patterns in a corpus. Lexicons can be created automatically or manually. Automatically built lexicons perform well across all domains (Taboada et al. 2011) because it uses general knowledge resources which have wider term coverage (Muhammad et al. 2013). While manually built lexicon approach may be preferred because it simulates the effect of contextual word or phrase. However, the latter approach is mostly domain specific, expensive and time consuming to build. There are several lexicons developed before the use of social media such as WordNet. WordNet is an online English lexical database developed by Miller (1995). These available lexicons tend to perform poorly when used to classify social media data due to the ambiguous and colloquial text present in social media data. Considering that, previous studies have developed lexicons like Sentistrength to recognise the presence of ambiguity and colloquial words in social media data and was shown to have performed well in different domains. There are other popular lexicons namely, SentiWordNet, AFINN, SO-CAL, LIWC and NRC which are described below.

- SentiStrength

  Thelwall et al. (2010) proposed Sentistrength to detect polarity of text, and their strength value ranges from 1 to 5. The lexicon was built to detect informal words, negation, emoticons, and intensifiers in text. Thus, the algorithm recognises negation when classifying sentiment better than machine learning classifiers (Thelwall et al. 2012). The original lexicon wordlist, phrases, and their respective strength value were enhanced in 2012 to improve on performance of the lexicon. The aim was to capture the commonly used language and slangs on social networks. The enhanced lexicon was validated on six different web dataset and results indicate the lexicon was successful. However, the lexicon performed less accurately in classifying positive sentiment in news-related dataset (Thelwall et al. 2012). The performance of Sentistrength attracted various researchers and thus, the lexicon was employed to suit different domains. For example, Islam & Zibran (2018) developed software engineering version of Sentistrength called Sentistrength-SE which was reported to have outperformed other software engineering lexicons.

- SO-CAL

  Semantic orientation calculator (Taboada et al. 2011) was developed from dictionary of words annotated with their semantic orientation. The lexicon was manually constructed using words from general inquirer, Epinions reviews, and movie reviews. The sentiment bound words were assigned values of -5 as high negative to +5 for high positive words. The author tested the lexicon on different review dataset and showed the performance of the lexicon was consistent across different domains.

- SentiWordNet

  This popular lexicon was created by Esuli & Sebastiani (2006) from WordNet synset which is an English lexical dictionary. SentiWordNet was automatically built by annotating sets of synonyms from WordNet with their semantic orientation. The lexicon associates the set of synonyms with three scores namely, negative, positive, and neutral. Many studies have successfully used the lexicon to extract sentiment. For example, Sohangir et al. (2018) showed lexicon-based approach such as SentiWordNet outperformed the machine learning approach like Naïve Bayes and SVM in extracting sentiment in financial social media dataset. Igbal et al. (2019) applied SentiWordNet as part of their proposed hybrid approach to sentiment analysis. Asghar et al. (2019), developed Urdu lexicon successfully using SentiwordNet as part of the lexicons adopted.

- NRC

  NRC was developed by Mohammad & Turney (2010). This manually built lexicon contains over 14,000 unique English words which are emotionally tagged. NRC fetched frequent words from Macquarie Thesaurus, google n-gram corpus, Wordnet affect lexicon and general inquirer. The lexicon contains words coded into two sentiment classes (positive and negative) and eight emotion categories namely, joy, anger, trust, surprise, fear, disgust, anticipation, sadness. The NRC emotion lexicon is commonly used when both emotion and polarity of sentence need to be understood.

- Linguistic Inquiry and Word Count (LIWC)

  This lexicon was developed firstly in 1993 (Pennebaker, 1993), updated in 2001 as second version (Pennebaker et al. 2001) and third version in 2007 (Pennebaker et al. 2007). The modern version was introduced in 2015 (Pennebaker et al. 2015) which uses entirely new dictionaries and software rather than just an updated version done in the previous versions. LIWC is a lexicon that includes common words from other dictionaries, and social media frequently used text like lol, 4eva, and b4. However, the lexicon has no potential to identify abbreviations, misspellings, colloquialisms, negation, part of speech tagging, and foreign words as they are not included in the dictionaries (Dhaoui et al. 2017). LIWC does not restrict users from adding up customized words into the lexicon which enhances the default program in polarity detection. However, the lexicon is not free to use. Studies like (Tausczik & Pennebaker, 2010; Hutto & Gilbert, 2014; Crossley et al. 2017; Dhaoui et al. 2017) have employed the lexicon successfully for sentiment classification.

- AFINN

  Nielsen (2011) introduced AFINN lexicon. The lexicon can identify social media text like Twitter data including slangs. It was created as an updated version of affective Norms of English words (ANEW) to capture social media data and slangs due to the recent increase in usage of ambiguous and colloquial text. The lexicon algorithm scores text in the range of 1 to 5 for positive words and -1 to -5 for negative words.

3.5.3 Hybrid Approach

Studies have proposed different hybrid approaches to sentiment analysis to improve on the performance of the machine learning (ML) and lexicon-based approaches discussed earlier. These approaches combine two or more of ML methods or lexicons or both methods to yield an enhanced performance in terms of accuracy, precision, and recall. For example, Kundi et al. (2014) proposed a lexicon-centric hybrid approach that combined lexicons of emoticon, slangs, and English dictionaries for Twitter-sentiment analysis. Their study showed sentiment classification system with 92% accuracy. Machová et al. (2020) used lexicon and Naïve Bayes to solve SA problem. They employed particle swarm optimization (PSO) for automatic labelling of words to develop their lexicon. Huang et al. (2020) proposed SentiCNN (lexicon and CNN). Ma et al. (2018) proposed SenticLSTM (SenticNet and LSTM). Appel et al. (2016) utilised SentiWordNet and fuzzy sets for sentence level sentiment classification. Their report showed the proposed approach outperformed ML methods like Naive Bayes and maximum entropy when used in isolation. Asghar et al. (2018) proposed a hybrid system that combined lexicon and probability-based domain-specific classifier to improve on sentiment detection of slangs, emoticons, and domain specific terms. In comparison, their study showed improvement in the performance of Twitter-based sentiment analysis systems. However, many of the hybrid approaches are not easily interpretable and deployable.

3.6 Comparative Study on Sentiment Classification

To critically understand and identify the research gaps in sentiment classification task, this study conducts a comparative study on techniques and challenges of sentiment classification. This will help identify the best approach for this study.

3.6.1 Comparison of Lexicons

Over the years, several lexicons have been developed to classify sentiment or emotions. Few of the lexicons were described in section 3.5.2 above. Table 3.1 below provides a summary of the key characteristics of the lexicons.

Lexicon	Author	Total	No of Positive	No of Negative	Scale	Label Method	Algorithm	General/ Domain Purpose
AFINN	Nielsen (2011)	2477	878	1598	-5 to +5	Manual	Score individual words and sum	General
Loughran	Loughran & McDonald (2011)	3917	354	2355	-1 to +1	Manual	(No of Pos – No of Neg)/Total	Domain Specific
Bing	Hu & Liu (2004)	6789	2006	4783	-1 to +1	Machine Learning	(No of Pos – No of Neg)/Total	General
SentiWordNet	Esuli et al. 2010	20093	8898	11029	-1 to +1	Machine Learning	Score individual words and sum	General
SOCAL	Taboada et al. (2011)	5971	2438	3530	-5 to +5	Amazon Mechanical Turk	Score individual words and sum	General
NRC	Mohammad & Turney (2010).	5555 (without emotion tags)	2312	3324	0 to 1 when associated	Amazon Mechanical Turk	(No of Pos – No of Neg)/Total	General
WKWSCI	Khoo & Johnkhan (2018)	10221	3121	7100	-3 to 3	Manual	Score individual words and sum	General
Vader	Hutto & Gilbert (2014)	7500	____	____	-4 to 4	Amazon Mechanical Turk	Score individual words and sum	General
SentiStrength	Thelwall et al. (2010)	58119	____	____	-5 to +5	Manual	Score individual words and sum	General

Table 3.1: Summary of lexical resources

Lexicon based methods have been praised to perform well across domains and can be easily deployed (Moussa et al. 2020; Mehmood & Balakrishnan, 2020; Li, 2020; Turner et al. 2020; Huang et al. 2020). However, their performance still varies. Traditional lexicons like General Inquirer (GI), and MPQA have been criticized as not well suited for social media data due to presence of colloquial words i.e., does not analyse sentiment of words not in the lexicon which is a problem for Twitter data that consists of continuously changing malformed words (Saif et al. 2016). Sentistrength lexicon was built specially for social media data and thus received much of praises for its applicability and consistent performance across domain (Thelwall et al. 2012). Sentistrength is considered one of the best English lexicons available (Ahmad et al. 2017). Studies like Islam & Zibran (2018) benefitted from the strength of Sentistrength lexicon and customized a software engineering version of Sentistrength called Sentistrength-SE. Ahmad et al. (2017) compared Sentistrength with eleven other lexicons using Twitter dataset, BBC comments and DIGG comments and found out Sentistrength performed best across the three different domains with performance of 92 - 95% accuracy. Sentistrength performs great because the method applies lexical rules to solve the problem of malformed words (Islam & Zibran 2018). However, Sentistrength is considered a fixed or static lexicon restricted by the lexicon and the word sentiment values regardless of the context the words are used.

Sentistrength was applied in the health sector (Bhattacharya et al. 2014). However, studies like Gohil et al. (2018); Gabarron et al. (2019) emphasized Sentistrength has not yet been validated to accurately classify healthcare context-based words because the lexicon was trained with general knowledge, online, informal, and social media terms. Sentistrength and other traditional lexicons are context independent which can be problematic in cases where a word means different thing in another context (Saif et al. 2016; Islam & Zibran, 2018; Gohil et al. 2018). Saif et al. (2016) proposed SentiCircle to improve on the lexicon context problem. The SentiCircle lexicon was developed to detect contextual semantics of words and the algorithm tunes the sentiment strength and polarity of the word to match the semantic captured. Saif et al. (2016) examined the performance of SentiCircle in three (3) different twitter datasets (Obama debate, health care reform and STS-Gold) and compared to three other lexicons. Hence, performed a runtime analysis to show their method is scalable and better. Their result showed SentiCircle is different to other lexicons based on its ability to recognise words pattern in different domain. The lexicon captures semantics and updates the pre-assigned strength and polarity in sentiment lexicon accordingly. SentiCircle outperformed MPQA and SentiWordNet in accuracy for both entity and Tweet level sentiment detection.

However, was outperformed by Sentistrength in terms of F1- measure. SentiCircle showed a high variance due to the imbalance class in dataset used for validation and thus affected the performance of the lexicon in terms of recall. This implies the performance of the lexicon cannot yet be ascertained until it's been validated across different dataset. In addition, it was observed SentiCircle system does not classify emoticon and slang which will affect their tweet level sentiment classification performance when compared to Sentistrength.

Other lexicon such as VADER has been deployed by Sohangir et al. (2018) in financial social media dataset (StockTwits) with good performance. Igbal et al. (2019) used SentiWordNet to enhance their hybrid approach in extracting sentiment. Asghar et al. (2019) used SentiwordNet to develop Urdu lexicon. However, Khoo & Johnkhan (2018) ranked SentiWordNet poorly when compared with other lexicons. Khoo & Johnkhan (2018) developed WKWSCI (Wee Kim Wee School of Communication & Information) lexicon for sentiment classification across different domains. WKWSCI lexicon contains 3121 positive words, 7100 negative words and 19500 neutral words. The negative and positive words in the lexicon were scored from -3 as very negative to +3 for very positive to identify word intensification in sentences. In comparison, they showed WKWSCI and SO-CAL performed well in analysing amazon product review and news headline dataset. However, WKWSCI lexicon has not yet been tested across different domains apart from the original domains which were used for validation. In addition, the lexicon is lacking in terms of informal words, emoticons, booster words and multiword expressions. The WKWSCI lexicon conflicts other lexicon with some words coded as neutral which WKWSCI lexicon scored as positive or negative. Furthermore, the positive to negative contradiction with other lexicons was also observed. For example, WKWSCI lexicon coded the word "proud" as negative which contradicts lexicons like GI, SO-CAL and NRC that coded "proud" as positive word. Thus, Khoo & Johnkhan (2018) acknowledged that lexicons are not perfect due to the human coding involved and admits their proposed lexicon WKWSCI had wrongly coded the word "unequivocal". They further stated lexicon-based method are better than machine learning method in classifying sentence level and aspect level sentiment analysis when there is small number of features. Their study therefore recommends WKWSCI sentiment lexicon for non-review dataset however researchers should be aware the lexicon might be biased towards narrow context. It is also good to note that WKWSCI lexicon was built with American wordlist and has not been compared extensively.

In sentiment classification, emotions are important because sentiments are created based on emotion. For example, when negative emotions such as sadness, anger, disgust, and fear are high then it is understandable that the resulting negative sentiment polarity will be high. Kusen et al. (2017) investigated emotion classification of social media messages using NRC, DepecheMood and EmoSenticNet lexicons. Their study stated there are available literature that have investigated the sentiment polarity of articles, blogs, and social media text. However, individual emotions in the text have been understudied due to challenges of complexity and the ambiguity of natural language expressions. They compared the performance of three word-emotion lexicons on social media text to human emotion classification received by means of a questionnaire survey to evaluate the performance of the lexicons. Kusen et al. (2017) found out NRC performed better than other lexicons with the highest precision and recall especially in identifying emotions of anger, fear, and joy. However, their study found a high number of false positive when performance was evaluated which might be because of choice of words unavailable in the lexicon. Zimbra et al. (2018) evaluated twenty-eight sentiment analysis techniques which includes NRC and Sentistrength across five different dataset and found out NRC performed better than Sentistrength in four domains in terms of accuracy. Kusen et al. (2017) ascertained emotion lexicons such as NRC can adequately identify the affective valence of a text. The NRC emotion lexicon is commonly used to classify emotion and sentiment polarity. However, the lexicon is limited in coverage.

AFINN lexicon had also been compared to other lexicons such as SentiWordNet, Opinion lexicon, and SentiStrength. AFINN showed equal or better performance (Koto & Adriani, 2015). Ribeiro et al. (2016) conducted an extensive comparison of twenty-four sentiment lexicons to understand how the lexicons vary across different domain using eighteen datasets from social network, online reviews, and comment. Their study showed Sentistrength and AFINN performed well especially with the social network datasets. However, Ribeiro et al. (2016), suggested popular lexicons such as AFINN and Sentistrength needs continuous comparison and validation to ascertain their performances in different domains. Todd et al. (2019) stated many studies have used general purpose lexicons such as AFINN and LIWC in the market and health research studies without considering the classification accuracy or if any of these lexicons had been validated in the applied domain. Thus, opens the need for validation of these lexicons. To conclude, the superiority of these lexicons is difficult to determine randomly because it is unclear the best applicable lexicon in this context. Therefore, there is need to review literature on what has been done in the financial domain (see section 3.6.4).

3.6.2 Comparison of ML Approaches

Traditional machine learning algorithms have been used widely for sentiment classification task. For example, Anis et al. (2020) compared support vector machine (SVM), Naïve Bayes (NB), random forest (RF), logistic regression (LR), ensemble learning model and K-nearest neighbours (KNN) for binary sentiment classification task. Their study performed data pre-processing such as removal of URL, punctuation, stop words, and stemming on 37,827 Egyptian hotel reviews (26,521 positive reviews and 12,411 negative reviews) collected from Kaggle. They used Word2vec as the feature engineering technique and showed SVM outperformed all the other algorithms. They reported performance in terms of accuracy (86.3%) and F1 scores (85.9%). Jadav & Vaghela (2016) compared NB, SVM and optimised SVM for binary sentiment classification task in a movie review, Twitter, and publicly available gold dataset. Their result showed SVM performed better than NB. The optimised SVM (using Gaussian Radial Basic Kernel function) showed the best result of all. Al-Hamoud et al. (2018) compared SVM and NB in a binary classification task. In their experiment, they used bag of words vector model for feature representation and chi-square to select prominent feature selection. They used 2004 English political tweets to show SVM has a better performance with F1 score of 88.5% and accuracy of 88%. Said & Muqrashi (2020) classified 23,999 Oman hotel reviews collected from Tripadvisor.com into binary sentiment polarities. In their experimentation, research data was labelled positive if customer rating is 4 or over and negative if below 4. They used term frequency inverse document frequency (TF-IDF) for feature representation and thus, compared SVM, Gaussian Naive Bayes (GNB), Random Forest (RF), and Bootstrap aggregating (Bagging). Their result showed SVM outperformed others with F1 score 75% and accuracy of 75%. Guo et al. (2020) conducted a multi-class sentiment classification of student feedback. The feedback collected are from July 2018 to June 2019 on C++ programming course from 1324 students. They used Word2Vec for feature numeric representation and compared SVM, NB, Logistic Regression (LR), and Random Forest (RF). Their result showed SVM with F1 score of 75% as the best performed model to predict student attitude towards blended learning class. Studies like Pang & Lee (2004), Mullen & Collier (2004), Go et al. (2009), Devi et al. (2016), Al-Smadi et al. (2018), Batra & Daudpota (2018) also used support vector machine (SVM) to classify sentiment and achieved an accuracy score equal or better than 80%. Other studies like Tripathi & Naganna (2015), Kalaivani & Shunmuganathan, (2013), Shoukry & Rafea (2012) also showed SVM performed better than NB. SVM have been praised for their resistance to overfitting and effectiveness in high dimensional spaces (Valencia et al. 2019;

Shoukry & Rafea, 2012; Cao & Tay, 2001). In sentiment classification task, literature suggest SVM is an effective and popular classification model. This is because the text data used contains newly formed and malformed words which results to high dimensionality data problem. Fortunately, SVM has been long known for their effectiveness in dealing with such problem (Gaye et al. 2021) which provides the rationale for their success in sentiment classification task.

In contrast, Balakrishnan et al. (2020) compared SVM, NB, RF, and decision tree for multi-class sentiment and emotion classification task. Their study performed data pre-processing such as removal of emoticons, punctuation, non-English text, and lemmatization on the 2463 mobile payment app reviews collected in June 2019. From the 2463 reviews, 1054 were selected and annotated by human. Their study used TF-IDF as the vector space model and SMOTE to rebalance the class distribution. They showed random forest to have outperformed all the other algorithms for both sentiment and emotion analyses. They reported performance in terms of accuracy (75.62%) and F1 scores (71.99%).

Past studies (Goel et al. 2016; Bhargava & Rao, 2018) have employed NB to achieve high accuracy of 80% or above. In comparison, NB has been compared to perform better than other models such as SVM (Wawre & Deshmukh, 2016; Gautam & Yadav, 2014), decision tree (Mostafa, 2020), KNN (Dey et al. 2016) and Maximum Entropy (Parikh & Movassate, 2009). In general, the traditional ML approaches have been used in the field of SA, but a common challenge is the unavailability or limited training data for the classifiers to perform greatly. It is also good to note that when these trained ML classifiers are deployed in a different domain, their performance drops significantly. This reason led studies to employ more sophisticated methods such as the multi-ML systems or the deep learning methods. For example, Catal & Nangir (2017) built a multiple classifier system using Naïve Bayes, Support Vector Machine and Bagging to classify the sentiment of three different online customer reviews. Their study showed the system performed better than ML methods when used in isolation for the three (3) datasets.

Recently, deep learning methods have been applied successfully to sentiment analysis task (Shijia et al. 2018; Zhang et al. 2018; Sohangir et al. 2018; Jangid et al. 2018). For example, Ray & Chakrabarti (2020) proposed CNN and some rule-based methods to improve on the performance of aspect level SA. In their experiment, they made use of CNN with seven layer to tag the aspects in the sentence. An approach of clustering the aspects detected into aspect categories was employed. Their implementation used dataset like electronic product review from Twitter, movie reviews from Stanford Sentiment Treebank and restaurant reviews from SemEval Task 4 and thus, achieved overall accuracy of 87%. Colón-Ruiz & Segura-Bedmar (2020) showed the strength of Bi-LSTM, CNN, and BERT in classification of 215063 drug review. Their study used TF-IDF features and SVM as baseline model for the multi-class classification task. They developed hybrid models using CNN & LSTM, BERT & LSTM, BERT & Bi-LSTM in their experimentation. Their result showed BERT & Bi-LSTM with micro-F1 score of 90% outperformed other models. However, they reported BERT considerable increases the computational cost while the CNN model produced acceptable result in less training time. Liang et al. (2022) proposed SenticGCN for sentiment classification task. Their approach used LSTM layer to learn the hidden contextual representation of the input sentence, this is then passed to the GCN (Graph Convolutional Network) layer to capture the sentiment dependencies, thus enhance the dependencies by leveraging SenticNet. Their model outperformed variants of BERT & LSTM models in restaurant and laptop review domain with F1 score around 75.91% and 74.71% respectively. Li et al. (2022) proposed Bi-directional Emotional Recurrent Unit (BiERU) for conversational sentiment analysis. However, studies like Ma et al. (2019), Song et al. (2019), He et al. (2018), Poria et al. (2016) have shown the performance of deep learning models are better than that achieved by traditional machine learning models. Deep learning approaches have been criticized for complexity of training data, difficult to interpret and deploy. Dohaiha et al. (2018) argued that deep learning methods do not always perform than traditional ML methods and the performance depend on domain and deployment of the methods. To support this, few studies have shown traditional ML methods outperforms the deep learning methods. For example, Xu et al. (2017) compared SVM and recurrent neural network (RNN), convolutional neural network (CNN), and long short-term memory (LSTM) in an aspect-based SA. Their study showed SVM outperformed all the deep learning approaches employed in the laptop and restaurant domains. Xing et al. (2020) result showed SVM performed equally well or better than bi-directional long short-term memory (BI-LSTM) and bi-directional encoder representations from transformers (BERT). Rasool et al. (2020) compared logistic regression, artificial neural network and support vector machine using heart disease dataset. They showed SVM was the best model. Similarly, Al-Smadi et al. (2018), also showed SVM outperformed RNN in an Arabic hotel review dataset. Xing et al. (2020) result showed SVM performed equally well or better than deep learning model. Specifically, SVM was better than Bi-LSTM and equal with BERT. In general, it is worth noting that the presence of malformed new words in social media is ongoing and thus contributes to the problem of high dimensionality. SVM has been effective and popular because of the algorithm's strength in dealing with dataset with high dimensionality problem (Gaye et al. 2021; Maldonado et al. 2014; Noble, 2006; Kudo et al. 2001; Osuna et al. 1997). In summary, different machine learning methods have been applied to various domains for application of sentiment classification. Unfortunately, there is no fixed or proven approach that performs best in all domains. Thus, there is need to compare approaches to identify the best performing model.

3.6.3 Lexicon Based versus Machine Learning Approaches

Lexicon based approach had been observed to perform well across all domains because lexicons are built with lexical resources. The lexical resources consist of different dictionaries which furnishes the lexicon with general knowledge of all domains. This makes the lexicon-based approach to perform well across domains than the machine learning methods. For example, Sohangir et al. (2018) compared VADER and SentiWordNet in a financial social media dataset to Linear SVM, and Naive Bayes classification. Their study showed both lexicons outperformed the ML methods. Saif et al. (2016) stated supervised machine learning approaches are domain dependent, thus requires retraining the algorithm to use in a different domain which affects its applicability and portability to another domain. They further stated the automatic labelling of training data introduces error which affects the performance of the ML classifiers. Thus, ascertained lexicon methods are better in performing SA across different datasets. Melo et al. (2019) explained supervised ML models are more appropriate to build domain specific classifier. However, ML classifiers react differently when tested in different context which results to high variability of the sentiment classifier performance across domains. Their study also criticized previous approaches that used small size of training set for supervised sentiment classification approach. For example, the study of Siddiqua et al. (2016) and thus, tagged their approach as weakly supervised. Melo et al. (2019) presented 10SENT, an unsupervised learning method for general purpose sentiment analysis classification. 10SENT was designed to provide a stable sentiment analysis method across different dataset and domains. In comparison, 10SENT was tested for sentence level sentiment classification and the method performed equal or better than thirteen other general-purpose methods like Sentistrength, SO-CAL, and AFINN. The performance of 10SENT was consistent across different domain dataset.

To mention few, studies have also shown the ML methods perform better than the lexicons. The justification for this is that lexicon-based approach performs well because they classify based on the word present in the lexicon that can be matched to the document. This means if words in a document is not present in the lexicon, the lexicon is limited to classify accurately. While machine learning algorithms effectiveness and efficiency relies on the nature and characteristics of the data, feature representation performance and nature of learning algorithm applied (Sarker, 2021). This is evident in studies like, Krouska et al. (2017) that compared five ML methods against Sentistrength lexicon in a binary sentiment classification task. Their result showed Naïve Bayes and SVM were the top performers in the three (3) datasets obtained from Obama-McCain debate, health care reform and the Stanford Twitter sentiment gold standard dataset. Law et al. (2017) employed AFINN lexicon, neural network, logistic regression, and decision tree classifiers to detect the defect in online dishwasher review. They complemented the cross-domain SA techniques with domain terms and thus results indicate an improvement in the approach utilised. However, out of all the techniques employed neutral network was the best performing classifier. In summary, review of literature evident the mixed outcome when lexicons are compared to ML models. Literature showed lexicon methods perform across different domain due to the general knowledge acquired from the lexical resources they were built from, but supervised ML models drop in performance when used in another domain. Literature also showed ML models applied in a specific domain yields better accuracy compared to lexicons however there are studies that showed otherwise. Deep learning methods yield improved accuracy of upto 92% but requires huge amount of training set. The findings from the review of literature shows there is mixed outcome when comparing machine learning models to the lexicons. Therefore, it is necessary to review literature on these methods applied in the financial domain.

3.6.4 Domain Based Approaches

In the financial sector, SA is not just looking to classify sentences by adjectives into positive or negative sentiment. Financial analysts will look more into sentences that contributes into a negative or positive operation or product aspects. Therefore, it is important to consider tweets towards bank service or performance measures. This study agrees with Krishnamoorthy (2018) that stated SA tools accurately classify financial text better when performance measures of financial terms are considered. For example, lexicons like AFINN classified the sentence *"ATM in Abuja central is not working and not dispensing cash"* as a neutral sentence. Unfortunately, this is a negative tweet towards the bank ATM performance, and this is what the banks need to understand while considering sentiment analysis. SA techniques need to understand and detect financial terms to avoid misclassification. Early dictionaries like Harvard General Inquirer (GI) scored financial terms like tax, cost, and liability as negative. This leads to misclassification when used in a financial context while more recent dictionaries like AFINN had improved on this and scored these words neutral. Loughran & McDonald (2011) warn future researchers in the finance sector not to use any sentiment lexicon created and evaluated outside the financial business context and consequently compared generic sentiment lexicon and domain specific lexicons. They showed general purpose dictionaries like the Harvard General Inquirer (GI) perform poorly when used for sentiment prediction of financial text. This is due to misclassification of financial terms as negative. Therefore, they proposed LM lexicon applicable in the financial sector which was adopted by several studies (Das, 2014; Kearney and Liu, 2014; Gandhi et al. 2019). The study of Afolabi et al. (2019) was conducted in the banking context. They utilised SentiWordNet and domain ontology concept to classify bank tweets. Their wordlist is limited (to corpus terms from 5,934 tweets) in coverage and their approach is sub-optimal. This is because they attempted to expand social media slangs and bank words. In doing so, ignored the context of some word and thus, the approach was prone to error. For example, term like atm was expanded to at-the-moment. Such words in banking context could be automated teller machine.

Bhargava & Rao (2018) presented a comparative study of sentiment analysis (SA) techniques on different types of cryptocurrencies. Their study found out that SA is useful in making real time applications on social media data of different cryptocurrencies. They utilised naïve Bayes (NB) and support vector machine (SVM) on the different cryptocurrencies dataset, and results showed NB was more accurate than SVM. In addition, it was observed that both algorithms were able to recognise more negative tweets adequately than positive tweets. The machine learning (ML) models have been used in SA to monitor or predict stock prices (Sagala et al. 2020; Nofer & Hinz 2015; Li et al. 2014). More specifically, SVM has been shown to perform better or equal to other models for stock market prediction (Liu et al. 2020; Huang et al. 2005). However, it was observed that ML performance drops when used in financial context. For example, Xing et al. (2020) stated model performance in financial domain is less accurate compared to other domains. Their result showed model performance with F1-score of 80.6% and accuracy 71.6%. There was more false positive error than false negative. This is because their training set have more positive classes. They analysed the common mistakes and stated ML methods struggle with class imbalance problem. In summary, using the lexicon-based approach, there is need to identify the domain dependent terms to classify sentiment polarity more accurately. While previous study also identified the need to be aware of class imbalance problem when using ML approaches.

3.6.5 Language in Sentiment Classification

The most spoken and used language in Nigeria is English and Pidgin English which also applies to social media (Chiluwa, 2016; Udofot & Mbarachi, 2016). Pidgin English was developed to bridge the communication gap among citizens in Nigeria due to over 520 languages spoken in the country. Sentiment lexicons are essential resources to the application of SA. Unfortunately, existing non-English lexicons are limited. Jacobo et al. (2019) stated most lexical resources and lexicon software available are developed in English and very few are available in other languages despite the huge number of languages that exist in the world. Studies found to have analysed Nigeria language content in natural language processing (NLP) suffered from lack of lexical resources for Pidgin-English. For example, Ogbuju et al. (2020) used NRC lexicon for emotion and sentiment analysis of Nigerians during the Covid19 national lockdown. Their implementation was successful. However, was limited to English terms which was acknowledged in their study. The misclassified tweets were attributed to lack of Pidgin terms in NRC lexicon. To overcome this shortcoming, Afolabi et al. (2019) utilised SentiWordNet and domain ontology concept to classify bank Tweets. Their study expanded Pidgin, slangs, and bank terms to develop their ontology pre-processed file. However, this study considers their wordlist very limited in coverage and the approach is sub-optimal as explained earlier (in section 3.6.4). Oyewusi et al. (2020) presented an updated VADER lexicon (with 300 Nigeria Pidgin terms) using translated VADER English terms. The Pidgin tokens which include their sentiment scores were used to improve the performance of the VADER lexicon applied to Nigeria Pidgin publicly available tweets. They demonstrated improvement with their approach. Interestingly, they made the Pidgin terms publicly available which will benefit this study. However, it is worth acknowledging that the Pidgin terms were crawled from generic sport context. In addition, details were not provided to understand the performance of the updated VADER lexicon in terms of evaluation metrics such as accuracy, precision, recall or the F1 score. In summary, there is that research gap in SA in Nigeria banking context.

3.6.6 Challenges of Sentiment Classification

Studies that performed SA using social media data considered the use of this data challenging. For example, Kharde & Sonawane (2016) conducted a survey of Twitter SA techniques and identified language barrier, sarcasm, thwarted expressions, entity recognition, negation, domain dependence, handling comparison and the continuous use of novel words in social media data as major challenges faced by researchers to recommend an "off the shelf" sentiment analysis method. Similarly, Zimbra et al. (2018) outlined common challenges of Twitter SA as; frequent use of slangs, limited length of tweets which intensifies communication behaviour and promotes short novel words used among twitter users which are difficult to analyse. Hussein (2018) also conducted a survey of SA challenges and ascertained, language, sarcasm, negation, and domain dependence as the major challenges. However, lexicons like Sentistrength had been argued to handle some of these challenges well. Yue et al. (2018) conducted a survey of SA studies done on social media data and identified negation and domain dependence as major challenges. They further stated research had been done to limit these challenges through hybrid approaches proposed by different authors. However, there is room for improvement on the performance.

3.6.7 Summary of Sentiment Classification

In general, studies have proposed individual lexicon tested in different domains. These sentiment lexicons perform well across domains (general purpose lexicons), however, these lexicons are insensitive to contextual semantic orientation in domains such as financial domain. Kaity & Balakrishnan (2019) stated non-English words are challenges to sentiment analysis and unfortunately, lexicons that detect non-English words are limited. There are few lexicons that have been developed in other languages. For example, Spanish (Perez-Rosas et al. 2012), Arabic ((Abdul-Mageed et al. 2011), German (Clematide et al. 2010), Urdu (Syed et al. 2010), Japanese (Kaji and Kitsuregawa, 2007), Chinese (Xu et al. 2010), and Romanian (Banea et al. 2008). However, Pidgin English lexicon found in the context of SA are limited. Studies (Afolabi et al. 2019; Oyewusi et al. 2020; Ogbuju et al. 2020) found have utilised English lexicon as their major lexicon to classify Pidgin text which is unsuitable based on variation of Pidgin language. Foreign language such as Pidgin English is sensitive and difficult for human to understand what some statement means. This is because some of the words are malformed, they are broken English words and thus, difficult to know if they are positive or negative. This study considers the lexicons of Henry Word List (Henry, 2008) and LM wordlist (Loughran & McDonald, 2011) not suitable for this study because their wordlist does not capture social media data. Extensive review of literature shows there is no already made approach to analyse banking textual dataset and the importance of domain dependent words in SA cannot be ignored. In this context, it is more complicated because of the language (English and/or Pidgin English). Thus, this study requires a lexicon that can detect the semantic scores of banking terms in English and/or Pidgin English.

Machine learning methods struggle due to the regular change in ambiguous words used on social network. This result in regular need of training dataset. ML perform well with large, labelled data. Unfortunately, there is no labelled dataset available for training purpose in the banking domain. Yue et al. (2018) explained lexicon-based method and machine learning method have their strength and weakness however both methods might struggle with social media data of different languages because there is no available dictionary for some language such as pidgin English. In terms of which ML method is more appropriate, literature showed mixed outcome for the ML performances as these varies in different studies. Therefore, there is need to compare these techniques to validate the performing models in this context. To conclude, this study requires the development of SA techniques to classify banking tweets more accurately. The rationale to develop a lexicon that best suits this context is because no efficient lexicon was found to capture banking terms in English and/or Pidgin English and there is no training set available for ML models. Thus, the key task can be summarized as,

- Develop from scratch a banking lexicon to classify bank customer tweets more accurately (for this purpose, this study will review literature on sentiment lexicon generation techniques in section 3.8)
- Develop training set for the ML models.
- Compare machine learning models in this context.
- Evaluate the performance of the SA techniques.

3.7 Sentiment Lexicon Generation

Subjective terms are important to detect the contextual semantic polarity in SA tasks. There are different types of subjective terms that exist in a sentence. For example, the base type such as beautiful and bad and the comparative type such as better and worse. Most lexicons tend to perform well with base sentiment terms except in cases where words were used differently. The comparative words are more difficult to analyse especially in a sentence level sentiment analysis. For example, *"the network is slow using ATM but better in the banking hall"*. The sentence was classified positive in the AFINN lexicon because the word "better" is labelled positive in the lexicon. However, the sentence is negative towards ATM which is also a banking channel.

The problem of identifying factual words that mean sentiment in a context remains unresolved. For example, *"my account has been debited since last month and till now, no refund @EFCC @CBN"*. The example demonstrates customer's tweet towards the bank which expresses poor service experienced. In this case the customer tagged the economic and financial crime commission due to frustration to escalate the case. This type of sentence indicates sentiment analysis research work should consider both explicit and implicit statements. Literature on SA application focused on explicit statement (sentiment words directed to an entity) while research on implicit statement (objective sentences that indirectly demonstrates opinion) features are limited. Thus, this study will take into consideration both explicit and implicit (opinionated) statement because both statements are used in service industries. Other notable problems are dealing with; words that are domain dependent, co-occurring words with similar sentiment orientation and word semantic variance in a specific domain. For example, in a car review domain, the sentence *"the car is quiet and radio system in the car is quiet"* indicates opposing review of the car. Towards the car the sentiment is positive and towards the car radio the sentiment is negative. General purpose sentiment lexicon such as Sentistrength has shown promising result. However, due to inability to detect contextual sentiment orientation in domains, the performance varies across the domains. Thus, considered unsuitable for narrow context unless redesigned to suit that domain. Feng et al. (2018), stated customers' sentiment analysis of product depends largely on the quality of the lexicon. This agrees with the study of Zhao et al. (2014) that stated high quality lexicon yields a better and effective sentiment classification performance. Thus, this study concludes high quality lexicon is highly important and thus review approaches to generate lexicon.

3.7.1 Lexicon Generation Approaches

There are two main approaches to create sentiment lexicon which are manual, and automatic approaches. More often, studies have utilised both methods (hybrid). These approaches will be discussed into details in subsequent sections.

3.7.1.1 Manual Approach

This method involves gathering and labelling of sentiment words manually. Several lexicons have been built manually and been applied to different domains successfully. For example, MPQA (Wilson et al. 2005), LIWC (Pennebaker et al. 2001) and VADER (Gilbert 2014). Abdul- Mageed et al. (2014) manually built a sentiment lexicon and thus presented SAMAL, a supervised ML system. Manual approach can be used to improve other types of sentiment generating methods because it provides more accurate result (Kaity & BalaKrishnan, 2019). However, it is labour intensive, difficult to build, time consuming and suffers from low coverage (Wu et al. 2019).

3.7.1.2 Automatic Approach

This method involves using set of seed words to generate synonyms and autonyms through the help of online dictionaries or corpus. The synonyms and autonyms generated are therefore labelled according to their seed word. For example, Julian-Brooke et al. (2019) developed Spanish sentiment lexicon using cross linguistic automated approach. They employed the automated translation of English words to Spanish. However, this approach is prone to error. Thus, they utilised manual verification by native speaker. Wu et al. (2019) presented an automatic approach to construct target specific sentiment lexicon. In general, automatic approach can be best described by methods employed. Automatic approach can be categorised further into two approaches namely, dictionary-based approach and corpus-based approach.

3.7.1.3 Dictionary Approach

Dictionary based approach involves building lexicons using online dictionaries. Lexicons built from this approach tend to perform across all domains because they are built from general lexical knowledge. Thus, dictionary-based approach is assumed better than corpus-based approach (Kumar, 2019; Hamilton et al. 2016). The approach is beneficial because it is easy to find large numbers of sentiment words, useful to build general purpose sentiment lexicon. However, their sentiment orientation of words gathered are general i.e., domain independent. Several studies (San-Vicente et al. 2014; Gatti & Guerini, 2012; Saif et al. 2009; Peng & Park, 2011; Xu et al. 2010; Hassan & Radev 2010; Dragut et al. 2010; Rao & Ravichandran 2009; Peng & Park 2004; Valitutti et al. 2004) employed this method to generate sentiment lexicon effectively. More specifically, Kamps et al. (2004) utilised the distance-based approach to determine the semantic orientation of words from WordNet. Dragut et al. (2010) & Blair-Goldensolm et al. (2008), both studies utilised different bootstrapping method to generate sentiment words from WordNet. Their approaches were based on directed and weighted semantic graph such that the neighbouring words (synonyms and antonyms) were captured.

3.7.1.4 Corpus Approach

Corpus based approach involves generating of sentiment terms from available domain corpus. The effectiveness of corpora-based sentiment lexicon depends largely on the scale and quality of corpora employed. Medagoda et al. (2015) stated non-English lexicon receive less attention due to lack of resources. They further stated corpus-based approach appears a useful approach to build less resourced language lexicons. Past studies (Wang et al. 2017; Choi & Cardie 2009; Hatzivassiloglou & McKeown 1997; Kanayama & Nasukawa 2006) have employed this method to generate sentiment lexicon. Few studies have used more sophisticated approach. For example, Elhawary & Elfeky (2010) proposed graph-based method that clusters words based on their similarity to develop Arabic sentiment lexicon. Wang et al. (2020) proposed graph-based approach to sentiment generation to detect the domain sentiment polarity changes of words. There are other techniques such as exploring the syntactic relation between opinion and words (Volkova et al. 2013) or the use of word co-occurrences (Alemneh et al. 2020; Yang et al. 2014; Rothe et al. 2016). Some studies have employed the help of statistical methods. For example, Jhav et al. (2015) used pointwise mutual information (PMI) to label the words from their seed list developed from the corpus in context and thus developed Hindi sentiment lexicon. Twairesh et al. (2016) also used PMI to generate Arabic lexicon using corpus-based approach. Feng et al. (2018) considered the relationship between sentiment words and product features. They used PMI to calculate the relationship between sentiment words and product review. The lexicon showed an improved sentiment classification of mobile shopping review. Haniewicz et al. (2014) used random walk method to develop polish sentiment lexicon using various web documents. Remus et al. (2010) developed SentiWS from product review corpus. SentiWS determined the sentiment polarity of words using pointwise mutual information.

In comparison, Kumar (2019) criticized corpus-based approach to lexicon generation for lacking in detecting context variation of sentiment words. Feng et al. (2018) argued that the corpus-based approach to develop sentiment lexicon is better than dictionary based in terms of detecting context-based sentiment words. Corpus based approach is beneficial because it is easy to detect the syntactic relations of opinion and target. However, limited because it leads to domain specific lexicon, lacks pre-processing tools that can support other languages, encounters problem with context dependent words such as quantity adjectives. For example, in mobile phone domain, *"long battery life"* has a positive sentiment polarity. This is different to negative sentiment statement like, "*it takes a long time to focus"* in a camera domain. Thus, the approach is open to critic.

3.7.1.5 Other (Popular) Approaches

The translation generating technique is increasingly used lately for non-English lexicons. For example, Sazzed (2020) developed sentiment lexicon in Bengali language using the translating system. Studies found to have utilised this method explored existing English lexicon and used translators such as the Google translate and bi-lingual dictionary to change words from English to non-English words. Google translate API is free to use, fast, large coverage and accurate while Bi-lingual dictionary is labour intensive but also accurate. After the translating process, a native speaker or annotator verifies the translated words to check if they are correctly translated and labelled. The problem with this approach is it takes away the context value of the words and some words lose their semantic meaning which affects the performance accuracy of the lexicon especially in a strong context-based dataset. The available dictionaries have been resourceful in terms of translating and building sentiment lexicon to other non-English lexicons. For example, Abdaoui et al. (2017), translated NRC word emotion lexicon (English) to French. Hammer et al. (2013), translated AFINN lexicon (English) to Norwegian. Perez-Rosas et al. (2012) translated Opinionfinder and SentiWordNet lexicons (English) to Spanish and Basile & Nisim (2013) translated SentiWordNet & MultiWordNet (English) to Italian. Steinberger et al. (2012) employed triangulation translating method amongst three languages using English and Spanish as the source to generate lexicon. The semi-automatic approach outperformed abstract translation method used in other studies in terms of building sentiment lexicon. Kaity & Balakrishnan (2019), praised the translating method for providing a fast and easy system of building non-English lexicon however criticised the method for loss of context, unable to translate slangs, and abbreviation commonly found on social media.

Few studies have utilised both approaches (dictionary and corpus) and/or improve the approaches manually or by adopting/adjusting existing lexicons. For example, Kaity & Balakrishnan, (2019) attempted to overcome the challenges faced by sentiment generating methods by building lexicon automatically. This was aimed to be language independent to accommodate non-English vocabularies from any part of the world. Their study proposed an automatic language independent generation method for non-English lexicon. Their lexicon outperformed other existing Arabic sentiment lexicon which were built using translation methods. Xing et al. (2019), presented an approach to adapt existing sentiment lexicon such as Opinion lexicon (Hu & Liu, 2004), SentiWordnet (Esuli et al. 2010), L & M (Loughran & McDonald, 2011), SenticNet (Cambria et al. 2018) to the target domain by a sequence of trained data. The polarity score was determined by random walk using seed words like love, hate, good, bad to propagate via a constructed graph. Chen & Skiena (2014) designed sentiment lexicon for 136 languages by integrating various linguistic resources to construct language knowledge graph useful in lexicon generation.

Kumar (2019) proposed Lexscore (co-occurrence with word embedding and part of speech tagging) to improve on general domain lexicon. The approach detects context-based sentiment words to improve on lexicon-based classification across different domains. Kumar (2019) showed 10% of sentiment words changed score due to change in domain. Therefore, there is always that need to re-score words or adjust lexicon to use in a different domain for good lexicon performance. The approach combined the use of co-occurrence with word embedding and part of speech tagging to transform commonly used lexicons to domain specific. The approach combined the use of co-occurrence with word embedding to construct domain sentiment word vector representation from a corpus based and identify their semantic relation using part of speech tagging from a general-purpose lexicon. Medagoda et al. (2015) developed Sinhala language sentiment lexicon with the help of SentiWordnet. Their study mapped sentiment words in SentiWordnet to an online Sinhala dictionary. The Sinhala words were then expanded by adding their synonyms and respective sentiment score from SentiWordnet which now formed the Sinhala sentiment lexicon.

Several statistical methods exist for the measurement of seed words to their synonyms to detect their semantic orientation value. For example, Huang et al. (2014) constructed a similarity graph that automatically detect sentiment polarity of words using their contextual and morphological relations between words. Fernandez- Gavilance et al. (2018) utilised syntactic features of words to create an emoji lexicon. Mohammad et al. (2018) extended ALGA algorithm to generate lexicon and showed their lexicon outperformed AFINN and NRC. Zhao et al. (2019) developed sentiment unit context propagation framework to analyse implicit and explicit sentiment terms. Jijkoun et al. (2010) employed bootstrapping method to construct a general-purpose sentiment lexicon. Kevin-labille et al. (2017) developed sentiment lexicon using probabilities and information theoretic technologies. The information theory-based measure called TF-IDF evaluates the importance of a word in a document. Their approach developed a domain specific lexicon which they showed perform better than general purpose lexicon. Dirk-Reinel et al. (2018) employed graph-based approach to generate sentiment phrases successfully. Their approach used word co-occurrence measures and word frequency to structure sentiment word phrases into graph construct. Yazidi et al. (2015) showed the structural balance theory to generate sentiment lexicon outperforms the common approach of classical label propagation. They criticised label propagation technique that the sentiment words which are further hops away from sentiment seed words are assigned very low sentiment values. This leads to misclassification or chop in accuracy when the lexicon built is used in different domain. The inaccuracy occurs when the meaning of two synonyms is not overlapping. When such assumption is not put into consideration it causes mislabel of sentiment values. This show in cases where accuracy of the synonym-relationship is not taken properly into account, the sentiment lexicon classification is inadequate. In summary, both dictionary and corpus approaches have their strength and weakness as stated earlier. Both approaches have been widely used and combined sometimes to complement each other. For example, Jacobo et al. (2019) identified automatic method to generate wordlist and manual improvement as the major steps to build a lexicon. The approach of combining these lexicon generating methods is thus aimed for the methods to complement each other to build a reliable sentiment lexicon. This study will thus explore these approaches to generate a lexicon with optimal performance.

3.8 Aspect Extraction

In section 3.3, while discussing the concept of sentiment analysis, this study agreed with Liu (2015) that identified aspect extraction and sentiment classification as two different tasks in sentiment analysis system. This study considers the sentiment classification not enough to generate in-depth insight in the customer tweets without aspect extraction. This is because the analysis will only provide an overall Tweet-level sentiment analysis. A more functional system where these sentiment polarities can be attributed toward topics (bank aspects) will be more valuable in the banking industry or academia. This can be achieved using the aspect extraction techniques. Thus, subsequent sections will review literature on aspect extraction techniques.

3.8.1 Rationale of Aspect Extraction

In the area of sentiment analysis, past studies (Huang & Carley, 2019; Huang et al. 2019; Li et al. 2019; Munikar et al. 2019) on fine-grained sentiment classification task performed aspect-based sentiment analysis (ABSA) with the aim to improve the accuracy of the sentiment models using the extracted aspects. For example, the performance enhancement was successfully done using benchmark customer review (IDMB) dataset (Thongtan & Phienthrakul, 2019; Xie et al. 2019; Johnson & Zhang, 2016; Miyato et al. 2016). However, studies found to have applied this concept to investigate customers' sentiment polarities are limited. This is more valuable for organisations or companies to understand customers attitude towards a product or service. For example, a customer banks with Bank A does not mean the customer has a positive sentiment or attitude towards Bank X's ATM service or interest rate. Thus, it is essential for organisations to understand not just an overall opinion about their bank but also about their product and services.

Aspect extraction involves the process of extracting topics in the context of Nigeria bank product and services. There are two forms of aspects namely, explicit features and implicit features (Liu, 2015; Hu & Liu, 2004). The explicit features are the ones expressed in the form of noun or noun phrases while the implicit features are all other expressions that indicate aspect such as adjectives and adverbs. In general, there are several methods used for aspect extraction such as frequency-based approach, syntactic/dependency relations, supervised and unsupervised learning techniques. To explore the pros and cons of these techniques. This study will review prior studies on explicit and implicit aspect extraction.

3.8.2 Extraction of Explicit Features

There are several ways to extract explicit features. This can be achieved by, determining the frequency of occurrence (frequency-based approach), exploiting syntactic relations, and/or using supervised learning. The frequency-based approach is the process of counting the occurrence of nouns and noun phrases in the corpus (called term frequency - TF). This study considers frequency-based approach useful because it easily identifies popular features. Feature frequency is an important factor when ranking features to determine their importance or relevance (Liu, 2015; Kastrati et al. 2020). There are different approaches to extract features by frequency. For example, Hu & Liu (2004) applied a pre-determined frequency threshold to an association rule mining to extract highly frequent features. Ku et al. (2006), used TF-IDF (Term Frequency Inverse Document Frequency). Moghaddam & Ester (2010) applied basic frequency count to extract features and removed less frequent ones due to the assumption that infrequent nouns or noun phrases are not likely to be features or they are less important. The frequency-based approach has received praises for its effectiveness and simplicity (Hu & Liu, 2004; Liu, 2015; Kastrati et al. 2020). However, it should be noted that this process might fail to identify important features that are less mentioned.

The syntactic relation approach relies on opinion/sentiment words. The process involves using syntactic parser to identify syntactic relation between sentiment words/expressions and their target or referring noun or noun phrases. This approach had been praised for identifying infrequent features (Hai et al. 2012). Several studies have used this approach. For example, Zhuang et al. (2006) used dependency parser to identify the dependency relation of sentiment words to their target. Somasundaran & Wiebe, (2009) also extracted features by determining the opinion target pair using dependency parser obtained from Stanford parser. Qiu et al. (2011) used bootstrapping method called double propagation to extract features through dependency relations of sentiment expressions and their target. However, this approach is considered less effective with short objective sentences (because these sentences can sometimes be of negative opinion).

Supervised learning like Hidden Markov model (HMM) are used as direct sequential learning model to extract features. Like all other supervised models, this approach uses manually labelled features and non-features training set. It is worth understanding the assumptions of supervised learning like HMM is not adequate (Liu, 2015). Thus, affects the performance accuracy when used in real life problems. The general procedure involves breaking down sentences into token and apply dependency relations of different association rules depending on the model. For example: Yang & Cardie (2013) used conditional random field (CRF) to extract opinion target, opinion expression and opinion holder using the associated opinion linking relations.

3.8.3 Extraction of Implicit Features

Implicit feature extraction is more challenging and largely depends on the domain. Li et al. (2015) highlighted the importance of verb expressions in ABSA because verbs express opinions too. They stated complains or issues about a product or service are expressed using verb expression. These expressions demonstrate a product malfunctioning and thus the customer feels the service has failed or her experience is bad. Unfortunately, there are limited methods known for implicit feature extraction. The most common method is by mapping verb, adverb, and adjectives to their noun feature. For example, adjectives like expensive can be mapped to “price”, so as beautiful can be mapped to “appearance” and so on. However, mapping complex statement to noun feature remains challenging due to the random nature of implicit features expressions. For example, in a phone review dataset, customer tweeted ‘’the phone fits in my pocket very well’. The mapping of “fit in my pocket” to its noun feature “size” might be challenging. Therefore, the need for background knowledge of the domain and tuning of standard techniques is needed to extract implicit feature. To support this, Liu (2015) stated the major problem of extracting implicit feature is that it is domain dependent and needs manual inspection for improvement. This is expensive and thus suggests the reason for limited research carried out in this area. An alternative method is to detect the explicit features present in the raw dataset and afterwards, use online dictionary such as WordNet, to extract their synonyms and antonyms. Thus, those words can help find the explicit (noun) features of the implicit features.

Topic modelling methods is more applicable in aspect extraction context because it extracts both implicit and explicit aspects. A brief description of these methods is as follows; Latent

Semantic Indexing (LSI) was proposed by Deerwester et al. (1990). This topic modelling technique uses a singular value decomposition (SVD) of large term-document matrix to identify a linear subspace (in the space of term frequency inverse document frequency features) such that the relationship between the term and document are captured. The method transforms dataset into different space and thus map semantic association (structure) of terms with document. LSI assumes similar terms are used in the same context and hence tend to co-occur more. LSI captures the conceptual content and has been applied successfully in previous studies (Deshmukh et al. 2012; Kumar & Ravi 2017; Aquino & Chavez 2018; Dewangan et al. 2020). However, the technique is limited as it does not express topic with the probability. This lead Hofmann (1999) to develop the Probabilistic Latent Semantic Indexing (pLSI) for this purpose.

Blei et al. (2003) developed Latent Dirichlet Allocation (LDA) in the context of topic modelling. The statistical model was proposed from the background of improving on previous models such as LSI (Deerwester et al. 1990) and pLSI (Hofmann, 1999). pLSI received criticism that the model is incomplete as the model does not account for generative probabilistic model for the documents. Blei et al. (2003) stated LSI & pLSI leads to overfitting and unclear how the probabilities (pLSI) will be assigned when deployed or tested in a different data. LDA was thus, developed to consider the exchangeable representation for both documents and words. This aims to capture the important intra-word/document statistical structure via the mixing distribution. LDA has since then been successfully utilised for topic modelling (Jo et al. 2011; Moghaddam & Ester 2011; Lin et al. 2011; Kim et al. 2013; Hingmire et al. 2013; Panichella et al. 2013; Ghosh & Guha 2013; Barua et al. 2014; Binkley et al. 2014; Dermouche et al. 2014; Xiang & Zhou, 2014; Yang et al. 2015; Rahman & Wang 2016; Kaur et al. 2016; Beseiso 2019; Bertalan et al. 2019; Habibabadi & Haghighi 2019; Yiran & Srivastava 2019; Liu et al. 2020; Zhu et al. 2020; Kastrati et al. 2020; Qi et al. 2020). Prior studies have shown LDA is robust and efficient. For example, Albalawi et al. (2020) compared topic modelling techniques for aspect extraction. They showed LDA was the best performed model using evaluation metrics such as topic quality, precision, topic coherence, and recall. Ray et al. (2019) compared topic modelling methods using Hindi dataset and found LDA as the best TM method. Kherwa & Bansal (2020) compared topic modelling techniques and found LDA as the best topic modelling technique. However, LDA had been criticized that generated topic contain irrelevant features (Ko et al. 2017) and noisy topics from short text like social media text (Ali et al. 2019). Thus, various extensions of LDA were proposed. For example, the correlated topic model (Blei & Lafferty 2005), hierarchical Dirichlet process model (Teh et al. 2006), the hierarchical topic model (Blei et al., 2007), Constrained-LDA (Zhai et al. 2011), MaxEnt-LDA (Zhao et al. 2010), Automated knowledge LDA (Chen et al. 2014), LDADE (Agrawal et al. 2018), Ontology-LDA (Ali et al. 2019) and most especially, the popular Java-based Machine Learning for Language Toolkit-MALLET (McCallum, 2002). Hierarchical Dirichlet Process (HDP) is a non-parametric Bayesian model that recurse parametric function to cluster multiple grouped data (Teh et al. 2005; Wang et al. 2011; Chien 2017). HDP involves dealing with representation of multiple grouped data where each group is associated with mixture model from a local Dirichlet process i.e. Each local Dirichlet process governs the words generation for the group. Words from different group are represented by global mixture model from a global Dirichlet process. HDP is a Bayesian non-parametric extension of LDA where the represented groups can grow structurally as it captures the important intra-word/document statistical structure via the mixing distribution. Yan et al. (2013) proposed Bi-term Topic Model (BTM) for topic modelling specifically for social media data. BTM has been shown effective for clustering topics of social media text. BTM learns topics by modelling word pairs in the whole corpus. BTM was developed with the assumption that two words will be assigned the same topic label if they have co-occurred (Xia et al. 2015). However, it is worth mentioning that this study considers that assumption too strong for this context and thus will not consider the model for this study.

The use of language models for aspect extraction is evolving. This is because topic models such as LDA has been criticized for not able to extract topics that have low frequency of occurrence to related words. This led to the development of neural network stacked models and pre-trained language models. For example, Kumar et al. (2020) performed ABSA using Bi-LSTM and CRF with SemEval'14 Task4 dataset (3841 restaurant and 3845 laptop reviews). They showed their approach produced a good result in the restaurant domain with F1 score of 80% and 71.28% in laptop domain. Dandala et al. (2019) combined CRF and Bi-LSTM to detect medical aspects which are relevant to adverse drug event and achieved F1 score of 83%. Gandhi & Attar (2020) used CRF and Bi-LSTM to perform aspect term extraction in 5417 Hindi reviews to good effect. BERT (Bidirectional Encoder Representations from Transformers) is a deep bidirectional unsupervised language representation model developed by google (Devlin et al. 2018) and has shown good result in natural language processing task such as SQuAD question and answer test, and GLUE (General language understanding evaluation). BERT (base) is implemented based on transformer and attention mechanism which uses encoder to read in input and decoder to output. The BERT base model consists of 12 layers of transformer blocks, 768 hidden layer size, and 12 self-attention heads. The model comprises of two stages which are pre-training stage and fine-tuning stage. Yanuar & Shiramatsu (2020) used BERT for aspect extraction in an Indonesian tourism review from TripAdvisor. In their experiment, they added an Indonesian language specific pre-processing phase (conversion to lowercase, stemming, tokenization and normalisation) and thus, further pretrained BERT with 4220 review sentences (using train batch size 32, max sequence length 128, number of epoch 16 learning rate $3e^{-5}$, and Adam epsilon $3e^{-8}$). They tested their approach using 501 Indonesian amusement park reviews and showed their approach achieved an accuracy of 79.9% and F1 score of 73.8%. Bensoltane & Zaki (2022) compared variants of BERT against BiLSTM and CRF for aspect extraction using 2265 Arabic news post (retrieved from Facebook about the 2014 Gaza attack). They showed that the language model is suitable in a low-resourced setting and showed the combination of BERT, BiGRU and CRF outperformed all other models for aspect term extraction with F1 score (88%). Winatmoko et al. (2019) compared BiLSTM in the study of Fernando et al. (2019), BERT as a standalone model, and combination of BERT and CRF using 5000 Bahasa Indonesia hotel reviews for aspect detection task and thus showed the BERT models performed better especially BERT and CRF with F1 score of 92%. Xu et al. (2019) proposed a BERT post training approach for aspect extraction and aspect sentiment classification task. Their work found out that great performance boost comes mostly from domain knowledge post training, which indicates that contextualized representations of domain knowledge are very important for aspect extraction. Sun et al. (2019) fined-tuned the pre-trained language model BERT for Sem-Eval Task 4 data to good effect. However, Liu et al. (2019) stated BERT was undertrained and thus, proposed RoBERTa, a robustly optimized BERT pretraining approach for topic extraction. Their motivation is to help determine what aspect contribute the most. In their experimentation, they established that BERT is competitive with other language models such as XLNet, however, can be improved. They showed RoBERTa performed better than BERT on GLUE and SQuAD. Zhu et al. (2021) in their study, inserted topic layer into fine-tuned RoBERTa and thus, proposed topic augmented language model for topic extraction. XLNet (Yang et al. 2019), was also proposed to improve on BERT that suffers from the pretrain-finetune discrepancy. XLNet is an autoregressive model has been successfully applied to extract aspects in various domain. For example, Tao & Fang (2020) successfully applied XLNet for multi-label aspect prediction task in restaurant reviews (from Yelp), Wine reviews (from Winemag), and movie reviews (from Rottentomaties). Unfortunately, there is no literature found to have applied or compared these techniques in the Nigeria banking context.

3.8.4 Summary of Aspect Extraction

In summary, the review of topic and language models showed that prior studies have utilised them effectively. However, studies found to have compared these methods are limited. This motivates this study to consider a comprehensive comparison of these techniques. The frequency-based approach is considered simple and efficient. The technique outputs popular and important features in the dataset. However, does not detect some important features which are less mentioned. Studies have employed seed words or some association rules to detect the less-mentioned features but limited to domain. Supervised models have also been employed. However, labelled data for this use is unavailable. These problems identified, is the rationale to consider unsupervised models. This technique is preferred because statistical topic modelling has the advantage to extract both implicit and explicit aspects. However, topic modelling on social media data is challenging. This is so because the concept of unsupervised models is to cluster words. Unfortunately, social media data contains; frequently used malformed words, very short text, and are limited in content. Thus, this study contributes new strands of research in terms of aspect extraction in banking context.

3.9 Chapter Summary & Literature Gap

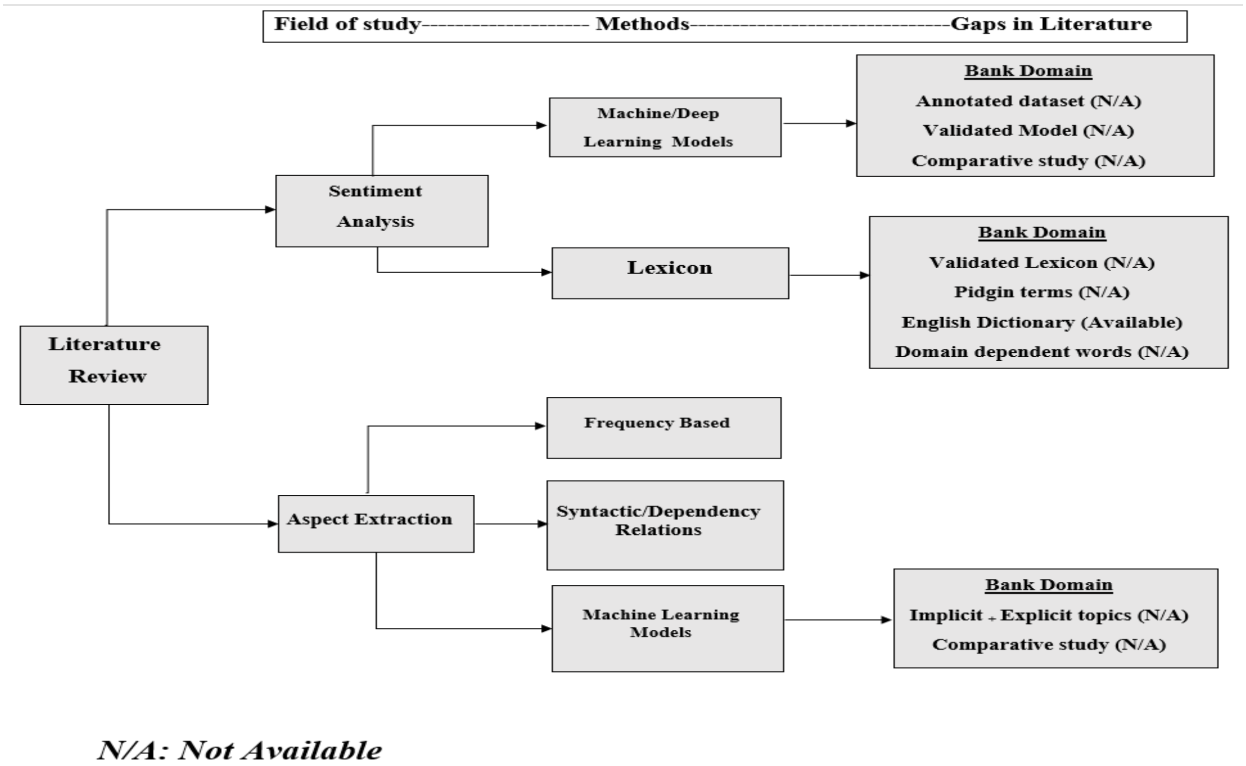


Figure 3.1: Gaps in the Literature

The chapter provided an extensive literature review of sentiment classification and aspect extraction for sentiment analysis. As regards sentiment classification, literature suggests there are two main approaches applied to sentiment classification namely, machine learning and lexicon-based approach. Unfortunately, as discussed in section 3.5 to 3.7 both machine learning and lexicon-based approach have been understudied in the banking context. As shown in figure 3.1 above, there is no lexicon or machine learning model validated in the bank domain. This implies that the best approach in terms of performance is yet to be known. Literature was reviewed as regards sentiment lexicon generation in section 3.7. This was done to provide background knowledge of approaches to lexicon generation.

Like sentiment classification, several studies as discussed in section 3.8.2 to 3.8.3 have utilised different techniques (such as frequency-based approach, association rule, part of speech tagging, topic models and language models) for aspect extraction. Unfortunately, there is no study found as shown in figure 3.1 above that extracted aspects in the Nigeria banking context. Aspect extraction in service industry such as banking is difficult due to their continuously changing product and services. This is more difficult with social media data. This is due to the text data sparsity, informal language, and very short text length which limits content and contextual cues (Asghari et al. 2020; Kharde & Sonawane, 2016). In addition, ABSA studies that employed aspect extraction techniques, focused their attention on extracting explicit features. Studies that considered implicit features are limited (Liu, 2015). This area of research is still growing, and literature reviewed suggest comparative study for aspect extraction is limited. The research gap thus, motivates this study to consider a comprehensive comparison of the statistical topic models in this context and thus to investigate customers' attitude using these aspects. This method is preferred because statistical topic models have the advantage to extract both implicit and explicit aspects. This will help assess customers' attitude towards the sentiment polarities and thus uncover hidden patterns. Thus, key tasks to fill the research gaps identified will be summarised as follows.

- Develop from scratch a banking lexicon to classify bank customer tweets more accurately.
- Develop training set for the machine learning models.
- Compare machine learning models in this context and evaluate their performance.
- Compare topic models techniques in this context and evaluate their performance.

Thus, the next chapter present the methodology for this study to achieve these tasks identified.

CHAPTER 4: Methodology

4.0 Introduction

In the previous chapter, inconsistent performance was reported on the performance of machine learning models and lexicon-based approaches to sentiment analysis. Literature reviewed showed there is no sentiment analysis (SA) framework available or validated in the banking domain. Based on this, this chapter will focus on the development of sentiment analysis approaches since it is yet unknown how these approaches will perform in this context.

Beforehand, this chapter introduces the research strategies and methods employed in this study. This study follows a quantitative and deductive research strategy which will be detailed in subsequent sessions. The chapter will provide a comprehensive overview of statistical and machine learning models applied in this domain to fill the research gap identified and to achieve the research objective. The rationale of chosen models is subject to their performance in text classification task especially in natural language processing. Table 4.0 below shows the research question, objective, and gap that will guide this chapter.

Research Question	Can machine learning models perform as an off-the-shelf sentiment classification method in the banking context?
Research Objective	To conduct a benchmark comparison of sentiment classification and topic models at Tweet level.
Research Gap	1). There is no sentiment analysis (SA) framework validated in the banking domain. 2). There was no labelled (bank) data found for training purpose.

Table 4.0: Research guide for this chapter

4.1 Research Strategy & Method

Social media provides the opportunity for users to freely express objective and subjective statement. Classification models can be applied to the data to predict user behaviour. To understand the philosophical and technical rules that underpin this. The study explores the details of the quantitative versus qualitative research design and deductive versus inductive research design.

4.1.1 Quantitative Versus Qualitative

Quantitative research design is primarily based on philosophical tradition associated with positivism. Positivism is focused on uncovering truth and the empirical presentation. This method involves using statistical procedures to obtain quantifiable measures of variables and inferences from sample of a population to produce a valid generalisable result (Queiros et al. 2017). The process starts from theory to data, narrowed to generalisable result. The research design is popular in the research community due to its statistically reliable outcome or generalisability. However, has been criticized to lack in depth to examine human and their behaviour (Amaturo & Punziano, 2017).

Qualitative research design is based on philosophical tradition associated with post-positivism. The method involves the understanding of the meaning of social phenomenon (Queiros et al. 2017). This type of research relies on data obtained through observations, interview, focus group, audio, and video recordings. No such case in this design uses statistical procedures of testing of hypothesis. The approach is effective to fully understand customers' experience and behaviour but subject to researcher's bias and lacks generalisation (Amaturo & Punziano, 2017).

Deductive research approach relies primarily on testing a priori hypothesis. This can be considered a 'top-down' method, that is, theory and hypothesis dependent. While Inductive research approach emphasizes on observation (data driven) such that conclusion can be made (exploratory). This can be considered a 'bottom-up' method and does not need pre-defined framework (Woo et al. 2017). Research methods are changing lately specifically in the field of natural language processing as data available is growing exponentially. The need for research approaches to complement each other is increasing since methods have their strengths and weaknesses. This agrees with Hesse-Biber & Johnson, (2013) that stated studies integrate both research methods as an investigation toolkit. Thus, broadens and allow emergence of independent results. Amaturo & Punziano (2017) emphasized on the strengths of mixed method in text classification problems. They outlined the benefits of using qualitative approach in capturing the essence of a phenomenon and the quantitative approach for prediction. To conclude, this study follows mainly quantitative and deductive research designs. However, employed qualitative approach for validation. This will enrich the study in having an in-depth understanding of the Twitter content and a valid generalisable outcome.

4.1.2 Proposed TSBS Framework for Sentiment Analysis

To fill the research gaps identified in chapter 3 and provide efficient sentiment analysis system, this study proposes a Topic-Sentiment Banking System (TSBS) shown in figure 4.1 below for sentiment analysis. The main tasks identified in this framework are sentiment classification and aspect extraction. As discussed in section 3.3 this study agrees with Liu (2015) that identified these two tasks to develop sentiment classification system. The sentiment classification task involves classification of subjective tweets into positive or negative and objective tweets into neutral. While the aspect extraction involves the process of extracting topics in the context of Nigeria bank product and services. Since it is yet unknown the best approach, this study will therefore employ the lexicon and machine learning approach for sentiment classification task and topic modelling for aspect extraction task. The rationale for chosen methods was discussed in the previous chapter and details of the methods will be produced in subsequent chapters. The proposed TSBS framework is made of three components namely, lexicon-based approach, machine learning approach for the sentiment classification task and topic modelling for the aspect extraction task. An explanation of the approaches to each component and justification for the selection of applicable method will be discussed rigorously at each stage. This will guide the study to comprehend the capability, limitation, and weakness of the methods and help in delivering a well performed sentiment classification system. In addition, the framework comprises of two software components, which are, the database system and programming language. The database system utilised is MongoDB. This is because MongoDB is schemaless and handles unstructured data well. In this case, MongoDB was used to store and query the Twitter data. The programming language adopted is Python. This is because Python is a highly sophisticated programming language that provides access to easy to write queries for data analytics. Specifically, Python provides access to build all benchmark models needed. In general, the framework benefitted from MongoDB and Python because both systems can be integrated using the PyMongo package. Thus, provides an easy to manage architecture for deployment.

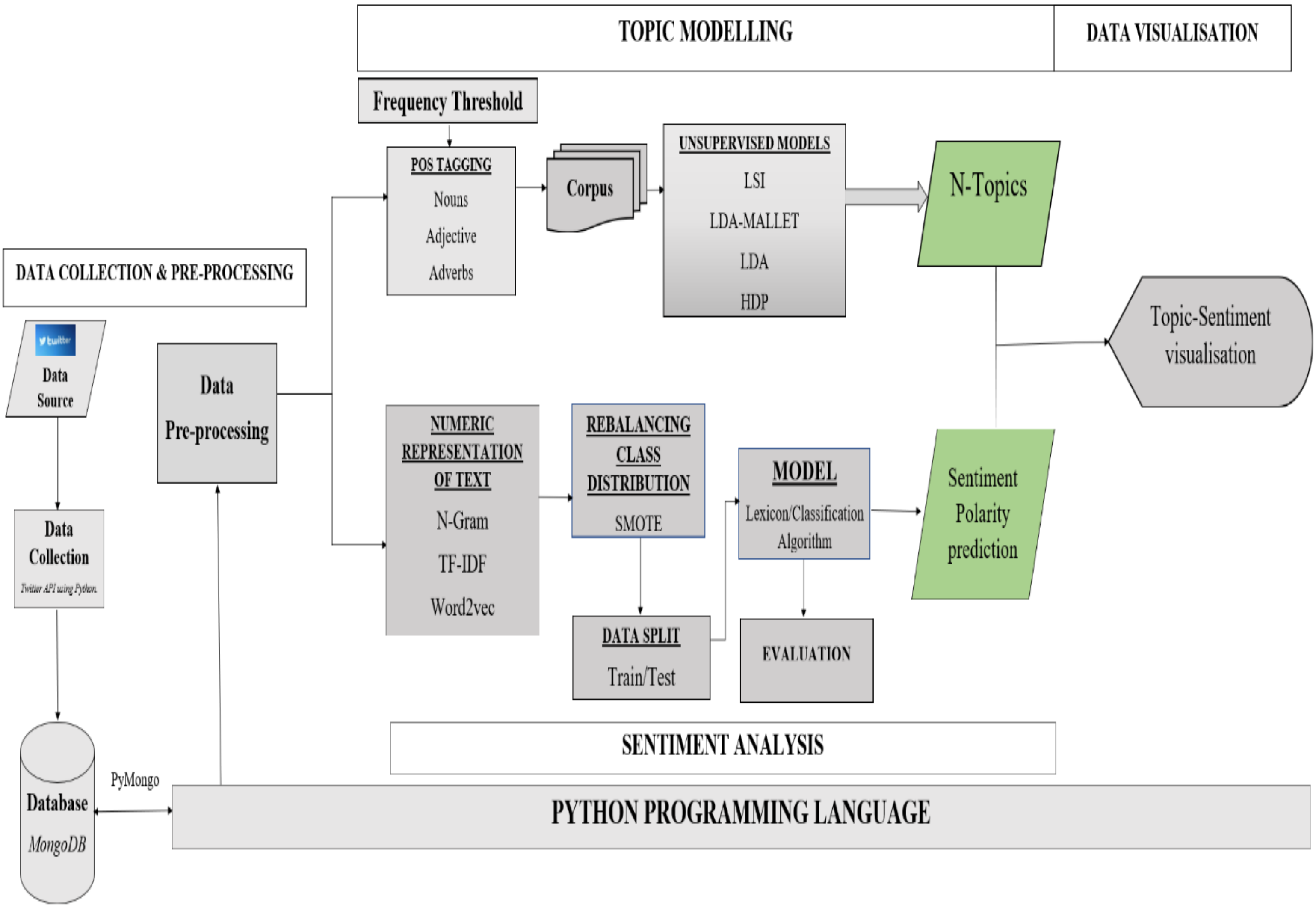


Figure 4.1: Proposed TSBS framework

In subsequent sections, the three components (lexicon-based approach, machine learning approach for the sentiment classification task and topic modelling for the aspect extraction task) identified will thus be discussed.

4.2 Lexicon Based Approach

Sentiment lexicons are essential resources to the application of sentiment analysis (SA). Unfortunately, existing non-English lexicons are limited. Jacobo et al. (2019) stated most lexical resources and software available are developed in English. However, only a very few are available in other languages despite the huge number of languages that exist in the world. For example, lexicons in Spanish (Julian-Brooke et al. 2019), and Chinese (Wan, 2008; Yang et al. 2020). A major limitation in this field, is that African countries that speak 32% of the total living languages in the world lacks lexical resources to build lexicon in their native language (Kaity & Balakrishnan, 2019). The limited non-English lexicons thus limits research in the field of SA applied to non-English text.

The percentages of internet users that communicate with English are lower than 27% (Internet World Stats, 2020). Based on Ethnologue (2021), there exists about 7139 languages in the world and out of these languages 62% are spoken in Africa and Asia. More specifically, Nigeria has over 500 different languages. Nigeria rank as the third in countries with most languages in the world. Statistics showed that Africa has the highest numbers of languages (32%) in the world. However, lexicon in their native language is limited or does not exist. There is no sentiment lexicon available for Pidgin English. Pidgin English is an unofficial language widely used across West Africa countries like Nigeria, Ghana, Cameroun, Equatorial Guinea, and Sierra Leone. In Nigeria, Pidgin-English is the second most used language across all tribes after English. The growth in usage of Pidgin English, motivates the need for organisations to render services using this language. For example, Google launched Nigeria Pidgin search interface in 2011. BBC also launched BBC News Pidgin as a news service in 2017. Since there is no validated lexicon in the Nigeria banking context. This next section focuses on approaches to sentiment lexicon generation starting from the data collection stage and thus proposes “SentiLeye” a novel lexicon algorithm.

4.2.1 Data Collection Stage

The growing usage of social media (as discussed in chapter three) by bank customers motivate this study to utilise the data to understand customers’ attitude towards their bank. This study considers social media data appropriate because the networks are increasing in coverage and usage. The data source provides an up-to-date information which are expressed freely. Twitter amongst other social networks (like Facebook) was deemed adequate because it is more open. Unlike, Facebook that follows a strict oriented policy to protect users' privacy which appears to be a limitation (in the research community) because data available depends on the level of authorization granted by the user. Twitter is free to extract tweets through their API. An application programming interface (API) is a component of object-oriented programming (OOP) language which developers can build software for an application using referenced program (Boillot, 2012). Twitter is a popular information resource that has been used in the development of different applications both in the academia and industry. Twitter has more than 330 million monthly active users, who can post messages of 280-character per tweet (Statista, 2021; Ayora et al. 2021). For the purpose of this study, a Twitter account was created, and the account was used to access the developer's portal. An application was created for this purpose which was approved by Twitter as shown in appendix A. The necessary keys like consumer key and token, which are needed to access, listen and crawl required tweets were generated. Thus, live tweets of Nigeria commercial bank customers toward their respective banks were extracted through the Twitter application programming interface (API) using Python programming language. The textual data are user generated content of customers' attitude to the banks. The criteria for tweets crawled were tweets towards Nigeria commercial banks' handle. The handles utilised can be found in appendix H. The most important information is the 'user-text'. However, the crawled data contains other details like time and location which were not considered in this study. The file is very large and therefore needs a well performing database architecture to avoid loss of data and most importantly, allow easy access, fast and effective queries.

A total of three hundred and forty-six thousand (346,000) relevant tweets were collected within three months. The criteria for tweets crawled were text towards commercial banks' handle in Nigeria. This is the first stage of data collection for this study, more details of the data collection process and statistics are provided in session 4.3. For the lexicon development purpose, tweets were collected from 22 May 2019 to 25 August 2019. The tweets were saved in JSON file format and imported to MongoDB (NoSQL). MongoDB is a non-relational database built to retrieve, store and query textual data speedily. MongoDB was used to store and query the data because it supports big unstructured textual data and are schema-less. Data were sorted in the database and only the field 'text' was used in this study.

4.2.2 Research Ethics

This study went through appropriate research ethics processes to get approval for the use of the Twitter data for the study. The approvals were from Sheffield Hallam University and Twitter. Kindly see approval letters in appendix A. This study ensured the tweets were anonymised to ensure the research is ethically complaint.

4.2.3 Data Pre-Processing Stage I

The text (tweet) data were cleaned and pre-processed in Python. Python was adopted due to its prevalence in data science. The programming language is syntactically simple, simple to learn, and straightforward. Additionally, Python is open source, free to use and has rich ecosystem of libraries for scientific computing (Lasser et al. 2021; Robinson, 2017; Koulouri et al. 2014; Ateeq et al. 2014; Jayal et al. 2011; Perez et al. 2010; Lutz, 2001). The data pre-processing steps taken can be summarized as follows:

- Lowercase: In text classification, most especially the natural language processing. Words are converted into lowercase to make matching of words or processes much easier. In this study, the text data were all converted into lowercase.
- Punctuation, Numbers, URL links: Punctuation, numbers and URL links tend to provide extra noise in text data. Thus, removed to make text easy to process. In Nigeria language context, the number 419 means 'scam'. Thus, the text data was looped and changed to that effect before removal of numbers.
- Stopwords: these are words that add little or no significant meaning to sentences or context in general when considering sentiment analysis (Dey et al. 2020). Thus, are deemed to be removed. Such words are "a", "the", "are" and "it".
- Irrelevant words like the special characters and symbols are non-alphanumeric characters found especially in social media text. These types of word or character are considered noisy and irrelevant to the sentiment in the context and are thus removed (Fithriasari et al. 2020).
- Stemming: this process involves converting all forms of word back to their root word. The pre-processing technique has been shown to have significant improvement on classification accuracy (Symeonidis et al. 2018).

The cleaned text data were deemed ready for processing after going through the steps stated above. Table 4.1 below shows example of input text against cleaned text.

Tweets	Cleaned Tweet
That is customer service. You dey force people to run around in circles because you and your staff are clueless	customer service dey force people run around circles staff clueless
Address the team in charge or your faulty machines. How can you be sending same message multiple times?	address team charge faulty machines send message multiple times

Table 4.1: Illustration of tweet versus cleaned tweet

4.2.4 Implementation of Lexicon Generation Method

There are three main steps taken to develop the lexicon. A comprehensive detail of the steps is provided below.

- The first task is to create corpus from the text (Tweets) data. To achieve this, corpus-based approach was utilised. The technique was chosen so that the context-based words in English and Pidgin English can be retrieved. In addition to the text data crawled, this study benefited from the study of Ogueji & Ahia (2019) that crawled Pidgin sentences and words from the BBC Pidgin English news website. The BBC corpus is a useful resource which captures the variations of Pidgin in the news-related context. Ogueji & Ahia (2019) obtained a corpus of 56695 sentences and 32925 words. The corpus is relevant and useful as sentiment Pidgin words were manually retrieved for this study. The corpus was constructed and manually checked for errors. Thereafter, the English terms were tagged with their respective part of speech (POS). The part of speech tagging was employed to identify the grammatical group of words. The natural language toolkit (NLTK) library (Bird et al. 2009) was used to achieve this purpose in Python. The NLTK was used because the library has extensive modules for text analytics which has been praised for efficiency and effectiveness.
- The second task involves creating labelled seed words. The adjectives, verbs and adverbs were extracted to form the seed words. Adjectives as recognised in the study of Hu and Liu, (2004) indicate the sentiment expressed in a document or sentence. The adverbs and verbs were extracted to help identify the factual words that express sentiment in objective sentences. However, to be more accurate manual inspection was conducted to ascertain the words that express sentiment. This approach was adopted from previous studies like Kim and Hovy's (2006) that expanded wordlist using three seed set of positive, negative, and neutral words to build their corpus. However, their expanded word list comprises of many errors. This study learned from their study and developed more seed list from the banking corpus and employed manual inspection to reduce error. A total of one thousand, one hundred seed words were generated. Thereafter, the seed words were labelled. The finite value within a range of strongly positive (+2) to strongly negative (-2). The most important part of the labelling was that the tone and semantic orientation of the terms were taking into consideration such that strong positive terms are labelled +2 to strong negative labelled -2.
- The third task is to expand the seed wordlist. This study employed the dictionary-based approach to expand the seed wordlist generated from the previous step. The seed words were expanded by retrieving their synonyms and antonyms from an online dictionary (WordNet). WordNet is a popular and rich online English lexical resource

developed by Miller (1995). Thereafter, the synonyms were labelled according to their root word in the corpus while the antonyms were labelled as opposite.

In summary, the Pidgin English words were generated using the corpus-based approach. The English words were generated using both corpus (tweets) and dictionary (WordNet) based approaches. The corpus-based approach was used to extract the bank domain dependent words, and the dictionary-based approach was used to increase the coverage of words. To make the lexicon suitable with social media text, this study adopted the emoticons, and slangs wordlist from the Sentistrength (Thelwall et al. 2010) lexicon due to the performance of the lexicon with social media data. In addition, intensifiers (adverbs) were added from the corpus and negating words were added manually. Finally, a manual inspection was conducted to remove any error encountered during the process. This was an intense process that involved three main annotators who are native speakers of Pidgin and English. The manual annotation and validation were chosen because it has been shown to produce the best result compared to other methods (Boukes, 2020). To have a clear understanding of the processes, this study presents a flowchart (figure 4.2) below to visualise the steps taken.

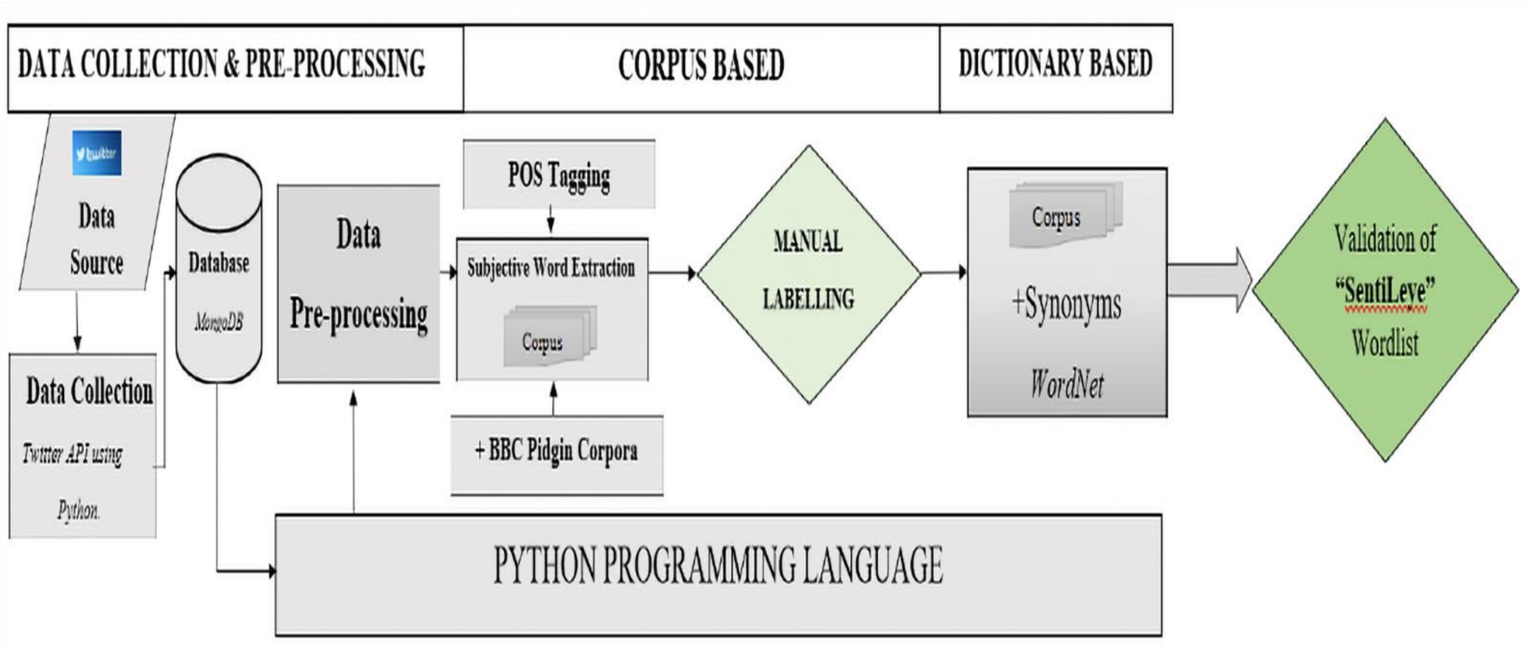


Figure 4.2: Flowchart of SentiLeye Wordlist generation

Figure 4.2 above shows the process adopted to generate the wordlist. These steps can therefore be summarised as follow:

1. Creation of seed words from text data, enhanced with POS tagging where each word was annotated with the part of the speech associated with it.
2. Labelling of the seed words.
3. Expansion of seed words using online dictionary for synonyms and antonyms.
4. Labelling of new words based on seed-word labels.
5. The iteration processes (3 & 4) continued till no more word was found.
6. Manual inspection of the words against their labels to remove errors.
7. Evaluation of the labelled words was conducted by three native speakers.

4.2.5 The Lexicon Algorithm

The lexicon algorithm utilised libraries like Csv, Ast, Re and Pandas (McKinney, 2010) to implement a basic approach to compute the sentiment scores. This basic approach can be summarised as follow:

- Identify words in the sentence/document present in the wordlist.
- The effect of negating word is to simply reverse the output of the functioning words (Xing et al. 2019). In this study, the previous or next two words (using index) will be affected. The consideration of previous words is to detect negation word that comes after the function words. For example, *"wahala no dey finish ooo see queue for XXX bank"*.
- The booster words are to increase the value of the functioning word either negatively or positively. The next two words (using index) will be affected.
- Sum the values of the words detected to identify if the entire sentence or document is positive or negative.

To initiate the process, the flowchart and pseudocode presented below were used to guide the development of the lexicon algorithm.

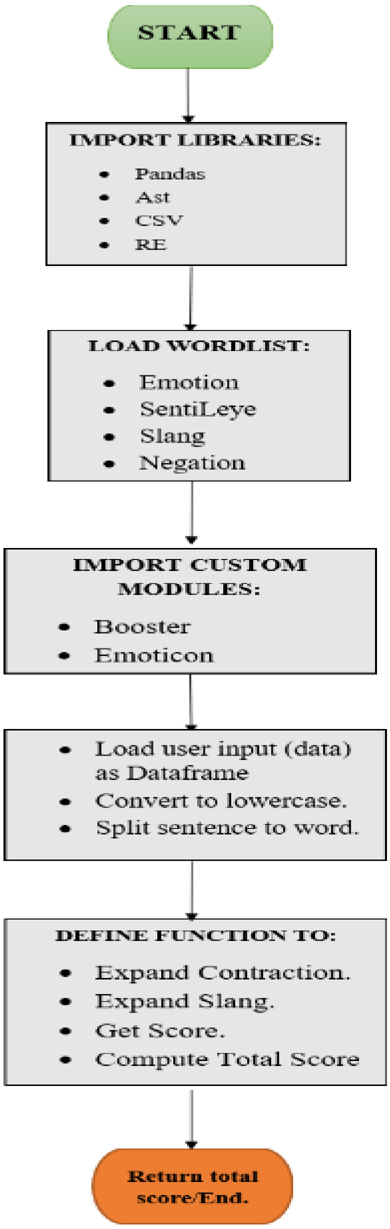


Figure 4.3: Flowchart of SentiLeye system

Figure 4.3 above presents the flowchart of the implementation which summarises the process of the system. The system utilised wordlist generated, custom module and developed simple functions to compute the sentiment class. The functions defined are expand contraction, expand slang, get score, and compute total score. The expand contraction function is to expand negation words such as “cant” to “can not” so that the algorithm can deal with “not” appropriately. The expand slang function is to expand slangs and informal words used frequently on social media. For example, “tx” which means thanks. The get score function is to get the scores of the words. The wordlist was loaded as a dictionary data structure. Thus, the words are the keys, and their label is the value. The compute total score function is to sum up the values as identified per sentence/document. These are detailed in the pseudocode shown in figure 4.4 below.

```
PSEUDOCODE
Import Pandas, csv, re, ast
Load SentiLeye, emotion, slang wordlist
Open negation list in read mode
Get text file from user and read in as dataframe.
Function expand_slang
        Parameter passed = text file, slang file
        Replace every word in text file, found in slang list by slang label.
        Return text file.

Function expand_contraction
        Parameter passed = text file, negation file
        Replace every word in text file, found in negation list by negation label.
        Return text file.

Function get_score
        The function accepts the text file using parameter denoted as s.
        Load custom modules booster and emoticon.
        Initiate variables (result, score, negation) to start from zero.
        Declare variable and pass the negation wordlist.
        Declare variable for the emotion pattern.
        Convert to lowercase.
        Split text file to words
        For each word in each line
                If word is found in emotion pattern, then pass the label into score.
                Or word is found in SentiLeye wordlist, then pass the label into score.
                Or word is found in emoticon, then pass the label into score.
                Otherwise return zero.
                If negation word is found, then multiply the next 2 scores (index) by -1
                Or negation word is found and score not in the next 2 scores (index), then
                multiply previous score by -1
                Or booster word is found, then multiply the next score by 2.

        Return result = sum of score

Function compute_total_score
        Create new column for sentiment score.
        If word in result is greater than zero, output positive.
        Or word in result is lesser than zero, output negative.
        Otherwise, output neutral.

Write file (DataFrame) to current working directory to see sentiment analysis result.
```

Figure 4.4: Pseudocode for SentiLeye Algorithm

4.3 Machine Learning Approach

The second component of the TSBS framework is the machine learning approach to sentiment classification task. Thus, this section aims to discuss the machine learning approaches employed in this study. However, before this can be achieved, it is useful to discuss the data pre-processing, data exploration, class imbalance and feature representation of text technique in subsequent sections. The processes will help feed data to the machine learning models since text cannot be entered directly into the models. The Twitter data collection continued for a duration of nine (9) months to have reliable sample representation for generalisation. A total of nine hundred and fifty-nine thousand (959,000) relevant tweets were collected from May 12th, 2019, to February 2020. The tweets were saved in JSON file format and imported to MongoDB (NoSQL). MongoDB is a non-relational database built to retrieve, store and query textual data speedily. MongoDB was used to store and query the data because it supports big unstructured textual data and are schema-less. Data were sorted in the database and only the field 'text' was used in this study.

4.3.1 Data Pre-Processing Stage II

The text (tweet) data were cleaned and pre-processed in Python. This study considers data pre-processing highly important due to the unstructured and noisy nature of the research data. The data is considered noisy because it contains URL, repeated words, retweets, slang, and hashtags (Curiskis et al. 2020). In any natural language application, pre-processing stage is sensitive because application or systems relies directly on text components obtained (input) from this stage. The text components such as words or phrases are fed into the system that performs complex analysis like identifying learning patterns or semantics to detect information. Data pre-processing contributes to the performance of the analytical system or model. This study utilises Twitter dataset, which are informal short text that contains slangs, abbreviations, emoticons, and irrelevant words. Thus, needs thorough pre-processing to improve the quality of the data. The following steps below describe the steps taken to clean and manipulate the dataset:

- Case Conversion: In natural language processing, words are converted into a single type of letter case to make processes much easier and faster. In this study, words are converted into lowercase to make matching smooth. For example, the word 'Scam' is converted to 'scam'.
- Removal of Punctuation, Numbers, and URL links: Since Twitter data were crawled online, the text data contains html <iframe> tags and URL links. In general, real-world text data contains punctuation and numbers. All these tend to provide extra noise in text data and are thus removed to make text easy to process. However, in Nigeria language context, the number 419 means 'scam'. Thus, the text data was looped and changed to that effect.
- Removal of Stopwords: From the non-linguistic point of view, stopwords are words that do not pass information. These words add little or no significant meaning to

sentences or context in general when considering sentiment analysis (Dey et al. 2020). Thus, are deemed to be removed. Such words are "a", "the", "are" and "it".

- Removal of Irrelevant & unwanted words: words like special characters and symbols are non-alphanumeric characters found specifically in social media text. These types of word or character are considered noisy and irrelevant to the sentiment in this context and are thus removed (Fithriasari et al. 2020).
- Tokenization: Tokens are individual text component with definite syntax and semantics. The process of breaking down text document (such as articles, electronic text, or any form of textual data) into tokens (words, phrases, or clauses) is called tokenization. There are two forms of tokenization which are sentence-tokenization and word-tokenization. The former is applicable when breaking down articles into sentences while the latter is commonly used in building corpus, bag of words (BOW) or lexicon. An example of word-tokenization is the sentence *"ATM machine in SHU-hubs is not working well"*. This can be broken down into tokens as; *"ATM", "machine", "in", "SHU-hubs", "is", "not", "working", "well".* In this case, word-tokenization is applied to break down tweets into tokens.
- Lemmatization: Is the process of transforming words back to their standard words by considering the vocabulary. This pre-processing technique has been shown to have significant improvement on text classification accuracy (Symeonidis et al. 2018). The main aim is to remove inflectional endings in word however considers the part of speech of the words. Words like good, better, and best results to good when lemmatized. In natural language processing, one of the issues that makes analytics more difficult is the text incorrectly spelt which often occurs. One way to overcome this challenge is through lemmatization. For example: the word 'fraudd' can be lemmatized to 'fraud'.
- Stemming: Is the process of transforming words back into their root (Liau & Tan, 2014). Word stemming is like lemmatization but differs with how they transform words into their root. For example: words like "Leaves" can be stemmed to "leave" and, "leafs" to "leaf". Both methods (stemming and lemmatization) are good but differ in result. It will be more appropriate to explore both techniques and chose the better one at different stages of this analysis.

4.3.2 Labelled Dataset

Supervised learning models have performed reasonably well in the field of sentiment analysis (Ghosh et al. 2020; Milu et al. 2020). However, these models are yet to be applied or validated in this domain. A thorough search for labelled bank dataset was unsuccessful because was no labelled dataset available. This unavailability of annotated dataset might have contributed significantly to the lack of SA research in this domain because the models rely on labelled dataset to train and learn for prediction. Therefore, this study selected 7037 tweets randomly from the research data. The tweets were labelled by two human annotators who are native speakers of English and Pidgin English. Due to the presence of Nigeria local language in the text data, this study was against using online annotators (for example, Amazon Mechanical Turk) to have a reliable and accurate class label. At the validation stage, a third native speaker was employed in cases where there was disagreement on class labels. This annotator was used to decide on the label in such scenario. The annotated bank dataset was used to train supervised models in this study. The labelled set is made available for research use (please see links in appendix C). The annotation was done at tweet level and examples of the labelled set is provided in Table 4.2 below.

Tweets	Label
bank na sure bank	Positive
disappointed called contact centre minute no answer call	negative
need get mobile app back please guy help without requesting card number	Neutral

Table 4.2: Sample of labelled set

The labelled dataset will help future work when proposing or enhancing performance of sentiment classification models in the banking context.

The Contribution of this chapter:

- The study provides a standard labelled Twitter data useful to build and improve SA model performance in the banking context.

4.3.3 Data Exploration

Exploratory data analytics (EDA) is commonly used in the field of statistics to understand the main characteristics of data. In this study, this approach was employed to familiarise and generate insights from the bank customer tweets. The text data was explored to check for missing values and produce wordcloud to understand the frequently used words in the tweets. The data exploration helps unveil hidden pattern and key terms customers tweet about. It is useful to identify characters or words with high frequency but has no significance contribution to avoid skewed analytics. Thus, these words can be included in the stopwords list to be removed. In cases of missing value, the row will be dropped since this study is only considering the tweets.

During the pre-processing stage, findings showed that bank customers tweeted a lot using numbers. An explanation for this is that customers used numbers to demonstrate amount of transaction and/or in the case of sentiment. For example, "*@ABC bank, I was debited 5,000naira yesterday and cash was not dispensed"*. In such instance, the number does not influence the semantic meaning of the sentence and thus can be removed. Another example, "*@customer A, that ur bank na just 419 @ABC bank"*. In this case the number in the sentence demonstrate sentiment because 419 in Nigeria context means scam or fraud (Isacenkova et al. 2014; Whitty 2020). The sentence simple implies @customer A, that your bank na just scam @ABC bank. Therefore, before removing numbers. The 419 were converted into scam.

In the case of punctuation, there are several common words that are used to negate in a sentence and these words come with an apostrophe. For example, words like *"haven't"*. If the apostrophe in the word is removed during cleaning. This can change the semantic meaning of the sentence. For instance, tweets like *'I haven't received the refund'* and *'I have received the refund'* means different things. This is where data pre-processing has huge influence in language processing. In this case, *"haven't"* is expanded to *'have not'*. Otherwise, applying punctuation removal will change the word to *'havent'* and stemming or lemmatization process will further change to *have*. To solve this problem, a list of English words (in Appendix) was created with apostrophe and expanded before removal of the punctuation process. The study utilised a cleaned dataset after passing through the pre-processing stages stated in previous chapter (lower-case conversion, removal of punctuation, numbers, unwanted words, and URL links).

Data exploration techniques such as bar plot and wordcloud were used to familiarise and understand the characteristics of the data. The initial plots (in appendix) showed words that were frequently used in the tweets. However, it was observed the tweets still contained unwanted words like the bank handles, malformed words, and slangs like asap. These words can skew the analysis. Thus, the bank handle and malformed words were removed, while the slangs were expanded. This study considered the use of wordcloud significant to help the system identify unwanted words which were frequently used in the tweets. The bar plot and wordcloud of the cleaned data are produced below to provide a visual understanding of the data.

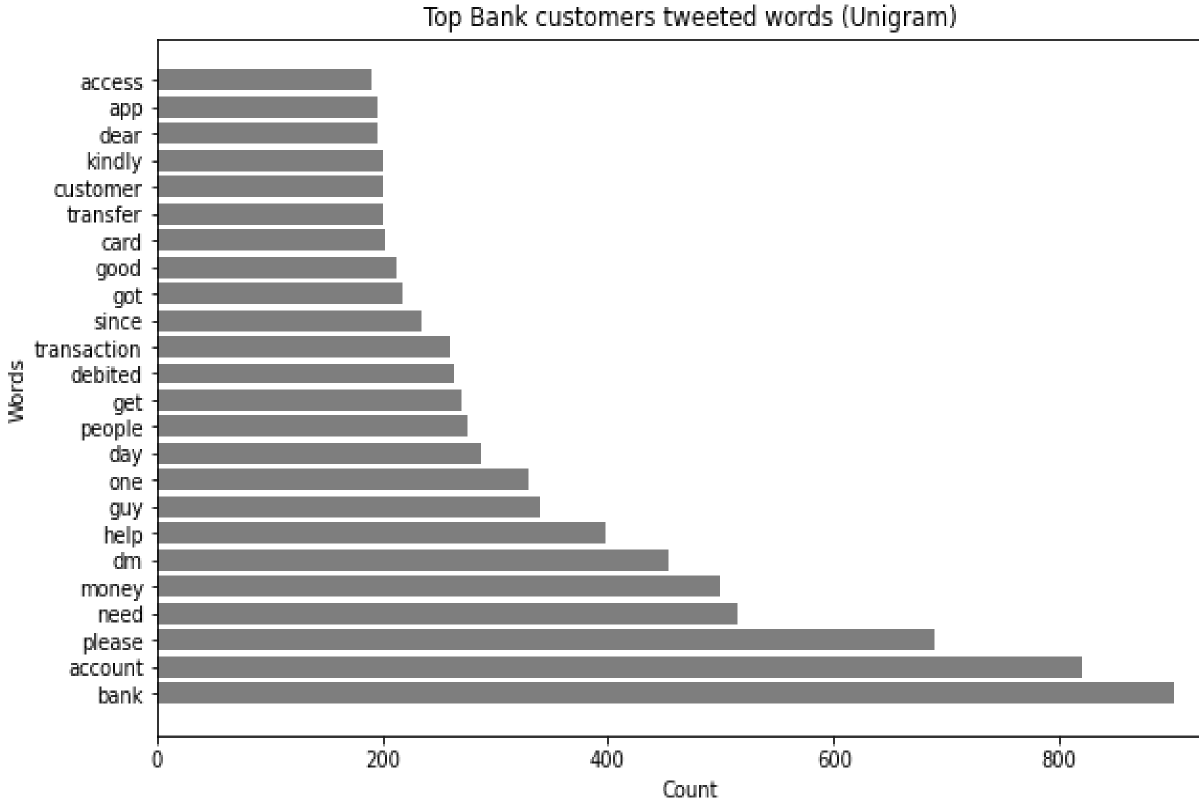


Figure 4.5: Plot of word (unigram) frequency

Figure 4.5 bar plot above, shows the words (unigram) with highest frequency tweeted by customers. These words vary in their interpretation. For example, words like 'good', 'morning', 'dear', 'customer' and 'kindly' suggests recognition or welcome. Words like 'account', 'bank', 'money', 'transaction', 'card', 'atm', and 'app' are domain (bank) specific nouns. These words are used mainly as target nouns when communicating a view or feelings towards bank operation or experience. For example, *@XYZ bank, your atm at Garki park is terrible but I managed to use your app*. Other notable domain specific words are 'debited', and 'transfer'. These words are bank operation terms that can be used to express different kinds of bank transaction. Time-bound words such as 'since' and 'day' were also frequently tweeted. These terms can be used to give more details on time of transaction or any other banking operation experience. Specific to this study, is the sentiment word like 'good' which also appeared to have been frequently tweeted. However, for now it is unknown if such word was used to express sentiment or just greetings (like good morning) as earlier stated. To have more insight into the order of these words or how it was used, a bi-gram bar plot was produced. The plot provides an understanding and order of two words and their frequency in the tweet.

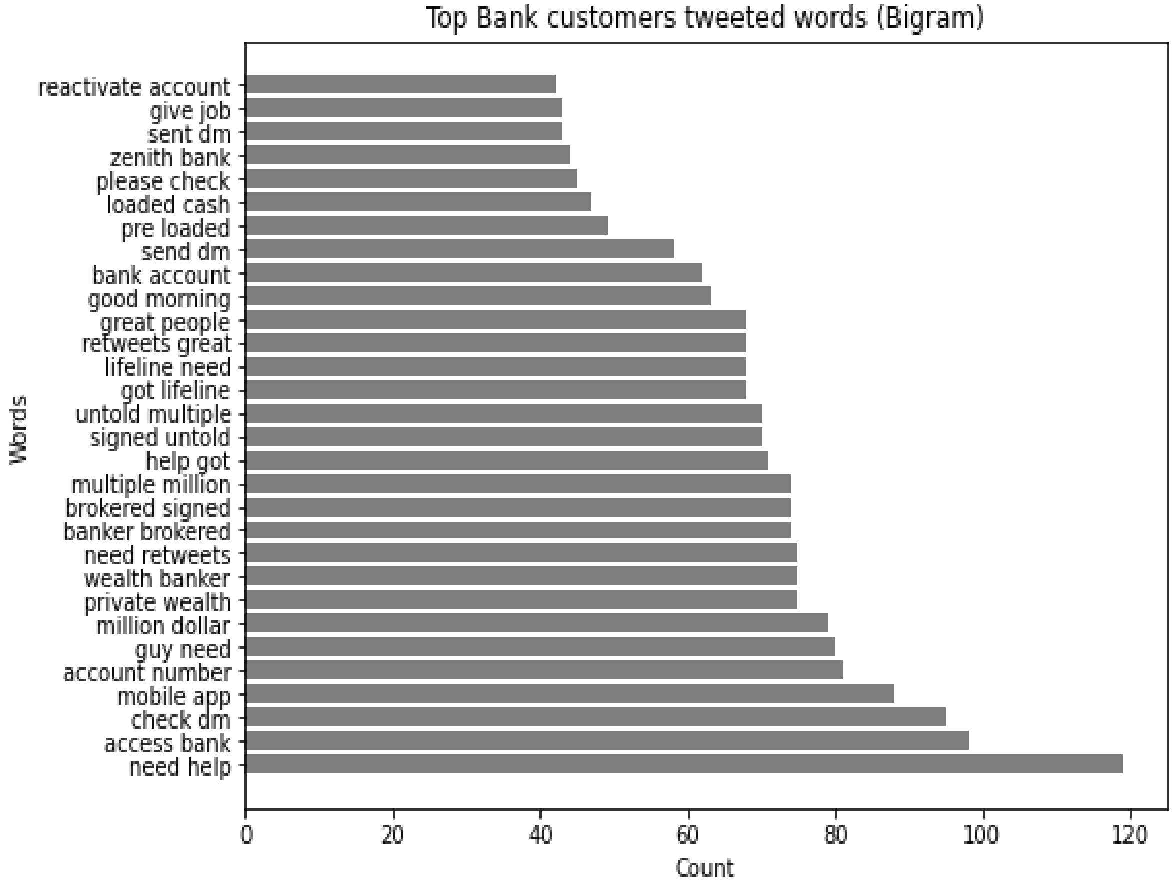


Figure 4.6: Plot of word (bigram) frequency

Figure 4.6 plot above shows the frequency of two words (bigram) used together in the tweets. Following on from the findings of the unigram plot, the bigram plot showed the term 'good' was frequently used as 'good morning' to express recognition or welcome. The domain specific (target) nouns identified earlier such as 'account', 'bank', 'money', 'transaction', 'card', 'atm', and 'app' were used frequently together as 'bank account', 'mobile app', 'access bank', and 'account number'. This implies target nouns such as access bank was highly tweeted. Despite the bank handles were removed during the pre-processing stage, there were a lot of tweets about the bank. Another very important target noun from the bigram keyword is the mobile app. This is a mobile digital application where banks offer their services to customers. It is interesting to know there a lot of Tweets that talked about 'mobile app'. This evidences the adoption and usage of the technology. Thus, the banks can use this channel as a competitive edge and fast delivery of service.

The bigram plot also shows bank operation terms such as 'reactivate account'. This objective bigram term is quite important in this context. It suggests customers tweeted a lot about reactivating their bank account. This occurs when accounts are inactive or dormant. Inactive accounts are bank account with no transaction for a period. When accounts are inactive for a long period, depending on the bank policy. The accounts go dormant which is a sign of customer churn in the Nigeria banking industry (Oyeniyi & Adeyemo, 2015). Thus, the banks will be glad to know their customers' tweet aims at reactivating their bank account as this is a positive 'objective term'.

Other objective terms also shown are 'please check', 'send dm', 'sent dm', 'need help', and 'check dm'. It appears these terms were used by customers to direct the banks to the post on their handles. Some terms were also observed in relation to customers' directing the banks to their post which were unattended to or needs urgent action. For example, 'need retweets', 'retweet great' and 'great people'. To have an in-depth knowledge on the bigram top words Table 4.2 below was produced.

Subject	Bigram Words
Positive (sentiment) terms	Refunded thanks, everyone thanks, best bank,
Negative (sentiment) terms	Useless bank, worst bank, failed transaction, account debited,
Target nouns:	ATM point, Credit card, ATM card, Verve card, Debit card, Mobile banking, Internet banking, USSD service, USSD code, Customer service, ATM service, Service issue, Mobile app.
Information	Account details, card details, minimum balance, bank balance,
Name of banks	Diamond bank, Stanbic IBTC, Access bank, Zenith bank, UBA, Sterling bank, Union bank
Objective terms	Reverse money, sent money, transferred money, airtime transfer, refund money, made payment, withdraw money, debited money, send email, debit without, pls respond, SMS alert, open account, app working, fraudsters could, please reverse, money reversed
Time bound terms	Till date, yesterday debited,

Table 4.3: Bigram top words not shown in bar plot.

Table 4.3 above provides a summary of the bigram (top) words which can be attributed to the subject in relation to this context. For example, it was observed that customers tweeted on positive sentiment terms (refunded thanks, everyone thanks, best bank,). This implies some customers have thanked the bank on service earned and/or Twitter users on how they have helped retweet their posts. The negative sentiment terms (useless bank, worst bank) indicate some customers considered their bank bad. This might be due to customers' growing frustration on service earned. For example, "*@XY bank, is a useless bank, they are yet to refund my account since last month"*. Furthermore, it was observed customers compared their banks using keywords like 'best bank' and worst bank'. Using knowledge gained from SentiLeye development, customers' often use factual words to demonstrate opinion. An example is the *"account debited".* The word implies a transaction in a customers' account. However, it was mostly used in this context to express frustration on service experienced. This explains why it is classed as negative sentiment bigram term. The bigram top word analysis also showed services (target nouns) like ATM, mobile banking, USSD service and internet banking as one of the key banking channels customers tweeted about. This is quite useful for the banks to know which of their services customers are using and thus, can be used as competitive edge or fast service delivery.

In summary, the unigram and bigram plots showed frequently tweeted keywords relating to welcome greetings, domain (target) nouns, bank operation (objective) terms, time-bound and sentiment words. To have a broader knowledge of the text data. The study further produces a wordcloud to uncover more keywords for extensive insight to the analysis. The visual plot can help provide more interpretation on keywords the customers tweeted about.

Figure 4.7: Wordcloud of bank customers' tweets

The wordcloud (figure 4.7) above shows a comprehensive visual representation of the keywords in the customers' Tweets. In addition to top words shown in the previous plots, keywords like branch, Nigeria, issue, million, naira and dollar are visible on the wordcloud. These keywords suggest customers tweeted about branch banking, banking issues, currencies, and problems encountered. All these keywords are in relation to service quality, delivery, and product such as ATM, mobile app, internet banking, transaction, airtime, and card which were also frequently mentioned. The keyword "airtime" suggest customers are using bank service to purchase phone airtime. It could also be that customers encountered problems using these services which should be resolved. Keywords like "transaction" and "refund" are words that can give insight of transaction error on account and thus, customers request for refund of money debited. These transactions might be through ATM, mobile banking, branch, or internet banking where money has been debited from customers' account. The wordcloud is useful to discover keywords (features) customers were tweeting about. The plots were helpful to understand these features, which the classification models use for training and prediction. However, data exploration is not an advanced analysis that can build model or make prediction. Therefore, this chapter will discuss the classification results of the sentiment analysis which can provide more definite answers to customers attitude, if positive or negative. Beforehand, it is helpful to explore the word length of the tweets to provide insight on the sentence length of the text data. Figure 4.8 below shows the word length of the tweets per class. The plot indicates the neutral class has the highest number of words, followed by the negative classed tweet and the positive class has the shortest length. Overall, the plot suggests the highest length of words per tweets is 17.

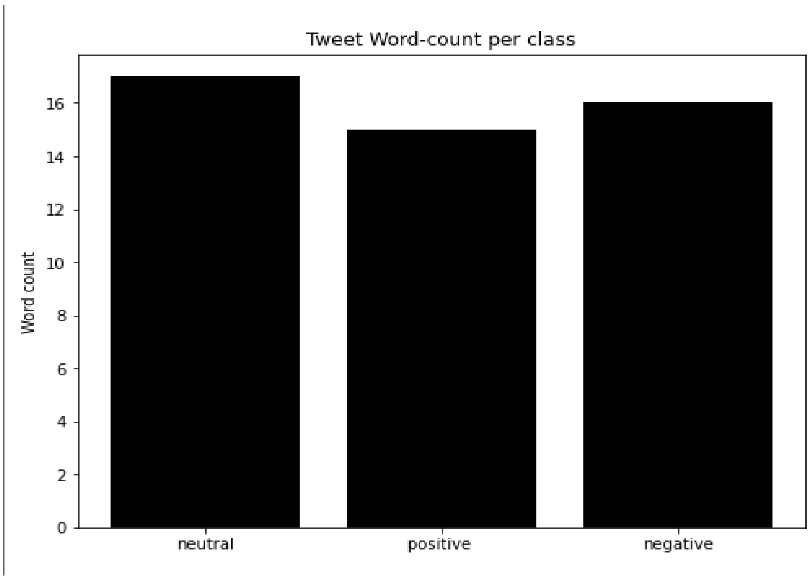


Figure 4.8: Tweet word length per sentiment score class

Since this study is focused on classifying the sentiment of bank customer tweets, it will be interesting to know the class ratio of the sentiment polarity labels. Basic visual plot (bar plot) was produced to check for class imbalance. For example: the number of negative sentiments class compared to neutral and positive classes. This was checked against each other to make sure these classes are equal or otherwise. In real-world dataset like this, sentiment classes are mostly imbalance. This is different to 'product' review dataset where people rate service or product from one to five as an example. In this case, it is a situation where customers generally do not just go online to praise their banks but more often to make a complaint or ask for information. For example, tweets like *"@XXbank, how much is the daily limit of your ATM card?"* can be classed as neutral because it is an objective statement that does not show any sentiment. Whereas sentences like *"@XXbank, your ATM card is too limited for me. I need to withdraw more than 90k naira"* can be classed as negative and *"@XXbank, your ATM card is fab cos it limits my lavish expenses."* can be classed as positive. These sentences demonstrate the difference between objective sentence (neutral) and subjective sentences (positive or negative) in this context. The plot produced in figure 4.9 below shows there is class imbalance problem.

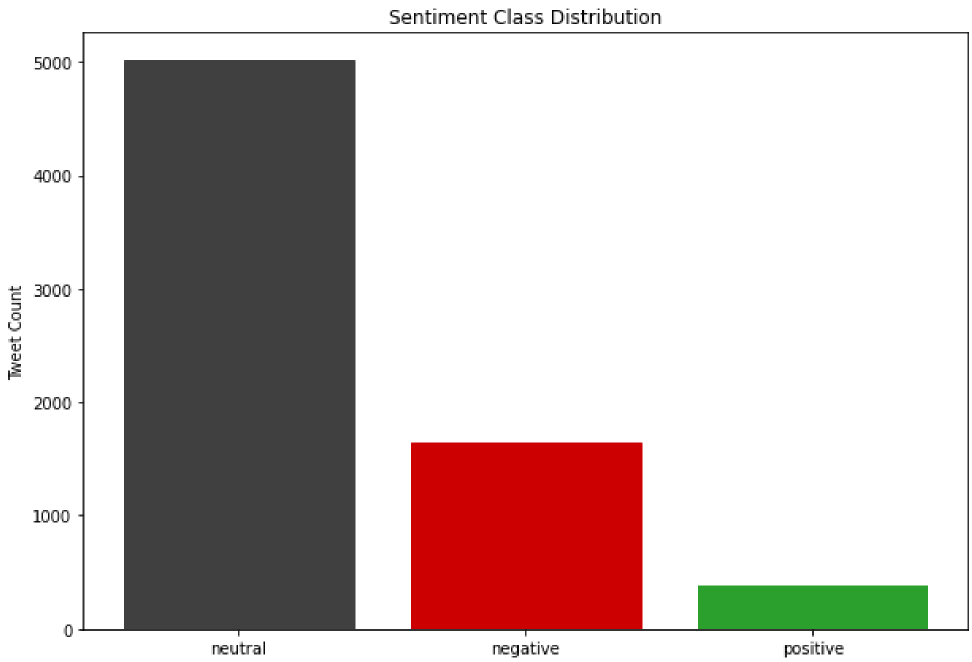


Figure 4.9: Class distribution of sentiment class.

Figure 4.9 above shows that the positive class has the lowest number of tweets in the annotated dataset. Followed by the negative and the highest is the neutral class. It is not statistically efficient to use this data directly on the classification models. To avoid the issue of classification models' biasedness towards the majority class which can result to high misclassification rate when deployed. There is need to re-balance the classes. Therefore, this study will investigate suitable approaches in handling class imbalance.

4.4 Handling Class Imbalance

This study defines class imbalance as the disproportion of numbers of the different sentiment classes present in the tweet (or data). Class imbalance is an issue of great significance in text classification problems because it hinders the performance of classification models (He & Garcia 2009; Longadge & Dongre, 2013). The supervised classifiers tend to be biased towards the majority class because the rules that correctly predicts the instances are positively weighted in favour of the accuracy metric or the corresponding cost function. Thus, the minority classes are more misclassified. For example, in a case where there is a disproportionate number of positive sentiment class compared to negative. The classifier tends to overlook the minority class when fitted and thus results in misclassification when the model is deployed. Hong, Nam, & Cai (2019) expressed their concern that previous SA studies fail to account for class imbalance in their approach and thus affects the performance of the classifiers. For example: Al-Omari, et al. (2019), trained logistic regression model with 353 negative and 2779 positive review dataset. Thus, reported the performance of the model was poor while predicting negative sentiment. This is untrue because the classifier is skewed towards positive reviews and there was need to treat the imbalance problem before fitting the model. Another example is the study of Xing et al. (2020) that trained supervised models with StockSen dataset (imbalanced) and their model result showed more false positive error than false negative. This is because their training set has more positive classes. To avoid such problem in this study, it is necessary to explore the data to check for class imbalance. Thus, rebalancing the class distribution is essential.

There are several approaches to handle class imbalance. This study does not look to dive deep into class imbalance. However, it is necessary to investigate different approaches and select the most appropriate method. A review of class imbalance studies indicate there are two main approaches which are, data level and algorithm level methods. Data level method is more popular and effective due to its historic existence in statistics which involves sampling. The method includes techniques such as undersampling, oversampling or combination of both (hybrid) which will be discussed below. While the Algorithm level methods involve using data mining and machine learning tools.

4.4.1 Undersampling

This type of data-level class rebalancing technique involves removing a subset of the majority class to balance with the minority class. In many applications, this technique has been applied (García & Herrera, 2009; Moniruzzaman et al. 2020). An enhanced approach called Random undersampling was also developed to randomly remove samples to balance the class distribution. Other notable undersampling techniques are Inverse Random undersampling, Diversified sensitivity undersampling and clustering based undersampling. However, in this context, these undersampling techniques are deemed inappropriate because they will discard potential useful information.

4.4.2 Oversampling

This technique involves replication of samples in the minority class or creating new samples using existing ones. A non-heuristic approach to oversampling is the Random oversampling method. This technique is applicable in this context however, it has been criticized as it causes overfitting due to samples replicated randomly. To overcome this problem, SMOTE (Synthetic Minority oversampling technique) was developed by Chawla et al. (2002). The SMOTE approach aimed to rebalance the class distribution with synthetic samples initiated by randomised interpolation between several minority instances that lie together. SMOTE iterative process can be summarized as follows:

- X_i minority class instance is selected at random.
- Based on distance metric, several k-nearest neighbours ($X_{i1 \ldots .}$ X_{ik}) are selected depending on the number of k (default number of k is 5).
- Randomized interpolation is conducted to obtain the new synthetic instances depending on the total amount of oversampling.

Notable studies have also enhanced the SMOTE technique to improve on the original SMOTE version which worked well in the domains applied. For example: Borderline SMOTE (Han et al. 2005), ADASYN: Adaptive synthetic sampling (He et al. 2008), Safe-level SMOTE (Bunkhumpornpat et al. 2009), MSMOTE (Hu et al. 2009), CURE-SMOTE (Ma & Fan, 2017), LVQ-SMOTE (Nakamura et al. 2013), and MWMOTE (Barua et al. 2012).

4.4.3 Hybrid

The hybrid approach is the combination of undersampling and oversampling techniques. Both approaches have their weakness and strength. To enhance the performance of existing sampling techniques, few studies developed the idea of complementing the strength of these approaches with each other. This has been used successfully to improve on existing class rebalancing techniques. For example: SMOTE-RSB (Ramentol et al. 2012) and RHSBoost (Gong & Kim, 2017).

4.4.4 Algorithm Level Methods

The method focuses on modifying the classifier's learning procedure. Not popular due to their difficulty in design and implementation. This approach adjusts the cost measurement during the training process to consider the class imbalance. The straightforward way is to automatically generate the cost matrix based on the class distribution. However, in real world dataset, it is not always straightforward so this can be done during the learning process. When using decision tree, the approach is to adjust the probabilistic estimate at the tree leaf. This technique makes the minority class samples more important than the majority class samples. There are algorithms which have been proven to be effective in dealing with class imbalance such as K-Nearest Neighbours and Support Vector Machine (Ganganwar, 2012).

4.4.5 Implementation of Class Rebalancing Techniques

This study considers the undersampling techniques unsuitable because they discard potential useful information. For example, in this case where the neutral class has the highest number of tweets in the annotated dataset. If undersampling approach is applied, the loss of tweets limits the dataset available to train the ML models. The algorithm sampling technique were not utilized because the study aimed at comparing different classification models. Thus, becomes practically difficult to implement this approach with all the models. The original SMOTE technique was considered to rebalance the dataset because it has been used successfully in several studies including domains with high dimensional data (Fernández et al. 2018; García et al. 2016; Fallahi & Jafari, 2011). Most specifically, used in the Indian banking domain (Shrivastava et al. 2020) and the way the technique works is very clear and appears more appropriate in this context. The oversampling (SMOTE) technique was implemented using the imblearn library (Lemaître et al. 2017) in Python. The number of nearest neighbours was set as k = 5. At K=5, the algorithm connects the inliers and outliers' features. This is used to construct the synthetic samples. The annotated dataset contains

7,037 tweets and were split in ratio of 80:20 of training and testing set, respectively. The training set amounts to 5629 and test set amounts to 1408 tweets. The training set contains 4022 neutral, 1306 negative and 301 positive tweets and were balanced using SMOTE technique as shown in figure 4.10 below.

```
Before SMOTE : Counter({'neutral': 4022, 'negative': 1306, 'positive ': 301})
After SMOTE : Counter({'negative': 4022, 'neutral': 4022, 'positive ': 4022})
```

Figure 4.10: SMOTE output of the annotated dataset

4.5 Feature Representation of Text

Generally, machine learning models do not understand text (input) data directly when inputted. These models can only recognise numeric representation. Therefore, this stage involves the description of different approaches to that effect. This is very important because it helps models in performing effectively in text classification task and contributes to the performance of the classifiers (Wang et al. 2019; Kasri et al. 2019). Vector space model is the mathematical model to represent unstructured text in a computational system as numeric vector (Salton et al. 1975). There are various methods such as one-hot encoding and bag of words model. However, this study will not consider these techniques because they are basic and do not consider the frequency and sequence of words in a document. For example, the term ‘not good’ means bad. When split into single words, that results to ‘not’, ‘good’ which does not interpret to bad. Thus, this study will consider using more advanced frequency-based vector space models such as N-gram model and the TF-IDF.

4.5.1 N-Gram Model

This is a popular language model that recognises the frequency and sequence of words in a document. N-gram model is a statistical language model that assigns probabilities (using frequency count-based system) to sentences and sequence of words. The N indicates the number of words in a document. For example, a unit gram word is ‘good’, or ‘bad’ which is known as 1-gram words. Bi-gram words are ‘not good’, ‘not bad’ known as 2-gram words. Trigram is ‘I love this’ also known as 3-gram words. This model in natural language processing is popular because it can predict the neighbouring words in a document.

4.5.2 TF-IDF

This frequency-based approach helps recognise the importance of words present in all documents of a corpus. The approach is more helpful in identifying words that were rarely used but are of high significance to the document. The model is the multiplication of two major steps namely, term frequency (TF) and inverse document frequency (IDF). TF is the ratio of the numbers of occurrence of a word in a sentence to the length of the entire sentence. IDF is the log of the ratio of total number of rows to the number of rows in which the word can be found. This measures the rareness of a word. That is, it helps identify the importance of words that might have been overshadowed by simple count-based system. For example, a word counted 4 times in a sentence of 20 words length compared to a word counted 4 times to a sentence of 100 words length. The word in the first instance is more important than the other instance. TF-IDF presents the importance of word to a document in a collection and thus normalise words that appear frequently in all documents.

4.5.3 Word Embedding

Word embedding has been shown to improve performance of text classification models especially when applied to deep learning model for sentiment classification (Kim, 2014; Ruder et al. 2016). This is because the semantics, structure, sequence, and context in which the words are used in a document is put into consideration. To train and generate word embeddings, Word2Vec is considered in this study because the model generates high quality vector representation of words that capture semantic and context (Mikolov et al. 2013). Word2vec is an unsupervised model in which word embeddings are learned using distribution of word co-occurrences in the context. This model was developed by Google in 2013. The algorithm takes text corpus as input, builds vocabulary of words, generate dense word embeddings for each word in the vector space and then produces the word vectors as output.

In summary, these approaches are appropriate numeric vector models which are widely used. They will be used in this study to input tweets into the classifiers discussed in the next section. It will be interesting to understand how they differ in their contribution in terms of their performance to the classifiers.

4.5.4 Implementation of Feature Representation of Text

The count vectorizer from Sklearn feature extraction library (Pedregosa et al. 2011) was used to implement N-gram model in Python. The model was set in the range of N = 1-3. This was used to generate maximum features of 5,000. This vector space model is frequency based however, is aware of neighbouring words where n=2 (bigram) and n=3 (trigram). The TF-IDF vectorizer from Sklearn feature extraction library (Pedregosa et al. 2011) was used to implement TF-IDF model. The parameters were tuned to maximum features of 5000 and the N-gram range was set as N= 1-3. The decision to choose this parameter was to give the bigram and trigram a chance to have a fair comparison between both N-gram and TF-IDF models. TF-IDF parameter chosen implies that the model will transform text of unigram, bigram, and trigram terms into TF-IDF vector model.

In addition, this study utilised Word2vec to train and generate word embeddings. The model was introduced in the study of Mikolov et al. (2013). This helps to capture the semantic relation of words and context. The Gensim package (Rehurek & Sojka, 2010) in Python was used to implement the embedding. The skip-gram model helps to predict the probability of a word considering the context. The algorithm takes text corpus as input and then produces the word vectors as output. The implementation used tokenized words of the cleaned Tweets. The context window size was set at 3 and the minimum word count of words was set to 1. These vector space models will be used to observe the differences to the performance of the classification models.

4.6 Learning & Modelling Stage

Model performance varies across different domains. Unfortunately, no study was found to have applied these models in the banking domain to ascertain their performance. To determine the most efficient sentiment classification model, there is need for a conceptual understanding of how these algorithms work. The study will conduct a comparison of well performing statistical, machine and deep learning models. Firstly, the study will review the iteration process of each of this model to understand their appropriateness in this context.

The selection of these statistical, machine and deep learning models is based on their successful application to text classification of natural language. The models have performed well in sentiment analysis application as shown in literature reviewed in chapter three. It is important to understand that the annotated data are of three classes (positive, negative, and neutral) which means binary classifiers like logistic regression needs to be adjusted as a multinomial classifier.

4.6.1 K Nearest Neighbour

KNN is a powerful supervised learning model that makes no underlying assumption of the data or any functional form. The model is widely used in sentiment analysis (SA) applications, specifically in hotel review domain (Dey et al. 2016) and has shown better performance in-comparison to other models (Jain & Katkar, 2015). The non-parametric classification technique classifies instances by calculating the nearest neighbour using distance metrics among training set and thus predict by choice of nearest majority to test set. The iteration process can be summarized as follows:

- Define K, the total number of nearest neighbours.
- Calculate the distance between the test data and all training data.
- Sort nearest neighbour by minimum distance to value K.
- Majority class of the neighbour is used to determine the class of the new instance

In most cases, the problem with KNN is the difficulty in assigning the best value for K. This depends on what domain or task at hand. However, it is good to note that tuning parameter can be helpful as different k value which yields different result.

4.6.2 Naïve Bayes

Naïve Bayes (NB) is a supervised learning model based on Bayes theorem and uses probabilities in classifying data. NB was formulated by Thomas Bayes (McCallum & Nigam, 1998). The model is fast, simple to implement, highly scalable, and widely used for SA classification (Abbas et al. 2019). The probabilistic model works perfectly with small training set (Rasjid & Jetiawan, 2017), has low storage requirement and dependable as baseline algorithm in text classification problems (Rennie et al. 2003; Feng et al. 2015).

NB class with maximum probability is chosen as the predicted value or class. The probability of a class is the ratio of number of documents in class to the total number of documents. The likelihood of a word given a class will thus be the ratio of the number of a word occurring in that class to number of all words in that class. For example: a word 'y' from test set and a window of n-words ($X_1, X_2 \ldots. X_n$) from a document. The conditional probability of given data point 'y' to be in the category of n-words from training set is given by:

$$P(y/ X_1, X_2 \ldots. X_n) \quad = \quad P(y) * P(x/y)/P(X_1, X_2 \ldots. X_n) \tag{1}$$

NB does not consider the relationship amongst features and that is why it is considered optimal classifier if features/predictors are assumed independent. However, the probabilistic classifier is also successful when feature dependencies exists because the classification error is not necessarily related to the quality of fit to the probability distribution. The model is robust to irrelevant features (the irrelevant features cancel each other without affecting results) and very good in domains with many equally important features (Abbas et al. 2019).

4.6.3 Logistic Regression (Multinomial)

Logistic regression (LR) is a traditional statistical model developed by David Cox in 1958 (Cox, 1958). LR has performed well in text classification problems (Milu et al. 2020). Specifically, in the publicly available review dataset (from Amazon, IMDB and Yelp). For example: Ghosh et al. (2020) showed logistic regression performed better than SVM and decision tree in the three different domains. LR uses a logit function to estimate probabilities between these classes. The model therefore predicts a instance to the class with highest probability.

The model can be extended to handle multi-class target variable. In this context, the Multinomial Logistic Regression (MLR) is considered because the dependent variable is of three different classes (positive, neutral, and negative). The 'OVR' (One versus Rest) function is used to implement the binomial classification. That is, positive vs (neutral and negative), negative vs (neutral and positive), and neutral vs (positive and negative).

4.6.4 Random Forest

Breiman (2001) introduced this ensemble method which can be used for both classification and regression problems. Random forest (RF) is the ensemble method that uses random subset of features (from training set) to train multiple independent decision trees (bootstrap) and predict instance by majority voting of each tree outcome. The model is easy to interpret, runs efficiently on large database, fast to train and scalable, performs well in complex dataset, and robust to irrelevant features (Misra & Li, 2019; Milu et al. 2020). However, it is sensitive to overfitting which can then be regulated using the numbers of trees. RF handles dataset with huge (input) features well and essentially suited for multi-class problems (Alzamzami et al. 2020). RF has performed well in text classification problems (Gupte et al. 2014). RF iteration can be summarized as follows:

- Select k data points randomly

- Build decision trees with the k data points
- Predict y value based on aggregate of class (outcome) voting

4.6.5 Support Vector Machine

Support vector machine (SVM) was proposed by Vapnik in the 1990s (Vapnik, 2013) and has been shown to be a reliable and well performed supervised learning model in sentiment analysis application (Chory et al. 2018).

SVM is a linear classifier that also performs non-linear classification problem using the kernel function. The model constructs a hyperplane between separated marginal lines of the nearest support vectors (input vector) of the classes. In the space, an optimal separating hyperplane is determined by maximizing the marginal distance. In non-linear classification problem, the algorithm maps the data into input vector and uses the kernel function to transform low dimensional input space to a higher dimension to solve non-linear problems. The maximum margin hyperplane is used for optimal classification of new instance (class). The higher the marginal distance between the classes the more generalizable the result is.

The model regularize parameter to naturally avoid overfitting and bias problems seen in other algorithms (Basari et al. 2013). The model is good at dealing with high dimensional dataset, works well with sparse document vectors, performed better in comparison to other models and across different domains (Tan & Zhang, 2008). For example: SVM outperformed deep neural network & CNN in Arabic Tweets on health services (Alayba et al. 2017), KNN (Istia & Purnomo, 2018) and RNN in Arabic dataset of SemEval 2016 task 5 hotel reviews (Al-Smadi et al. 2018). In this study the radial basis function (rbf) kernel was applied, and the decision function shape was also adjusted to OVR (One versus Rest) to perform multi-class classification.

4.6.6 Gradient Boosting

Gradient boosting (GB) is an ensemble machine learning method that combine weak classifiers together to construct a robust learning machine (Rao et al. 2019). In this case, the models in the bootstrapping technique are combined in sequence to iteratively minimize errors. The models reduce/complement failures of individual classifiers and thus improve overall performance of the model. The tree model is optimized by reducing expected generalization error. The model generalizes by allowing optimization of an arbitrary differentiable loss function. The loss function is a measure that indicates the efficiency of coefficients of a model that fit the underlying data. Thus, the loss function is to be minimized to improve performance. GB has proven to handle high dimensional and sparse dataset (Athanasiou & Maragoudakis, 2017). Alzamzami et al. (2020) proved the technique is useful to develop general purpose sentiment classifier system. Past studies showed GB handled class imbalance problem well by yielding better F-score for minority class and shown to outperform other models for multi-class sentiment classification task (Athanasiou & Maragoudakis, 2017; Alzamzami et al. 2020).

4.6.7 Convolutional Neural Network

Convolutional neural network (CNN) is a hierarchical network architecture (LeCun et al. 1998). CNN uses multilayer perceptron with three main components namely, convolutional layer, pooling layer, and the fully connected layer (with activation function and drop-out) to perform classification task (Kim, 2014). The input layer transforms words into input vectors. The convolution layer detects key features in a sentence irrespective of the position. The pooling layer conducts subsamples from the output of convolutional layer and increases the invariance. The maximum pooling operation takes the maximum value to the output according to the filter pooling size. The fully connected layer (output layer) presents the classification result.

CNN is popular in image classification problems but also works well for sentiment classification task (Widiastuti, 2019). The method performs well on natural language processing (Zhang & Wang 2015) and better in-comparison to other neural networks such as RNN (Vu et al. 2016) and traditional machine learning models such as Naïve Bayes and SVM (Zhang et al. 2019).

4.6.8 Recurrent Neural Network

In natural language processing, the order of word is important to capture the semantic relation of words in a sentence or document since words have different meaning in different context. The information retrieved from such sequence of words enhance sentiment classification performance (Wang et al. 2018). That defines why Recurrent Neural Network (RNN) is a preferred approach to ordinary Neural Network (Rao et al. 2019). Recurrent Neural Network (RNN) is a type of deep neural network which has performed well in language modelling due to its sequential processes. The model performs well and robust in sentiment classification task because the algorithm can model context dependencies (Tang et al. 2015). RNN architecture consists of the input, hidden and output layers. The network treats input and output layers independently while learning long term dependencies in a sequence. RNN is a nonlinear model that can learn deeper structured information. Purnamasari et al. (2017) showed RNN is better than CNN. However, despite the praises that the model has earned, it is worth noting that RNN suffers from exploding gradient problems while learning long range dependencies. The model experiences the phenomenon of gradient disappearance or explosion during the processing of long time series (Rao et al. 2019). That is, gradient blows up or decays exponentially over time.

4.6.9 Long Short-Term Memory

Long Short-Term Memory (LSTM) was proposed by Hochreiter & Schmidhmber (1997) as an improvement on the recurrent neural network (RNN). LSTM models sequential data effectively (Sundermeyer et al. 2012) and solves the problem of gradient exploding in RNN. The network has shown better performance in sentiment classification task than other neural networks such as CNN (Sun et al. 2019).

In LSTM network, the input layer maps word to real value vector like other neural networks. LSTM layer captures contextual semantics relations of the words. Output layer presents the classification result. Multilayer LSTM contains one or more layers that can learn both linear and non-linear function while the single layer contains one layer that learn only linear function (Sak et al. 2014). The most important element of the LSTM network is the memory cells. LSTM network includes the input gate, forget gate and the output gate used to interact with the memory cells. The gates use the activation function (For example, Sigmoid to transform input vector values to values between 0 and 1). The input gate determines important information from the input layer and add to memory cell. The forget gate removes unwanted information from the memory cell and the output gate fetches the important information from the memory cell to help in prediction. This memory cell is the advantage of LSTM over RNN because it can remember the long-term memory.

4.6.10 Bi-Directional Long Short-Term Memory

LSTM encoder reads input (x_1, x_2, ... x_n) of a sentence where x_i is the first term and x_n is the last term. However, only first few words of x_i terms are obtained in the correct context. Thus, leads to misclassification (output). The Bi-Directional Long Short-Term Memory (Bi-LSTM) was developed to solve this problem. Bi-directional LSTM has high level representation of forward and backward sequences incorporated in two layers of LSTM network (Polignano et al. 2019). The network reads input in the order of $x_1, x_2, \ldots. x_n$ in one layer and the other layer reads from the last term $x_n, \ldots\ldots x_1$ to capture more of all terms in the correct context. The model has also been applied successfully in sentiment classification task (Al-Smadi et al. 2019). Thus, will be utilized in this context.

4.7 Implementation of Machine Learning Models

K-Nearest Neighbour was implemented with five neighbours considered (K=5). The weight was set to uniform. This means all points in the neighbourhood have equal weights and the distance was calculated using the Euclidean distance. Linear classifiers such as Naïve bayes was applied using the MultinomialNB function. The MultinomialNB was used for the classifier to understand the multi-class labels. Logistic Regression was also implemented as a multinomial model to capture the multi-class labels. The maximum number of iterations taken for convergence was set to 1000. Support Vector Machine was implemented using two different kernel functions. Firstly, as a linear classifier and the other as non-linear classifier to understand which of the parameters perform better for the model. SVM uses kernel function 'linear' for linear classification and 'radial basis function' (rbf) for non-linear classification. The other kernel functions also tried for tuning the parameters were sigmoid and poly. The study employed GridSearchCV in tuning these parameters. The sklearn model selection package provides the GridSearchCV function. The decision function shape was set to ovr (one versus rest).

The ensemble models were implemented with different numbers of estimators to observe their performance. In the case of Gradient Boosting, the model was implemented with 10 numbers of estimator (tree) and the random state was set to 42. This controls the random seed given to each estimator. The Random Forest classifier was implemented with parameter bootstrap set to true. The numbers of estimator (tree) used was 100. This means the multi-decision tree (100) was trained with sub-samples of the training set.

The deep learning models employed are convolutional neural network, recurrent neural network, long short-term memory, and bi-directional long short-term memory. The neural networks were built in Python using the Keras package (Chollet, 2015). Chollet developed Keras in 2015 as part of the research project ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System). Convolutional Neural Network (CNN) was implemented using 128-dimensional embedding layer, two convolution layers with activation function ReLU (rectified linear unit) and two hidden MaxPooling layers. The network also consists of batch-normalization, flatten and dense layers. The network used dropout of 0.3 to regularize the model and SoftMax activation function to output the probability outcome for each sentiment class. The model was fitted with a batch size of 64, 20 epochs and applied Adam optimizer to train. The Recurrent Neural Network consists of 128-dimensional embedding layer, recurrent neural network layer and the dense layer with SoftMax activation function. The model was fitted with a batch size of 32, 20 epochs and applied Adam optimizer to train. Long Short-Term Memory network was built using 128-dimensional embedding layer and LSTM layer with dropout of 0.3 and recurrent dropout of 0.4. The network also consists of batch-normalization and dense layers. The network used the SoftMax activation function to output the probability outcome for each sentiment class. The model was fitted with a batch size of 32, 20 epochs and applied Adam optimizer to train. Bi-Directional Long Short-Term Memory consists of 128-dimensional embedding layer and Bi-directional LSTM layer with dropout of 0.1 and recurrent dropout of 0.1 The network also consists of a convolution layer, GlobalMaxPool and dense layers. The network used the SoftMax activation function to output the outcome for each sentiment class. The model was fitted with a batch size of 32, 20 epochs and applied Adam optimizer to train. In the deep learning models, a dictionary structure was created for class weight to balance the weight of the target classes.

4.8 Model Evaluation

This section discusses the evaluation metrics of the models to understand how well they have performed in this context and help decide on the best model. There are several evaluation metrics such as accuracy, precision, recall, and f1-measure. The proportion of correctly and wrongly labelled tweets will be described in the form of true positive, true negative, false positive and false negative shown as a 2 by 2 confusion matrix in figure 4.11 below. However, the implementation recognises it is a multi-class in this context. Thus, measures in a one-versus-rest manner.

		MANUAL CLASSIFICATION	
	SENTIMENT CLASSIFICATION EVALUATION	Positive	Negative
SENTIMENT ANALYSIS TECHNIQUE	Positive Sentiment tweet	TP	FP
	Negative Sentiment Tweet	FN	TN

Figure 4.11: Confusion matrix

Where: True Positive (TP) i.e. Positive tweets classified correctly

False Positive (FP) i.e. Negative tweets wrongly predicted as positive (Type I error)

False Negative (FN) i.e. Positive tweets wrongly predicted as negative (Type II error)

True Negative (TN) i.e. Negative tweets classified correctly

Accuracy: is the ratio of number of samples predicted correctly to the total number of samples.

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Precision: denoted as P is the percentage % of selected items that is correct.

$$\frac{TP}{TP + FP} \tag{3}$$

Recall: denoted as R is the percentage % of correct items that are selected.

$$\frac{TP}{TP + FN} \tag{4}$$

F1 measure provides the balance between the precision and recall and can be denoted as

$$2PR \,/\, P + R. \tag{5}$$

The primary focus of the banks is to have a reliable and accurate prediction of negative tweets. Since customers proactively engage banks when they are poorly served (with negative tweets) or need information (neutral) but seldomly tweet to praise (positive) banks on service earned. The imbalanced distribution of the real-life dataset limits accuracy and error rate to decide on best performed model as these measures favour the majority sentiment class (He & Garcia, 2009; Bowes et al. 2012). Thus, accuracy and error rate are not sufficient to decide on the best performing model. Considering this, recall will be appropriate. However, a good precision score will also be considered to understand how reliable the classification model has performed. Thus, this implies the F1-score will play a significant role in choosing the best performed model as this provides the balance between the precision and recall.

4.9 Topic Modelling

The sentiment classification components have been discussed in the previous sections. Thus, this section will focus on the aspect extraction for sentiment analysis. This is the third component of the TSBS framework. The statistical topic modelling methods will be adopted since both explicit and implicit aspects are needed to form the banking themes in this context. In the experiment, this study will compare the unsupervised topic models namely, latent semantic indexing (LSI), latent Dirichlet allocation (LDA), and hierarchical Dirichlet process (HDP). LSI uses a singular value decomposition (SVD) of large term-document matrix to identify a linear subspace (in the space of term frequency inverse document frequency features) such that the relationship between the term and document are captured. The method transforms dataset into different space and thus map semantic association (structure) of terms with document. HDP is a non-parametric Bayesian model that recurse parametric function to cluster multiple grouped data (Chien, 2017). HDP involves dealing with representation of multiple grouped data where each group is associated with mixture model from a local Dirichlet process i.e. Each local Dirichlet process governs the words generation for the group. Words from different group are represented by global mixture model from a global Dirichlet process. LDA is a generative probabilistic model that was developed to consider the exchangeable representation for both documents and words. The aim is to capture the important intra-word/document statistical structure via the mixing distribution. These topic models were chosen due to their performance evidenced from literature reviewed (section 3.8 of literature review chapter). The frequency-based approach and part of speech (POS) tagging were deployed to enhance the statistical topic models. The frequency-based approach was used to announce the frequent important aspects. The POS function can help detect rarely used words which are also important. Thus, both approaches will be used to complement the unsupervised topic models. The Twitter dataset from previous chapter was employed.

4.9.1 Implementation of Topic Modelling Method

This section describes the implementation of the statistical topic models (latent semantic indexing, latent Dirichlet allocation, and hierarchical Dirichlet process) employed in this study. The MALLET version of LDA was considered because it has shown satisfactory or better performance in this field of study (Akhtar et al. 2017; Habibabadi & Haghighi, 2019; Karami & Elkouri, 2019; Wang & Li, 2020; Asghari et al. 2020). The methods were implemented in Python programming language and highly sophisticated Python libraries were utilised for the procedures. For example, NLTK (Bird et al. 2009; Perkins 2010) and Gensim (Rehurek & Sojka, 2010) packages. The text data was pre-processed as detailed in the methodology chapter. The pre-processing pipeline includes the lemmatization, stop word removal, and tokenisation procedures. The NLTK package was used for stop-word removal, lemmatisation, tokenisation, and part of speech (POS) tagging. POS are the lexical categories in which words are assigned based on their syntactic context. The process of mapping words with these parts of speech is called part of speech tagging. This technique is popular in natural language processing (NLP) to understand syntax and semantics of words. For example, a case where a system needs to detect the most frequent noun in a document. This technique is quite useful to extract any POS (such as noun) from the document and afterwards apply a count function to understand the frequency. In the experiment, this study made use of tagged nouns, adjectives, and adverbs. The POS were extracted in the form of unigram, bigram, and trigram (n-gram = 1-3). The tokenised text data formed thus formed the corpus which each of the models were built on. The frequency threshold was set at 20. The Gensim package was used to construct the topic model pipeline which consists of the LSI, LDA and HDP. The parameters of the models were adjusted. For example, the chunksize represents the number of documents used in each training chunk was set at 1740. The alpha represents the number of expected topics that expresses the prior belief for each topic probability. The eta represents the prior belief on word probability. These were set to auto to learn as asymmetric prior from the corpus. The passes represent the number of passes through the corpus during training and was set to 20. The iterations number is the maximum number of iterations through the corpus when inferring the topic distribution of a corpus. This was set to 1000. The unsupervised models were iterated to classify 20 terms per topic and 20 number of topics. The techniques produced cluster of terms regarded as topic and is a probability distribution over words. To have a clear understanding and summarise the topic modelling processes, this study presents a flowchart in figure 4.12 below to visualise the steps taken.

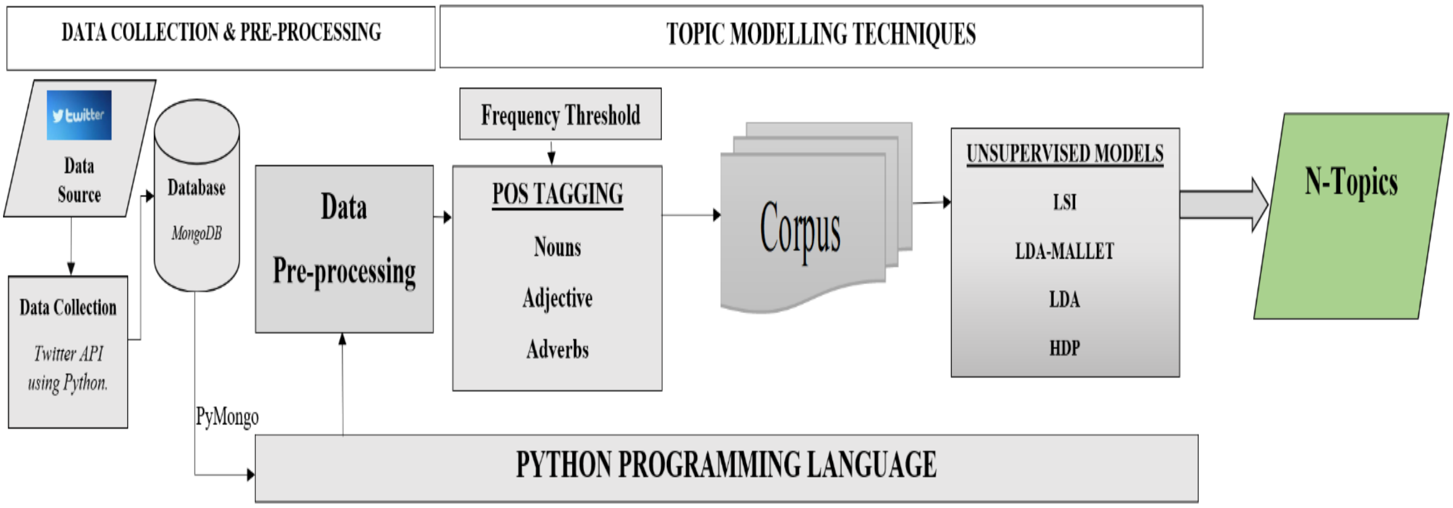


Figure 4.12: Flowchart of Topic modelling process

4.10 Chapter Summary

This chapter presented the research strategies and methods employed in this study. The rationale for choosing quantitative and deductive research strategies were discussed. The focus of the chapter was on the development of sentiment analysis approaches and thus, proposed TSBS framework for this purpose. The framework comprises of three main components namely, lexicon-based approach, machine learning approach for the sentiment classification task and topic modelling for the aspect extraction task and the techniques for each of the components ranging from data collection, data pre-processing, feature representation of text, modelling to model evaluation were discussed.

With the lexicon-based approach, sentiment wordlist was generated using both the corpus-based approach (tweets) and dictionary (WordNet) based approaches. The corpus-based approach was used to extract the bank domain, pidgin English and English words. The dictionary-based approach was used to increase the coverage of these words. To make the lexicon suitable with social media text, this study adopted the emoticons, and slangs wordlist from the Sentistrength (Thelwall et al. 2010) lexicon due to the performance of the lexicon with social media data. Human labelling of words was conducted using three main annotators who are native speakers of pidgin and English. The SentiLeye algorithm was thus developed which is one of the contributions of this chapter. To utilise the supervised machine learning algorithms and due to unavailable training dataset, this study labelled 7037 Nigeria bank customer tweets which is the second contribution of this chapter. The labelled dataset is useful to build and improve SA model performance in the banking context. Exploratory data analysis (EDA) tools such as wordcloud and barplot were employed to understand the characteristics of the text data. Techniques such as N-Gram, TF-IDF and Word2vec used for the feature representation of text were discussed. Furthermore, during the EDA, it was observed that the labelled data has class imbalance problem and thus SMOTE which is an oversampling approach was applied to rebalance the dataset. This will help the models against being biased towards the majority class. The supervised classification models adopted were also discussed. Lastly, the third component of the TSBS framework is the aspect extraction for sentiment analysis. Statistical topic modelling methods such as latent semantic indexing (LSI), latent Dirichlet allocation (LDA), and hierarchical Dirichlet process (HDP) were adopted since both explicit and implicit aspects are needed to form the banking themes in this context. Therefore, subsequent chapters present the result of the components.

CHAPTER 5: Lexicon Based Approach

5.0 Introduction

In the previous chapter, lexicon-based approach was identified as one of the components of TSBS framework and since one of the literature gaps identified was lack of sentiment lexicon to suit the Nigeria bank domain. This led to the development of "SentiLeye" a novel lexicon algorithm. The methodology chapter detailed the approach to the sentiment lexicon generation of "SentiLeye". Therefore, this chapter presents the lexicon statistics and performance evaluation result.

5.1 Lexicon Statistics

Language	Count
English	3734
Pidgin-English	352
Slang	51
Emoticon	116
Emotion	185
Booster	22
Total	**4460**

Table 5.1: Count of SentiLeye WordList

Table 5.1 above presents the statistics of the wordlist generated. It can be seen from the table that the Pidgin words are limited compared to the English wordlist. This is due to lack in lexical resources for pidgin English. Thus, one of the areas which this study contributes significantly. This study adopted the emoticons, emotion, and slang wordlist from the Sentistrength (Thelwall et al. 2010) lexicon due to the performance of the lexicon with social media data. It is worth noting that the slang, booster and 12 emoticon list are not labelled. Table 5.2 below shows the distribution of the words by their scores.

Polarity score	Count	Example
-2	1482	Dormant
-1	1649	Lost
1	891	Wise
2	353	Sure

Table 5.2: Distribution of SentiLeye wordlist by scores

5.2 Result of Lexicon Evaluation

In text classification problems, one of the most well-known approaches to determine model performance is using evaluation metrics such as precision, accuracy, recall and f1-score (see section 4.12 in previous chapter for details). In this study, three annotators with banking expertise who are native speakers of Pidgin and English were employed to classify tweets into positive, negative, and neutral. A random sample of one thousand tweets were selected from the text data for this purpose and manually coded by human annotators. This study ensures all annotators agree on the class of each of the tweet for reliability using majority voting rule for any disagreement (please see links to labelled set in appendix C). The class statistics of the randomly selected one thousand tweets are provided in table 5.3 below.

Sentiment Class	Count
Positive	68
Neutral	236
Negative	696

Table 5.3: Class distribution of evaluation dataset

The coded classes were compared against the prediction of the lexicons to know how well the lexicons have performed. This study compared SentiLeye to lexicons such as Afinn, Sentistrength, Bing, NRC, Sentiwordnet, TextBlob, and Vader. This is because these lexicons have been shown to perform well across domains and validated with social media data. In addition, these lexicons are free to use for academic research. There are several other lexicons such as LIWC which this study could have considered. However, they are not free to use. The study was able to compare the performance of these lexicons against proposed SentiLeye lexicon. Table 5.4 below shows the performance of the lexicons and that proposed by this study.

Metrics/Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SentiLeye	**77**	**86**	**77**	**77**
AFINN	38	71	38	45
TextBlob	41	71	41	44
Vader	35	74	35	43
SentiStrength	61	69	61	64
NRC	32	69	32	39
Bing	50	71	50	55
SentiWordNet	39	66	39	46

Table 5.4: Comparison of lexicon algorithms

From table 5.4 above, the result shows SentiLeye performed better than other lexicons using all the metrics with accuracy and F1-score of 77%. To have a clearer understanding of the lexicon comparison, figure 5.1 plot is produced to visualize the lexicons' performances.

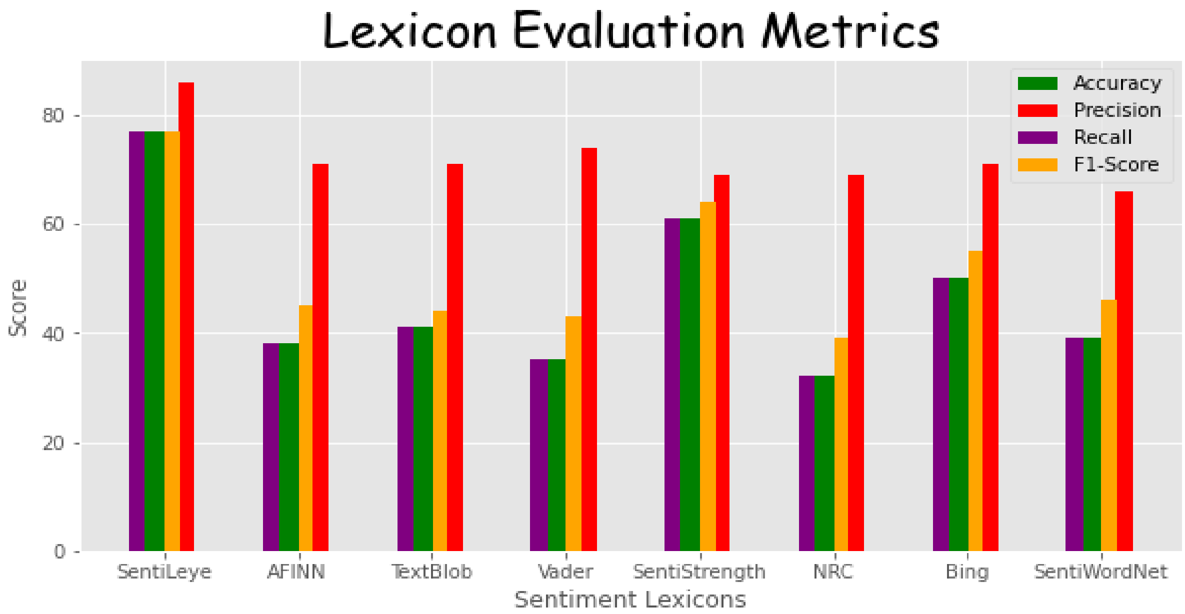


Figure 5.1: Evaluation of sentiment lexicons

As shown in figure 5.1 above, **SentiLeye and Sentistrength** are the top performing lexicons in this context. These lexicons showed good performance across all metrics. Their performance is attributed to the ability to detect bank domain terms and opinionated factual words. Secondly, their algorithms handled negation well. Another lexicon that performed well is Bing**.** Bing produced an accuracy of 50% and F1-score of 55%. The lexicon algorithm has good coverage of social media terms, malformed words, and opinionated factual words like failed, complain. However, did not handle negation and struggled with contextual words like deduct, debited. Others like NRCLex, Afinn, TextBlob and Vader performed poorly. This is largely due to these lexicons inability to detect bank context words. For example, *"Whats happening na another failed transaction this morning debited and yet to receive reversal for the previous one would discontinue to use this account"*. In addition, the algorithms of the lexicons were checked, and it was observed that the implementation **does not handle negation.** For example, *"not able to send dm please rectify issue someone made transaction to my account since yesterday received alert longer funny referred twitter send dm yet no one responded"* has been misclassified by NRC, Afinn, Vader as positive, and Textblob as neutral. A basic example to demonstrate these popular lexicons are limited in terms of handling negation is shown in figure 5.2 below using Afinn sentiment classifier. In the example, the bigram "not good" was classed as positive by Afinn.

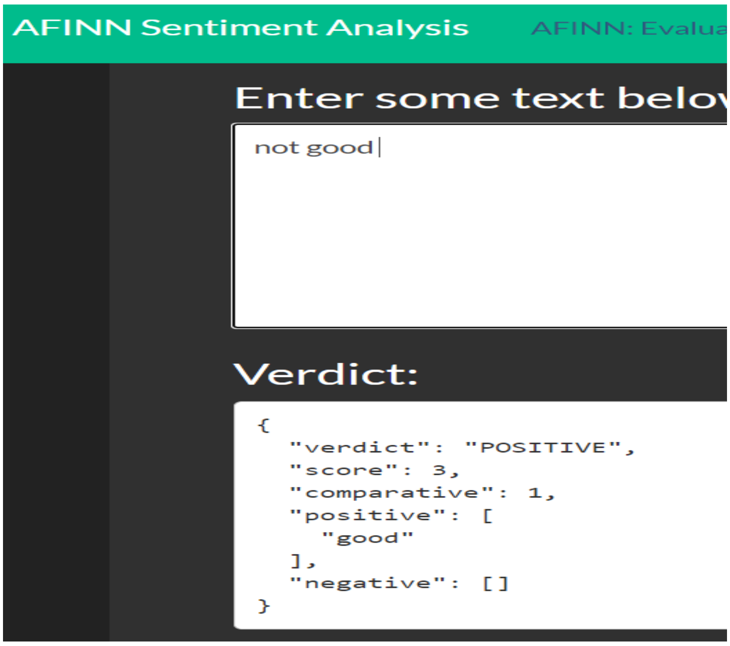


Figure 5.2: Afinn Sentiment classifier

To understand the performance of the lexicons further in terms of predicted classes. This study produced a plot of the correct predictions to show how many positive, negative, and neutral classes were correctly classified by the lexicon algorithms.

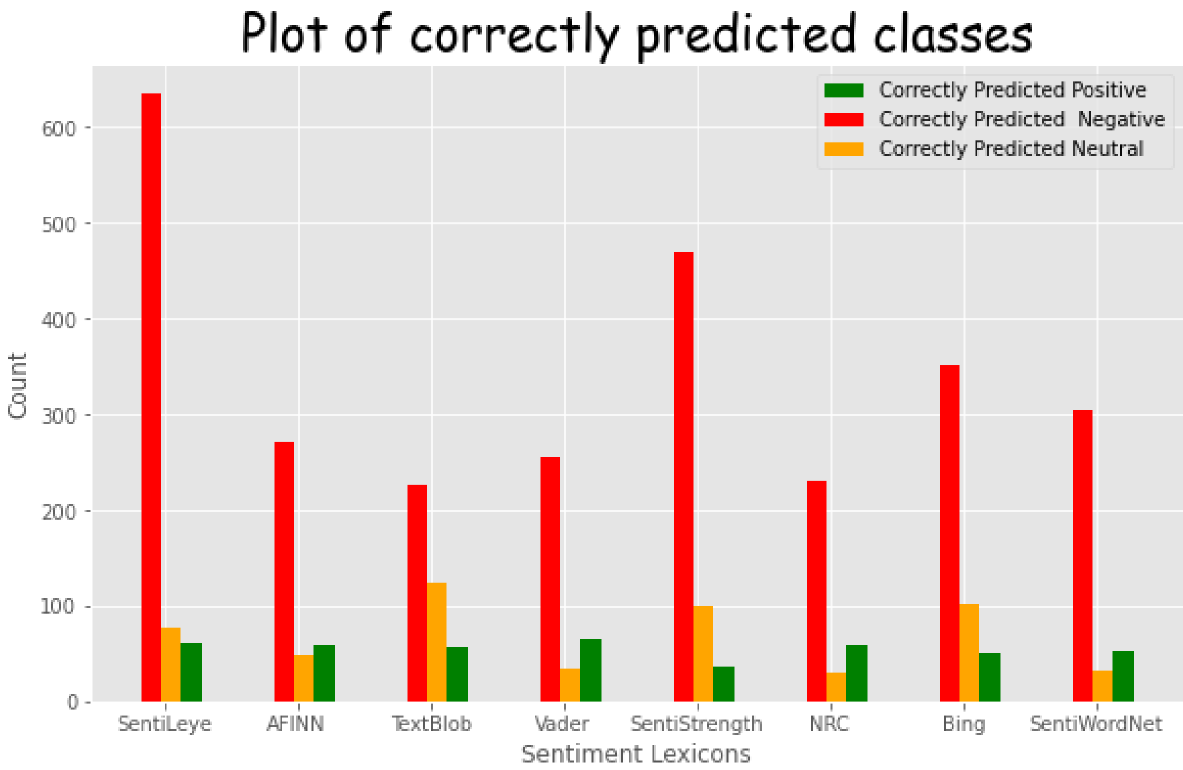


Figure 5.3: Plot of correctly predicted sentiment class

From figure 5.3 above, TextBlob produced the highest correctly classified neutral Tweets. Seconded by Bing and Sentistrength. Compared to others, SentiLeye performed averagely in classifying the neutral Tweets. For the positive tweets, SentiStrength struggled to predict the positive classes correctly compared to other lexicons. Interestingly, Vader produced the highest correctly classified positive tweets and seconded by SentiLeye. In contrary, SentiLeye and Sentistrength produced the highest correctly classified negative tweet. Others like TextBlob and NRC struggled significantly. In summary, considering the importance of classifying negative tweet correctly in the service industry such as banking. SentiStrength showed an able to compete performance. However, SentiLeye produced the best performance.

During the experimentation, lexicons such as Sentistrength, Afinn, Bing and NRC were updated to take the Pidgin terms into consideration. A slight improvement in terms of accuracy (approximately 2%) was observed. This was expected as the Tweets were of Pidgin and English. With respect to the second research question (RQ2), it was found during the experimentation that local terms & context-based words significantly influence the performance of the lexicons. An explanation for this is that lexicon-based approach relies strongly on their wordlist, and thus, if word does not exist the lexicon cannot capture the semantic orientation of such word. Some of the lexicon algorithms could have performed much better if the context words were considered. For example, the term 'debited' was coded into Afinn corpus as a neutral word. Surprisingly, Sentistrength coded 'debited' as negative which evidenced why the algorithm produced a better result. To conclude, the lexicons need to consider the context words to improve on performances of lexicon approach to SA. In comparison, general-purpose lexicons like Bing, SentiWordNet, Vader, TextBlob & AFINN are less accurate in narrowed context compared to SentiLeye. Thus, are limited because domain dependent words are not considered at all. For example, Tweet like, *"the thing swallow atm card no gimme money received debit alert like min ago"* is misclassified because those lexicons are developed using general lexical knowledge. This finding is consistent with that of Kevin-labille et al. (2017) that showed domain specific dictionaries perform better than the general-purpose dictionaries for SA, and specifically in financial context shown by Palmer et al. (2020).

5.3 Chapter Summary

This chapter set out with the aim to develop lexicon to capture the semantic orientation of Pidgin and English words in the Nigeria banking context. The findings suggest general purpose lexicons suffer in context such as banking because they are developed from general lexical knowledge and thus, there is need to update or recreate lexicon to suit domain terms. Secondly, there is need for existing lexicons to update their wordlist with opinionated factual words especially when working with service industry dataset such as banking. This finding is consistent with Liu, (2013) that buttressed the need for creation of opinionated factual words. Thirdly, the lexicons compared show there is need for lexical algorithms to consider negation. This is consistent with the findings of studies like Mukherjee et al. 2021; Gupta & Joshi 2021; Singh 2021; Garg & Subrahmanyam, 2021 that stated negation handling improves sentiment classification performance. Lastly, the finding shows there is need to create wordlist for non-English text where needed as this improves the performance of SA lexicons. This finding is consistent with that of Kaity & Balakrishnan (2019) that stated creation of non-English lexicons significantly enhances the performance of SA in domains where non-English terms exist. The main contribution of this chapter is SentiLeye, a novel lexicon algorithm developed to capture Nigeria Pidgin English, English, and localized words in the banking domain. In addition, this study contributes by providing labelled dataset which is suitable to validate lexicons in Nigeria banking context. Lastly, this study provides sentiment Pidgin wordlist to encourage further research in Pidgin language processing (please see links to dataset in Appendix C). In this chapter, the significance of the research finding demonstrates the need for future researchers to put domain terms, opinionated factual words, negation, and language into consideration when proposing a lexicon algorithm.

The implementation is useful for banks to monitor their products and service and most importantly, understand how their customers' feel. However, the corpus used in this study is limited in terms of word coverage and thus can be expanded in future research work. It is worth acknowledging that the implementation is limited to banking because the algorithm has not been validated in other domains. Secondly, the SentiLeye lexicon is limited to words in English and Pidgin English and is yet validated for words in Nigeria local languages such as Yoruba, Igbo and Hausa. Unfortunately, the local words were found in the tweets. Therefore, for future work, the implementation can be improved by applying some knowledge transfer techniques to suit other domains. In reviewing the literature, inconsistent performance has been reported on the performance of machine learning models versus lexicon-based approaches to sentiment analysis. It will thus be interesting to know which of these techniques perform better. In the next chapter, this study discusses the result of the machine learning algorithms to investigate their performance in this context.

CHAPTER 6: Machine Learning Approach

6.0 Introduction

This chapter presents the result of the machine learning models applied to the bank tweets. The result will be used to understand the efficiency of the methods employed including available alternatives. The performance of the sentiment classification models will be assessed based on evaluation metrics such as precision, recall, accuracy, and F1-score. Thus, this will help in answering the research questions and fulfil the goal of this project as stated in chapter one.

6.1 Result of the Machine Learning Evaluation

The rebalanced labelled dataset was used to train the classification models. The machine learning models employed in the study are K-Nearest Neighbour, Naïve Bayes, Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting. The test set was used to evaluate how well the models have performed by comparing the model prediction to the actual labels. Thus, the performance of these models is reported in terms of accuracy, precision, recall, and the balanced F1-score. Most importantly, the report presents the result of both vector space models (N-gram and TF-IDF) for each of the classification models.

		N (1-3) --- Gram				TF-IDF			
	Classes	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
KNN	Negative	0.38	0.73	0.50	0.54	0.35	0.82	0.49	0.41
	Neutral	0.89	0.46	0.60		0.97	0.23	0.37	
	Positive	0.24	0.73	0.37		0.18	0.82	0.29	
NB	Negative	0.52	0.88	0.65	0.73	0.53	0.86	0.66	0.72
	Neutral	0.93	0.69	0.79		0.94	0.67	0.78	
	Positive	0.49	0.69	0.57		0.42	0.76	0.54	
LR	Negative	0.59	0.70	0.64	0.76	**0.60**	**0.76**	**0.67**	**0.78**
	Neutral	0.88	0.79	0.83		**0.90**	**0.80**	**0.85**	
	Positive	0.49	0.67	0.57		**0.54**	**0.70**	**0.61**	
SVM	Negative	0.55	0.57	0.56	0.75	**0.57**	**0.77**	**0.66**	**0.77**
	Neutral	0.83	0.82	0.83		**0.90**	**0.78**	**0.84**	
	Positive	0.57	0.57	0.57		**0.58**	**0.70**	**0.63**	
RF	Negative	0.70	0.40	0.51	0.78	**0.68**	**0.54**	**0.60**	**0.79**
	Neutral	0.80	0.93	0.86		**0.83**	**0.90**	**0.86**	
	Positive	0.72	0.49	0.59		**0.64**	**0.57**	**0.60**	
GB	Negative	0.52	0.42	0.46	0.73	0.51	0.54	0.53	0.73
	Neutral	0.79	0.85	0.82		0.82	0.81	0.81	
	Positive	0.59	0.49	0.54		0.55	0.52	0.53	

Table 6.1: Performance result for the sentiment classification models.

Table 6.1 above presents the performance result of the statistical and machine learning models. It is worth noting that the F1 score reported are for individual classes (balanced F1 score). To calculate the macro F1 score, we divide the addition of the three F1 scores by the total number of classes. For example, in the case of LR, the macro F1 score is (0.67 +0.85 + 0.61)/3 = 0.71. This means LR achieved macro F1-score of 71%. However, the individual class F1- scores were produced to understand the model performance towards individual classes. For example, considering the negative tweet, the F1-score provides a balance between the precision (which gives a clue on the number of false positive) and the recall (which indicates the percentage of correct negative tweets that are selected). The higher the precision indicates how low is the false positive and the higher the recall indicates how low the false negative.

In general, the result shows K-Nearest Neighbour (KNN) performed poorly especially in terms of precision. Comparing both vector models, KNN yielded a better result with N-gram features than the TF-IDF. However, the model performed poor generally. The gradient boosting (GB) model showed a better performance. Considering the performance of other models, gradient boosting is not the best. However, GB with TF-IDF features produced a fair result. Naïve Bayes (NB) performed well with both vector space models especially in terms of recall. However, the model struggled to predict positive classes precisely as this was below 50% in both cases. This means there were a lot of positive tweets that the NB has falsely classified as negative and neutral. Comparing both vector models, NB yielded a better result with N-gram features than TF-IDF. In this study, the well performed models are Logistic regression (LR), Support vector machine (SVM), and Random Forest (RF). Logistic regression performed better with TF-IDF model especially in terms of recall. Though, it was observed the result for positive class is slightly lower than other classes, but the model performed well in all the classes using all the metrics stated. Similarly, random forest performed better with TF-IDF model especially in terms of precision. The N-gram model result was not great in terms of recall with both positive and negative classes having a performance lower than 50%. RF produced a lower recall for positive and negative when compared to LR and SVM results. However, the model has performed well. Lastly, SVM performed well with both vector space models across the classes. The TF-IDF model produced a better result. The classifier showed its robustness across the classes with the best recall for negative tweets when compared with LR and RF. This is important in this context because the banks are more likely to attend to, and/or inform their decision making with more emphasis on the negative tweets.

The findings suggest most of the models except SVM and RF struggled classifying positive classes but have good performance towards the neutral and negative classes. This might be due to the number of positive instances in the training set. Using majority rule, the TF-IDF produced better result than the N-gram model especially with logistic regression, support vector machine and random forest. This study considers using the F1-score to decide on the best performing models since F1-measure provides the balance between the precision and recall. The F1 scores of logistic regression (positive = 0.61, negative = 0.67, neutral =0.85) and support vector machine (positive = 0.63, negative = 0.66, neutral= 0.84) showed the two classification algorithms are the best performed learning models in this context. The models have good performance across the classes. Considering our interest in the negative class, the F1 score of the negative class were also good. To choose one of these models to build a system or for recommendation purpose, the parameters of the models will be tuned (hyperparameter) for optimal performance. Beforehand, the study will thus, compare the result of the traditional ML models with the deep learning models inquest of a better model.

Table 6.2 below presents the result of deep learning methods employed. The result shows the performance of Word2Vec embedding trained with bank dataset and the models. This was compared with deep learning models when used without Word2vec embedded layer.

		Neural Network (Word Embedding)				Word2Vec (Embedding) + Neural Networks			
	Classes	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
CNN	Negative	0.56	0.59	0.57	0.75	0.46	0.33	0.38	0.71
	Neutral	0.82	0.84	0.83		0.76	0.87	0.81	
	Positive	0.42	0.19	0.26		0.86	0.08	0.15	
RNN	Negative	0.53	0.58	0.53	0.71	0.10	0.10	0.10	0.71
	Neutral	0.81	0.79	0.80		0.71	0.99	0.83	
	Positive	0.30	0.40	0.34		0.15	0.11	0.13	
LSTM	Negative	**0.65**	**0.59**	**0.62**	**0.79**	0.62	0.24	0.35	0.74
	Neutral	**0.85**	**0.87**	**0.86**		0.76	0.94	0.84	
	Positive	**0.56**	**0.59**	**0.58**		0.83	0.20	0.32	
Bi-LSTM	Negative	**0.57**	**0.66**	**0.61**	**0.77**	0.58	0.18	0.27	0.73
	Neutral	**0.86**	**0.82**	**0.84**		0.75	0.94	0.84	
	Positive	**0.62**	**0.51**	**0.56**		0.80	0.11	0.19	

Table 6.2: Performance result for the deep learning models.

The result presented in Table 6.3 above shows that the deep learning models performed poorly when the Word2Vec embedding layer was added into the model networks. The idea was to build a more context informed network to improve the performance of these models. The approach had been used successfully in the studies of Kim (2014) and Ruder et al. (2016). Unfortunately, the models performed poor with Word2Vec layer which is not in agreement with the result of Wen et al. (2018).

Surprisingly, the models performed better without the Word2Vec embedded layer. It was observed that the deep learning required more time to train and were computational expensive. Amongst, the deep learning models used on their own, the long short-term memory (LSTM), and bi-directional long short-term memory (BI-LSTM) models performed well. While CNN and RNN did not perform greatly specifically with the positive class. Both models performed fairly with the neutral and negative classes but poorly towards the positive. This is similar to the result of the traditional ML models and thus suggest the poor performance with positive class is due to the number of positive instances in the training set. It is worth noting that SMOTE was applied to rebalance the dataset. The SMOTE approach rebalanced the class distribution with synthetic samples initiated by randomised interpolation between several minority instances that lie together. The approach utilises nearest neighbour words to instantiate the samples and thus, the new samples created might overlap with features of other classes which then becomes a noise to the models.

LSTM and BI-LSTM had a fairly balanced performance across the classes. The LSTM has the best result of all. Thus, with the aim to obtain optimal parameters; the batch size, epochs, and network were manipulated. The results for the deep learning models were changing at a range of ±0.015 which does not perform upto the ML models (SVM and LR). The findings suggest the context-based vector space models is a great idea to understand the difference in positive or negative sentiment context but could not outperform the frequency-based vector models (traditional methods) used with the machine learning models. This might be due to the syntax of the dataset. The result is consistent with the findings of Abercrombie (2021) that stated the use of traditional methods for text representation does outperform the word embeddings that relies on neural network or more complex model. In this case, the tweets are short and mostly objective statement which was shown in figure 4.8. Most notably, the positive tweets were the shortest in terms of length per tweets which significantly contribute to the performance of the models. The deep learning models used, are sequential models. These models tend to rely on more words in a sentence to grab the semantic orientation. For example, the Bi-LSTM aims to capture the context of the sentence in a forward and backward sequence. The model reads input from the first term to the last as a direction, and the last term to the first in the second direction. Such models perform better with longer sentences. Unfortunately, the data applied were short tweet. This might have contributed to why deep learning approach were outperformed by the traditional machine learning models with TF-IDF features. This result is inconsistent with the study of Wang et al. (2020) that showed LSTM performed better than traditional machine learning models with financial article text.

6.2 Parameter Tuning & Feature Selection for ML Models

The performance evaluation result shown in Table 6.1 showed the logistic regression and support vector machine are the best performed machine learning models. These models were considered the best using the accuracy and F1-scores. These models outperformed other models including the deep learning models. However, to enhance the performance of all the models. This study tuned the parameter of all the models using Scikitlearn grid search package in python. This helps detect the best parameter for optimal performance of the model. Majority of the models did not show significant improvement. However, LR produced results which the values changed at the range of ±0.01. To have a fair comparison, this study utilised different subset of the features with LR for improvement. Unfortunately, the result remained the same or poorer. SVM parameter tuning resulted in an improved model with accuracy increment of 0.05 as shown in Table 6.3 below.

Model	Classes	Precision	Recall	F1-score	Accuracy
SVM	Negative	0.75	0.56	0.64	0.82
	Neutral	0.83	0.93	0.88	
	Positive	0.80	0.52	0.63	

Table 6.3: SVM tuned parameter result.

The parameters of the above result are kernel = rbf, C = 10, and decision function shape = ovr. Whilst, tuning the parameters, it was observed that SVM produced a better score in terms of recall when the kernel function is set to be linear. This result aligns with the recall score of linear classifiers, Naïve Bayes which also produced a better recall. When the non-linear kernel function is used, SVM produces a better precision. The non-linear SVM produced an improved F1-score (neutral) and accuracy as shown above. Thus, this study ascertains support vector machine as a robust sentiment classifier in the banking context. This result is consistent with the findings of Adamu et al. 2021; Xing et al. 2020; Chory et al. 2018; Istia & Purnomo, 2018; Al-Smadi et al. 2018; Alayba et al. 2017; Tan & Zhang, 2008 that showed

SVM as the state-of-the-art method for sentiment analysis. To answer the research question in this study.

RQ3. Can Machine or Deep learning models perform as an off-the-shelf sentiment classification method in the banking domain?

There is no off-the-shelf sentiment classification model until validated. This study has therefore filled that gap by validating support vector machine (SVM) as an able to compete model in the banking context. However, it is worth noting that SentiLeye lexicon also provides an easy-to-understand sentiment analysis technique. In chapter 5, results showed SentiLeye outperformed other lexicons due to domain terms, opinionated-objective words, negation, and language which was put into consideration during the development. This is an area where the lexicon has the advantage over SVM, this is because machine learning models are black box, and they are difficult to understand because they do not produce human interpretable models (Liu, 2015). This means when there is a problem. It is difficult to detect the problem and solution.

With respect to RQ3, SVM can be used as an off-the-shelf sentiment classification model across different domains. This is because it is evidenced in literature that SVM has great or better performance in different domains such as hotel review (Addi et al. 2020; Said & Muqrashi, 2020), education (Guo et al. 2020), health (Rasool et al. 2020; Adamu et al. 2021), movie review (Sharma & Dey, 2012), stock selection (Liu et al. 2020) and stock price movement (Sagala et al. 2020). SVM has also been shown to perform better than other machine learning models (Bouchlaghem et al. 2016; Yu et al. 2020; Addi et al. 2020) and deep learning models such as CNN, RNN (Alayba et al. 2017; Al-Smadi et al. 2018), Bi-LSTM (Xing et al. 2020) and LSTM (Salehin et al. 2020). Xing et al. (2020) conducted their study in the financial context, and they showed SVM performed equally well as BERT and performed better than Bi-LSTM. In summary, the performance of SVM in this study has shown the model can be used as an off-the-shelf method in the Nigeria banking context and as baseline model for future work.

Contribution:

- *This study contributes by conducting an extensive comparison of sentiment classification models and thus, validates support vector machine (SVM) as the sentiment classification model with an able to compete performance in the Nigeria banking context.*

6.3 Causes of Misclassification

To improve on the sentiment classification and answer the fourth research question (RQ4), this study will perform a qualitative analysis of the predicted tweets. This is done to understand the reasons behind the misclassification of tweets. The test set were manually examined by comparing the result of the correct 'manual' label against the model 'predicted' label. Thereafter, examine the wrongly predicted tweets. Therefore, addresses the fourth research question below.

RQ4. What are the major causes of misclassification in this context?

6.3.1 Use of Localized Language

The bank dataset used in this study are of Nigeria customers whose official language is English and unofficial widely spoken language is Pidgin English. The tweets were majorly of these two languages. However, there were observed cases when customers tweeted in their local language. In Nigeria, there are three main local languages namely, Yoruba, Hausa, and Igbo language. It was observed that customers expressed their frustration with some of these languages since that is their primary language. For example, *@XY bank I go show una pepper since last night dm and no respond send email get generic response on the transaction hey see tweeting event eti ya werey*. This is a strong negative tweet which was misclassified as neutral. The case here is that the models were unable to identify the presence of the Yoruba language in the tweet. Another example is *@XY bank.... ode ni eleyi walahi cleared balance available balance na d same bros sharap.* The tweet was misclassified as neutral. This is also a strong negative tweet which demonstrates according to the customer how the bank staff cannot distinguish between cleared balance and available balance. Thus, resulted in a high tensed negative tweet. In summary, the model is unfamiliar with the local language and thus misclassifies. Since the language was not frequently used in the training dataset like English and pidgin English, the model did not learn enough to classify the tweet appropriately. Thus, a straightforward solution is to train the model with sufficient tweets that contains local terms. Another method is to enhance the classification model with annotated lexicon of the local languages. This approach was shown in section 5.2 where it was observed that inclusion of Pidgin terms significantly influences the performance of the lexicon.

6.3.2 Opinionated-Objective Sentences

Liu (2014) identified opinionated-objective sentences as one of the difficulties in a sentiment analysis system. This kind of objective sentences are factual statement that expresses desirable or undesirable opinion about the target aspect. They are fact implied opinions and are often classed as neutral by classification models. This problem was observed as one of the challenges the model faced in this context. For example, *@XY bank…. request to close account with account number charging card maintenance fee therefore not interested in banking anymore*. The tweet was misclassified as neutral. Another example, *@XYZ bank …bought glo airtime via mobile app yesterday debited till this moment glo line yet credited*. These kinds of tweets are strong opinionated objective statement which should be classed as negative. Liu (2014) emphasized the need for sentiment analysis system to classify opinionated-objective sentences as either positive or negative sentiment especially in the business context. Unfortunately, it was observed the model misclassified such tweets as neutral and thus, creates room for improvement. Again, this is another area, the lexicon "SentiLeye" proposed in chapter 4 of this study performed well.

6.4 Further Experimentation: Hybrid

An attempt to improve the sentiment classification model based on misclassification factors identified in section 6.3 above, a hybrid approach was employed. The SentiLeye lexicon words were used to enhance the classification model. This approach was adopted from the work of Lin & He (2009). The approach was implemented by using the word list and their polarity score as additional features for the classification model. This implies the word level polarity scores were feed in addition to the tweet level polarity scores to train the model. Unfortunately, this approach produced poorer result as shown in appendix. The poor performance was due to the fact that the approach created noise in the model as the algorithm did not learn any new feature. The model needs new features in Yoruba, Igbo and Hausa terms to learn from. For example, this tweet, "*@XY bank I go show una pepper since last night dm and no respond send email get generic response on the transaction hey see tweeting event eti ya werey"*. In this tweet, "*eti ya werey*" is a Nigeria local language from the Yoruba tribe. Unfortunately, these terms are currently unavailable in the SentiLeye lexicon. Thus, future work will be improving the lexicon with the local words (in Yoruba, Igbo and Hausa) so that this can be used as additional features to the models. This aligns with the need to increase word list in lexicons as social media words are constantly changing and contains new malformed words. In addition, this aligns with the need to create more non-English lexicons.

6.5 Chapter Summary

This chapter presented the result of the sentiment classification models. The performance result from the comparative study of the classification models validated support vector machine (SVM) with accuracy of 82% as the most performed classification model in the banking context. The findings showed the classifiers struggled to classify tweets with local language and opinionated objective sentence. This might be due to the low presence of local language in the training set. The problem of opinionated objective sentence remains a concern for machine learning classifiers. This creates room for future research work on how the model can be improved.

This study is beneficial to bank decision makers and will help them understand their customers' attitude. The sentiment analysis demonstrates the customers feelings on what they learnt through their banking experience. Therefore, with measurable attributes of positive or negative, the banks were able to know if their customers are happy or unhappy. This aligns with the definitions of customer attitude by Szmigin & Piacentini (2018); Oskamp & Schultz (2005). However, the banks are unsure of which banking product or service the sentiment polarities refer to. Liu (2014) buttressed the weakness of supervised machine learning as unable to determine to which target the opinion refers to in a tweet-level sentiment analysis. This is because the features used are target independent. Therefore, one solution is to perform aspect extraction to investigate the sentiment polarities of the customers' tweets. This will help to gain in-depth insight because it is more beneficial for the banks to understand to what target the customers have tweeted. This will help banks generate purposeful and actionable insight towards the bank product and services. Therefore, this study will perform aspect extraction for sentiment analysis in the next chapter.

CHAPTER 7: Topic Modelling

7.0 Introduction

This chapter focuses on aspect extraction to generate actionable insights from the bank customers' sentiment. This is the third component of the TSBS framework. This study considers the sentiment classification result presented in the previous chapter not enough to generate in-depth insight in the customer tweets. This is because the result was only able to provide an overall Tweet-level sentiment analysis. A more functional system where these sentiment polarities can be attributed toward topics will be more valuable in the banking industry or academia. The aim is to extract both explicit and implicit topics to understand what part of the banking product or services customers tweeted about. This chapter details the processes and result of the techniques employed and thus, provides a comprehensive process to investigate customers' attitude using these topics. This will help assess customers' attitude towards the sentiment polarities and thus help in answering the research question (RQ1) stated in chapter one.

7.1 Result of the Topic Modelling

During experimentation, the initial result contained irrelevant terms. This is due to the nature of the dataset. As mentioned in the literature review, Asghari et al. (2020) stated social media data is very difficult for topic modelling because the data is noisy and contains a lot of constantly changing informal words. They further explained a direct use of algorithms will not perform well with social media data. Therefore, in this study there was need to prune the corpus. This helped to get rid of the noisy data which are less important. The corpus was pruned using a frequency threshold of 10 and the irrelevant terms were removed. A background (banking) knowledge and adjusted parts of speech (POS) were employed to enhance in pruning the terms to important banking terms. The POS used are singular noun ('NN'), proper noun plural ('NNPS'), noun plural ("NNS"), proper noun singular ("NNP"), adjective ("JJ") and adverb ("RB"). These are all available in the NLTK module. After multiple iteration, it was necessary to get the optimal number of topics for the models. This was determined using their coherence scores and topic quality (interpretability). The N (number of topics) was set at the range of 5 to 250 for all the topic models. For example, the LDA model as shown in figure 7.1 below produced coherence score at a close range of 0.3 to 0.65. With coherence score the higher the better, however, topic quality is also considered. Thus, determining the number of topics relied on interpretability of the topics. Based on the latter, the number of topics N was set at 10 because the model produced 10 interpretable topics.

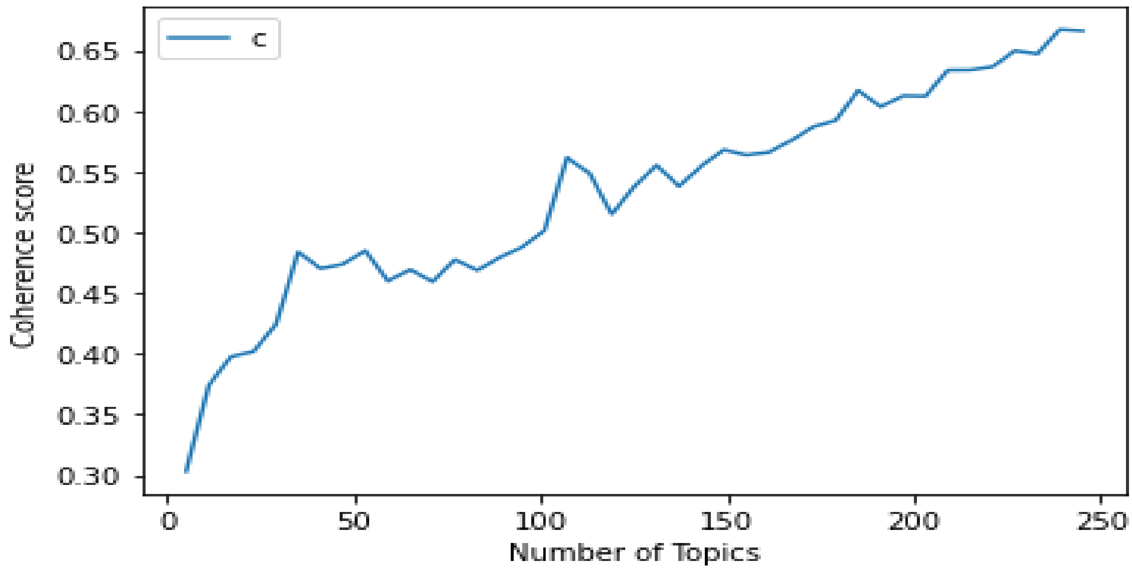


Figure 7.1: Plot of Topics versus Coherence score

Finally, the unsupervised models were iterated to classify 20 terms per each topic and 10 number of topics. Figure 7.2a – 7.2d below presents the result of the unsupervised topic models.

Index	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Term1	customer	job	help	loan	card	business	dm	line	always	credit_card
Term2	cenbank_cent…	morgan	access	help	charge	lagos	help	phone	poor	citi
Term3	branch	free	account	veilleur	never	great	detail	support	stop	bill
Term4	time	min	money	bbnaija	new	school	thank	ng	financial	state
Term5	month	company	transaction	cenbank_cent…	use	news	number	mobile	ng_union	debt
Term6	people	una	please	kindly_follow	open	monetizable_…	twitter	mtnng_mtn	cool	payment
Term7	guy	fossil_fuel	transfer	day	year	world	kindly	hope	youfirst	chase
Term8	atm	ready	pls	send_direct	debit	additional_i…	please	food	sport	omo_iya
Term9	even	hey_investme…	yet	officialefcc…	later	soon	complaint	consumer	business	mastercard
Term10	week	gas_coal	day	message_deta…	small	let	sorry	government	service	nthank
Term11	still	financing_oil	number	monthly	fee	part	send	international	finance	ng
Term12	day	earth_stop	kindly	thank	pay	good	tweet	buhari	internet_ban…	future
Term13	service	company_nofo…	yesterday	house	dey	best	transaction	sort	premiercooln…	true
Term14	call	nofossilfuel…	back	enable_assist	top	opportunity	handle	whatsapp	mind	alat
Term15	complaint	fargo_jpmorg…	today	salary	see	today	account	explanation	tell	abuja
Term16	nothing	change	issue	ph	right	deposit	information	enable	thing	pay
Term17	customer_care	actually	alert	advise_appro…	also	wish	send_dm	lagos	daily	india
Term18	money	head	need	heritagebank…	month	thanks	channel	foundation	single	atm_machine
Term19	customer_ser…	woman	airtime	eyin_ole	nyou	trump	phone_number	industry	fast	child
Term20	thing	country	fund	force	far	event	message	ubaat	vista_intl	product

Figure 7.2a: Latent Dirichlet Allocation (LDA) result

The LDA result in figure 7.2a above shows the technique produced a good result. The terms in the LDA topics are mostly banking terms. For example, customer_service, atm_machine, customer_care, complaint, account, transaction, transfer, mobile, branch, master_card, credit_card and internet_banking. These terms demonstrate banking channels and services. Interestingly, there are Pidgin English terms in the LDA topics. For example, alat. This term indicates bank alert. A surprising result is the presence of local terms. For example, eyin_ole. This term originates from the Yoruba tribe of Nigeria, but it is a term generally used by all tribes. The term indicates words like scam, fraud, or thieves. There are other unexpected terms such as Trump and India. However, in general, the LDA topics are very reasonable.

topics_df_mallet - DataFrame

Index	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Term1	access	loan	card	year	transaction	time	account	money	number	branch
Term2	guy	bankng	account	payment	day	people	transfer	back	dm	service
Term3	issue	staff	pls	business	cenbank_cent…	work	complaint	atm	detail	customer
Term4	problem	send	month	good	today	tweet	yesterday	credit	airtime	response
Term5	customer	code	debit	ng	refund	close	till	cash	app	call
Term6	week	cbn	naira	pay	central	network	morning	person	line	thing
Term7	dear	platform	alert	banking	fund	amount	balance	union	phone	customer_care
Term8	diamond	handle	charge	open	reversal	life	mobile_app	withdrawal	nplease	hour
Term9	email	officialefcc…	kindly	friday	successful	company	kindly	good_morning	mobile	customer_service
Term10	people	case	online	free	reverse	na	complain	withdraw	support	mail
Term11	system	world	message	experience	debited	god	asap	last_week	twitter	ubacfc
Term12	useless	message_deta…	transaction	credit_card	kindly	school	urgent	dey	recharge	long
Term13	care	information	acct	veilleur	week	fraud	wrong	polarisbankl…	team	lagos
Term14	agent	official	error	citi	data	financial	immediately	news	date	form
Term15	respond	enable_assist	june	give	time	ussd	bad	yet_receive	assist	ur
Term16	understand	morgan	change	nthe	wait	nyou	make	report	top	request
Term17	mtn	link	fee	state	monetizable_…	small	statement	debit_alert	check	office
Term18	kind	country	sum	big	additional_i…	lot	receive	matter	mtnng_mtn	job
Term19	feedback	dollar	una	show	july	hope	hour	urgently	amount	hr
Term20	brother	india	update	happy	ago	acc	recipient	bbnaija	nkindly	monday

Figure 7.2b: Latent Dirichlet Allocation (MALLET) result

The LDA-Mallet (figure 7.2b) produced a good result too. This is very similar to the original LDA result reported earlier. There are lots of banking terms observed. For example, customer_service, atm, customer_care, complaint, account, transaction, transfer, balance, branch, mobile, and credit_card. There are banking terms observed which were not seen in the original LDA topics such as ussd, code, app, mobile_app, debit_alert, withdrawal, network, fraud and cbn. These terms indicate banking channels, transaction, service, and service network. There are more Pidgin English terms in the Mallet topics. For example, na, and god. It is worth noting that the Pidgin English terms observed are irrelevant in this context. Surprisingly, there is no local term present. In summary, the Mallet result is good and presents topics which can compete with the original LDA topics.

topics_df_lsi - DataFrame

Index	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Term1	help	account	money	access	transaction	please	day	card	dm	today
Term2	access	access	account	help	help	transaction	transaction	day	number	refund
Term3	account	money	help	account	money	pls	card	debit	day	people
Term4	money	help	number	transaction	please	card	time	cenbank_cent…	card	news
Term5	transaction	please	access	please	account	number	guy	please	guy	day
Term6	please	number	back	money	card	day	money	guy	people	business
Term7	number	transaction	day	dm	day	dm	cenbank_central	dm	issue	transaction
Term8	day	transfer	guy	kindly	cenbank_cent…	help	branch	people	time	number
Term9	dm	yet	cenbank_cent…	number	yet	account	people	branch	nplease	citizen
Term10	transfer	kindly	dm	thank	kindly	transfer	still	pls	please	transfer
Term11	yet	back	refund	detail	time	money	customer	month	detail	global
Term12	guy	pls	detail	complaint	dm	time	issue	time	debit	pls
Term13	cenbank_cent…	yesterday	people	diamond	detail	access	month	transfer	kindly	need
Term14	kindly	guy	still	pls	still	thank	even	use	transfer	customer
Term15	pls	cenbank_cent…	yet	day	successful	guy	debit	charge	help	citi
Term16	time	naira	time	transfer	debit	kindly	today	number	check	kindly
Term17	card	credit	please	cenbank_cent…	access	even	complaint	atm	customer	guy
Term18	issue	detail	month	twitter	today	detail	service	yet	pls	month
Term19	still	card	atm	central	online	cenbank_cent…	please	eyin_ole	even	america
Term20	back	debit	reverse	response	number	people	year	ur_ur	account	please

Figure 7.2c: Latent Semantic Indexing (LSI) result

The LSI result (figure 7.2c) above shows the method struggled in this context. The method produced terms with banking semantic orientation. However, while investigating the terms across all the topics. It was observed that LSI topics contain repeated or closely related terms across the topics. For example, transaction, day, transfer, cenbank_central, customer, guy and please. This repetition of terms suggest LSI performed poorly. Despite the rigorous pre-processing and tuning of the corpus, LSI contains malformed words like ur-ur. Interestingly, the local term 'eyin_ole' which was found in LDA topics was also observed in the LSI topic. Surprisingly, LSI topics does not contain Pidgin English term.

topics_df_hdp - DataFrame

Index	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Term1	help	oct	facebook	product	black	help	true	account	hey_investme…	form
Term2	access	promise	association	notice	blame	duty	space	apology	experience	resolution
Term3	account	still	knowledge	wonder	joke	angry	fast	olu	help	report
Term4	money	baba	deduction	unclemohamz	help	samuel	head	fact	delete	concerned
Term5	transaction	dont	nhere	operation	branch	die	explorer	tomorrow	peter	cantstopwont…
Term6	please	thread	aware	banker	personally	fintech	star	news	allowance	nplease
Term7	dm	access	help	exchange	negative	sorry_experi…	v	power	told	anyone
Term8	number	longer	complete	cenbank_cent…	notice	worst	help	noodle	poor	sister
Term9	day	v	travel	plz	merger	wizkidayo_wi…	pastor	advance	access	offer
Term10	guy	gift	away	otherwise	tuesday	opay	meant	help	reflect	worth
Term11	still	swift	assistance	see	okay	nysc	etc	completely	unilorin	awareness
Term12	app	start	delivery	respectively	today	gov	place	document	brilafm	october
Term13	la	fccpcnigeria	goalcomniger…	better	aw	weekly	capital	pick	complaint_en…	keep
Term14	time	delay	review	blue	everyday	brilafm	still	lately	wallet	day
Term15	kindly	sent	nation	internet	market	glad	sept	mistake	deduct	till_date
Term16	director	reminder	fire	help	rectify	star	mbuhari_muha…	mastercard	chidi	last
Term17	blame	partnership	aw	tired	private	closing	aisha	okay	directly	problem
Term18	transfer	mind	fintech	customer_ser…	executive	ubacfc	war	life	imagine	ibrahim
Term19	banking	list	right	pain	receiver	question	sincerely	lie	enable_assist	shit
Term20	yet	help	peace	deal	appropriately	omo_iya	morning	several	capitalone	reno

Figure 7.2d: Hierarchical Dirichlet Processing (HDP) result.

The HDP result in figure 7.2d above suggests the model performed fairly. The model produced banking related terms such as mastercard, account, branch, customer_service, transaction, and transfer. However, HDP produced malformed terms despite pre-processing techniques applied. For example, la, aw, and v.

In summary, the LSI result produced the poorest result amongst the topic models. The original LDA and Mallet produced better results. It is worth knowing that this judgment is qualitative and thus just a suggestion. To ascertain the best topic model, this study presents the coherence score of these models in the section below.

7.2 Evaluation of the topic models

Model/Evaluation	Coherence Score (Cv)	Coherence Score (UMass)
LDA	**0.3919**	**-9.1174**
LDA_MALLET	0.3145	-6.9755
LSI	0.3912	-2.6630
HDP	0.6347	-16.5378

Table 7.1: Evaluation of Topic models

Table 7.1 above shows the coherence score of the topic modelling techniques. There is no standard performance cut off for the coherence score in deciding the best topic model. The coherence score is interpreted as the higher the coherence score (CV) the better. Using the coherence scores in Table 7.1 above, the models performed uniformly well in general. Most notably, the hierarchical Dirichlet process has a much higher coherence cv score of 0.6347 but lowest Cv (umass) of -16.5378. The original latent Dirichlet allocation has the next attractive performance by coherence with cv score 0.3919 and Umass -9.1174. While both latent semantic indexing and LDA Mallet performed marginally lower. This study will therefore investigate HDP and LDA to choose the best performed topic model in this context. The HDP and LDA models have been considered state of the art models in topic modelling. For example, Kherwa & Bansal (2020) compared topic modelling techniques and found LDA as the best topic modelling technique. While HDP was argued to be better (Wang et al. 2011; Srijith et al. 2017). It will thus be interesting to affirm the better one in this context.

Topic coherence is a common evaluation method for topic modelling techniques (Ray et al. 2019). However, evaluating topic models using coherence score only is not enough (Asghari et al. 2020). This needs to be complemented with topic quality/interpretability (Albalawi et al. 2020). Considering, the result showed in figure 7.2a – 7.2d, both original LDA and HDP produced quality cluster of terms per each topic. The LDA clustered terms are interpretable. For example, the topic 3 is transaction related. This makes a lot of sense and easy to identify as the terms in each topic are semantically related to customer. The HDP also produced interpretable terms however, it was observed that most of the terms are generic. In addition, HDP produced unwanted terms despite the corpus filtering, cleaning, and pruning. Such as ‘la’, ‘aw’, and ‘v’.

In conclusion, the comparative experiment conducted using latent semantic indexing (LSI), latent Dirichlet allocation (LDA), and hierarchical Dirichlet process (HDP), and MALLET version of LDA showed promising result. The count-based approach with the POS tagging enhanced the performance of the topic models. The coherence score and topic quality were used as evaluation metrics, and this ascertained LDA produced the top-quality terms which are interpretable with competitive coherence score. This outcome is consistent with that of Ray et al. (2019), Munir et al. (2019), Albalawi et al. (2020), M'sik & Casablanca (2020) and Kherwa & Bansal (2020) that ascertained LDA as the leading statistical topic modelling technique in their comparative studies.

7.3 Topic (LDA) Profiling

This section presents the descriptive analysis of the topic modelling result. The LDA model was used to classify the tweets into the 10-topics identified. This will help in understanding the topics as shown in figure 7.3 below.

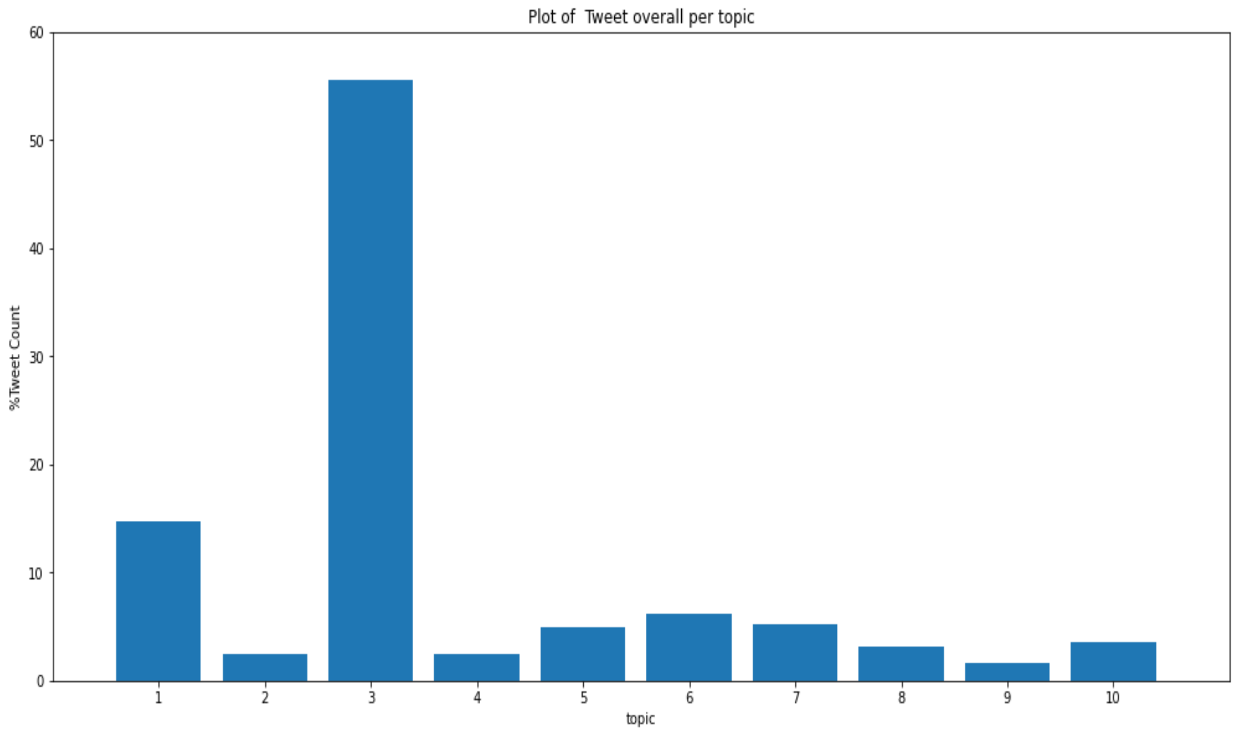


Figure: 7.3: Plot of overall Tweet against topic

Figure 7.3 above, shows the distribution of the tweets per topic. The plot suggests most tweets were classed into topic 3. A minority of the tweets (under 20%) are classed into topic 1 while the rest of the topics account for less than 10% of the tweets. To generate more insight into these topics. It is necessary to discuss the profiles of individual topics.

7.3.1 Topic 1

Topic 1	*Terms*
	customer, cenbank_center, branch, time, month, people, guy, atm, even, week, still, day, service, call, complaint, nothing, customer_care, money, customer_service, thing,

Table 7.2.1: Topic 1

From the plot in figure 7.3, Topic 1 accounts for 14.77% of the entire tweet (document). The topic consists of time, customer, and service experience related terms. The terms suggest tweets clustered are in relation to bank customers complaint, customer care and customer service. During our analysis, findings showed the most dominating tweet for Topic 1 was user's Tweet towards bank staff. The bank customer tweeted, *"get joy forcing customer drag their foot their annoying attitude customer photocopy form".* This tweet is demonstrating a customer's experience in a branch. The example illustrates a scenario of staff bad attitude to customer just to photocopy form and was dragging her foot. The tweet is a sign of frustration due to bank staff attitude. This example has the highest dominance by Topic 1 of about 77.5%. Another example with a dominance of 37.9% is *"allow unauthorized charge taken account given time investigation take month their entire staffing unprofessional"*. The tweet expresses the unprofessionalism of the bank staff towards their investigation of unauthorized charge on the customer's account which has taken months. These examples show that the tweets classed into Topic 1 are of customer experience. Thus, Topic 1 can be named ***Customer service experience***

7.3.2 Topic 2

Topic 2	*Terms*
	job, morgan, free, min, company, una, fossil_fuel, ready, hey_investment, gas_coal, financing_oil, earth_stop, company, nofossilfuel, fargo_jpmorgan, change, actually, head, woman, country

Table 7.2.2: Topic 2

From the plot in figure 7.3, Topic 2 account for 2.5% of the entire tweet. The topic consists of oil and gas, investment, and finance company related terms. Topic 2 can be named ***Investment bank.*** The term JPMorgan was significant in this topic. A possible explanation for this might be the ongoing crisis of the Nigeria government and JPMorgan on the money laundering case which involves the Nigerian banks. This case has a background story of oil and gas related monies. Social media users utilised Twitter to express their opinion about the case and how the Nigeria banks were involved.

7.3.3 Topic 3

Topic 3	***Terms***
	help, access, account, money, transaction, please, transfer, pls, yet, day, number, kindly, yesterday, back, today, issue, alert, need, airtime, fund

Table 7.2.3: Topic 3

A high proportion of tweets were classed into Topic 3. This topic account for 55.63% of the entire tweet as shown in figure 7.3. The topic consists of transaction related terms. Topic 3 can be named as ***Transaction problem.*** Findings suggest tweets allocated to this topic were mainly around day-to-day transaction concerns where customers either make enquiries, seek for help or have major issues with their bank transactions. For example, *"debited since Friday failed credit recipient account issue reverse money"*. This tweet has a dominance score of 91.8% by Topic 3. Another example with high dominance is, *"levity handling issue failed transaction without quick refund getting much guess closing account bad path"*. The tweet is concerning as customer expressed desire to churn. This topic is the most significant topic that the banks need to pay urgent attention as this represent a high proportion of the tweet and profiled as transaction problem.

7.3.4 Topic 4

Topic 4	***Terms***
	loan, help, veilleur, bbnaija, cenbank_center, kindly_follow, day, send_direct, officialefcc, message_details, monthly, thank, house, enable_assist, salary, ph, advise_appropriately, heritagebank, eyin_ole, force

Table 7.2.4: Topic 4

Topic 4 accounts for 2.47% of the entire tweet as shown in figure 7.3. The topic can be named ***Scam alert/advice.*** This is because the topic consists of enquiry and scam related terms. An illustration of tweets classed in this topic are scenarios where customers alert the economic and financial crime commission (EFCC) on potential unsatisfied/fraudulent banking act. For example, *"will continue to advise desist from doing business with eyin ole jati jati"* This tweet can be interpreted as a customer advising others to stop using a particular bank service because the customer feels the bank service is a scam or rubbish. The analysis showed the tweet has a dominant score of 77% by Topic 4. However, it is worth understanding that majority of the tweets in this topic are more likely to be on customer seeking information or advising. This is because terms in the topic are mostly related to information or enquiry seeking.

7.3.5 Topic 5

Topic 5	***Terms***
	card, charge, never, new, use, open, year, debit, later, small, fee, pay, dey, top, see, right, also, month, nyou, far

Table 7.2.5: Topic 5

Topic 5 account for 4.98% of the entire tweet. The topic consists of card, charge, and fee related terms. These terms suggest the topic can be named ***Bank charges.*** Dominant tweets observed are, *"nonsense without permission new atm card produced charged yet transferred card abeokuta still want pay another card go lagos collect card"* and another example, *"no fit laugh abeg dey para sha nno fear na na till next month deduct money for maintenance"*. These tweets express customers unsatisfaction about Atm card charges and fees deducted from their account. This topic is another important topic as customers are concerned about their fees and can influence customers banking decision.

7.3.6 Topic 6

Topic 6	*Terms*
	business, lagos, great, school, news, monetizable, world, additional_information, soon, let, part, good, best, opportunity, today, deposit, wish, thanks, trump, event

Table 7.2.6: Topic 6

This topic covers tweets towards events and programs organised by the banks. The banks aim to contribute to the society and thus have programs to suit this. For example, youth empowerment programs, entrepreneurship programs and scholarship in schools. There are other events such as award ceremonies and sponsorship events. Topic 6 clustered words in relation to these. This topic can therefore be named ***Bank events***. The topic account for 6.18% of the entire tweet. An example of tweet classed into this topic is, *"corporate responsibility sustainability week visited different school today awesome"* and another is, *"guy learn fashion den academy dey fund startups na opportunity"*.

7.3.7 Topic 7

Topic 7	*Terms*
	dm, help, detail, thank, number, twitter, kindly, please, complaint, sorry, send, tweet, transaction, handle, account, information, send_dm, channel, phone_number, message

Table 7.2.7: Topic 7

Topic 7 consists of Twitter handle, help, information, and complaint related terms. This topic account for 5.17% of the entire tweet. Topic 7 can be named ***Contact/complaint form***. Twitter users were observed to ask another user for bank Twitter handle/contact details. The tweets towards the bank suggest customers were requesting information or help from the bank. For example, *"good morning enrolled bvn abakaliki portharcourt currently wish correct update detail include phone number please advise help"*. As part of the information requested, some users were also observed to request for channel to log complaint. For example, “abeg send dm to log complaint”.

7.3.8 Topic 8

Topic 8	***Terms***
	line, phone, support, ng, mobile, mtn, hope, food, consumer, government, international, buhari, sort, whatsapp, explanation, enable, lagos, foundation, industry, ubaat

Table 7.2.8: Topic 8

Topic 8 account for 3.13% of the entire tweet. The topic consists of phone and mobile service-related terms. The terms can be interpreted as clustered tweets of bank customers who bought mobile phone airtime using their mobile banking app, internet, or online banking. Thus, expressed their experience or issue. Another explanation for the terms is the unstructured supplementary service data (USSD) code service. The USSD code is the code programmed with the phone SIM card to perform certain service. For example, In Nigeria, the USSD code is used as a channel where customers can transfer funds to another. The mobile phone SIM card service is provided by telecommunication companies like MTN, GLO and AIRTEL. It is therefore no surprise why the term “mtn” featured in this topic. There are USSD code functions which customer engage in. For example, WhatsApp bundle. Again, this explains why the term “whatsapp” is in this topic. Thus, Topic 8 can be named ***Mobile banking***. An example of tweets classed in this topic is, *“recent phone recharges left account never reflected line face tire mtn support”* and another is *“recharged mobile line debited line credited acct number XXX hope hear ur team soon”*.

7.3.9 Topic 9

Topic 9	***Terms***
	always, poor, stop, financial, ng_union, cool, youfirst, sport, business, service, finance, internet_banking, premiercool, mind, tell, thing, daily, single, fast, vista_intl

Table 7.2.9: Topic 9

Topic 9 account for 1.66% of the entire Tweet. The topic consists of performance, internet banking and network quality related terms. This topic can be named ***Internet service performance***. An illustration of Tweet classed in this topic is *"almost yr banking request otp almost month always come different story ur service poor"*. The tweet suggests the customer had requested for otp for certain months and now almost a year the bank kept giving different reasons for the delay. Other notable tweets are related to how the internet network is poor when accessing internet banking.

7.3.10 Topic 10

Topic 10	*Terms*
	credit_card, citi, bill, state, debt, payment, chase, omo_iya, mastercard, nthank, ng, future, true, alat, abuja, pay, india, atm_machine, child, product

Table 7.2.10: Topic 10

Topic 10 account for 3.51% of the entire tweet. The topic consists of atm_machine, credit_card and payment related terms. Topic 10 can be named ***Credit card payment.*** The qualitative exploration of the result showed most tweets classed to this topic are about getting credit card or loan, transactions using credit card online or via the ATM. For example, *"payment done paying credit card debt follow dm debt pay sugardaddywanted sugardaddyneeded sugardaddytwitter sugardaddy creditcard wellsfargo capitalone usaa"*. The tweets suggest the customer is acknowledging payment of debt on the credit card.

7.4. Summary of the 10-LDA Topics

Topic	%Total Docs	Topic (profile) Name
1	14.77	Customer service experience
2	2.5	Investment bank
3	55.63	Transaction problem
4	2.47	Scam alert/advice
5	4.98	Bank charges
6	6.18	Bank events
7	5.17	Contact/complaint form
8	3.13	Mobile banking
9	1.66	Internet service performance
10	3.51	Credit card payment

Table 7.3: Summary of Topics Identified

Table 7.3 above presents the summary of the 10-LDA topics identified in this study. It was observed that there were transaction related topics like transaction problem, customer service experience, scam alert/advice, bank charges, contact/complaint form, mobile banking, internet service performance and credit card. While the non-transaction related topics are Investment bank and bank events. The non-transaction related topics indicate there are customers that pay attention to non-transaction banking matters. The implication is that banks need to be aware of their reputation whilst engaging with other businesses and society schemes. This is because their activity outside of banking are directly linked to their reputation which influences customers banking behaviour (Aramburu & Pescador, 2019). This is consistent with the findings of Aishatu et al. (2017) that showed bank reputation and trust influences customers.

Prior studies highlighted the importance of the transaction related topics, for example, customer service experience (Ugwuanyi et al. 2020), internet service performance (Eke & Singhry, 2020; Oyelami et al. 2020), bank charges (Haruna et al. 2018), mobile banking (Inegbedion et al. 2019), and credit card payment (Aduaka & Awolusi 2020) as key influencing factors of customers' attitude in the Nigeria banking industry. Very interesting but surprising, previous studies did not pay attention or investigate "transaction problem" and "contact/complaint form" explicitly. In this study, findings showed that most tweets were in relation to transaction problem. A significant percentage of the entire tweets were also on contact/complaint form. The findings indicate when bank customers have transaction problem, customers tend to go online to express their frustration. Some customers go to the extent of copying CBN (Central Bank of Nigeria) and EFCC (Economic and Financial Crime Commission) in their Tweet. While some customers might request complaint form when their issues are not resolved for a period. Surprisingly, there was no study found to have investigated transaction problem as a major constraint. This is because previous studies were limited to survey or interview dataset (pre-defined variables). Thus, this study contributes by providing up-to-date themes which highlights influencing factors to customers' attitude in banking.

7.5. Human evaluation of LDA Topics

This study employed human evaluation to check the quality of the topics. In doing so, 300 tweets were selected for this purpose. Three human annotators were employed, and the performance were reported as shown in figure 7. 4 and table 7. 4 below.

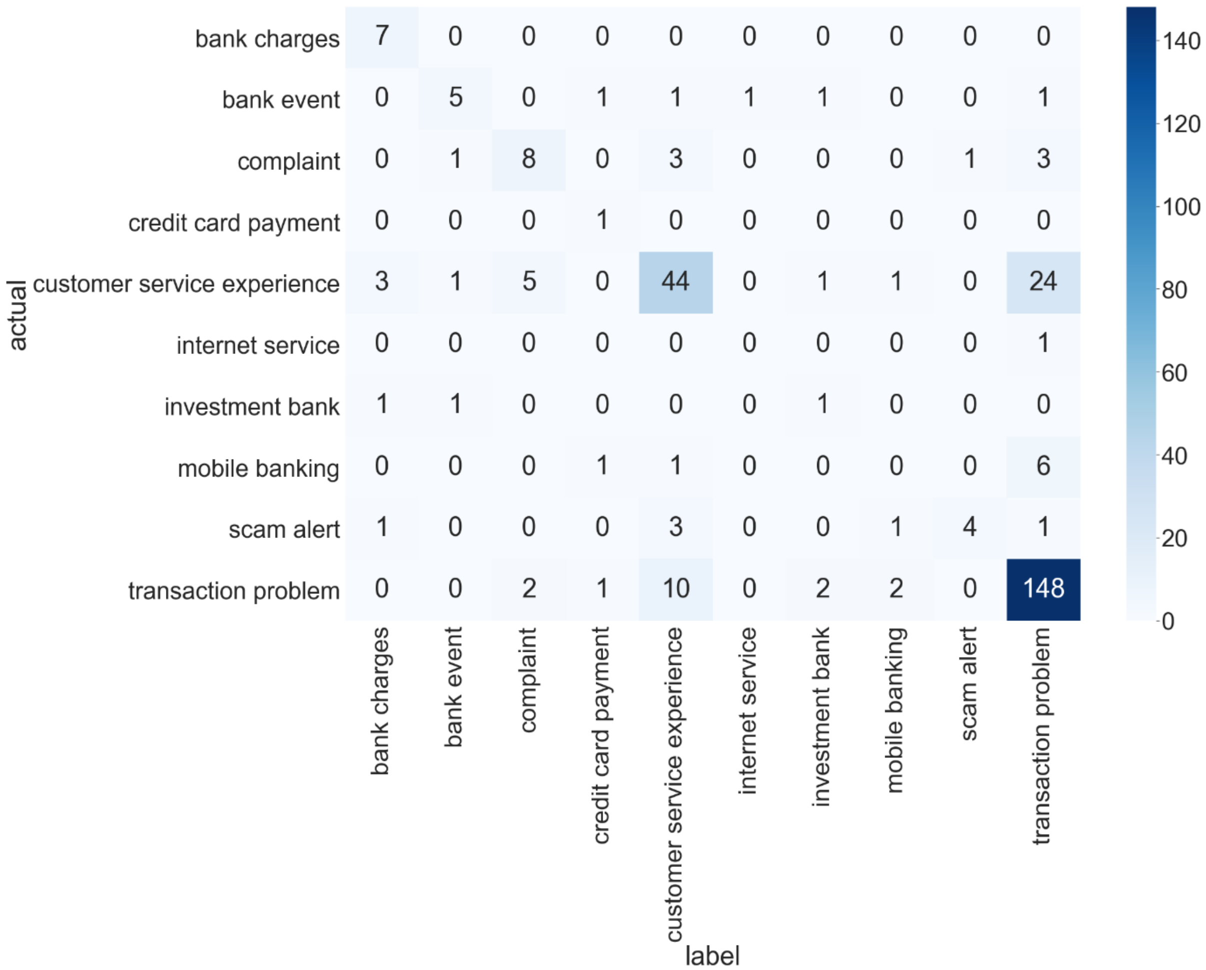


Figure 7.4: Heatmap of correctly predicted topic classification

The plot in figure 7.4 above provides a comprehensive understanding of the numbers of correctly and incorrectly classified topics. For example, internet service has a tweet which was misclassified as transaction problem. Mobile banking has a total of 8 tweets in which 6 was misclassified as transaction problem. Transaction problem has a total of 165 tweets and 10 of that was misclassified as customer service experience. Thus, table below presents the summary of the performance evaluation in terms of precision, recall and balanced F1 score.

	bank charges	bank event	complaint	credit card payment	customer service experience	internet service	investment bank	mobile banking	scam alert	transaction problem	accuracy
precision	0.58	0.62	0.53	0.25	0.71	0.0	0.20	0.0	0.80	0.80	0.73
recall	1.00	0.50	0.50	1.00	0.56	0.0	0.33	0.0	0.40	0.90	0.73
f1-score	0.74	0.56	0.52	0.40	0.62	0.0	0.25	0.0	0.53	0.85	0.73
support	7.00	10.00	16.00	1.00	79.00	1.0	3.00	8.0	10.00	165.00	0.73

Table 7. 4: Evaluation of LDA topics

The evaluation result showed the model as performed well with accuracy of 73%. The most interesting part is that transaction problem which has the highest tweeted topic was correctly classified by the model with a precision of 80% and recall of 90%. Unfortunately, there were some topics that were wrongly classified as shown in figure 7.4 above. For example, mobile banking and internet service performance. This is because customers had tweeted transaction problem using the mobile banking app and internet service performance. An example of such tweets is *“wahala dey with this internet oooo made transaction on my internet banking, dem debit me but dem no credit beneficiary”.* This kind of tweets were classified as transaction problem regardless of if the transaction was made on mobile or internet banking. This is because the tweet consists of terms that are profiled as transaction problem. The deployment of the framework depends entirely on the bank. This study grouped all transaction tweets together because it is easy for banks to understand the volumes of transaction errors and how it frustrates customers as some customers aimed to close their account based on transaction problem experienced. Overall, the LDA topic model has performed well. In the next section, this study will thus, detail how the sentiment and topics were assigned.

7.6 TSBS Visualization

The TSBS framework as discussed in chapter 4 (section 4.1.2) proposed components where customers tweet can be classified into a topic and sentiment polarity. In chapter 5, 6, and 7 the processes were discussed, and results were presented. Thus, this section provides an explanation on how the sentiment label to topics were assigned using example of customer tweets provided in table 7.5 below.

Bank Customer Tweet	Sentiment Feature	Sentiment Class	Topic Terms	Topic Class	Contribution to Topic
request to close account charging card maintenance fee no dey interested in banking anymore	close, charging, interested	Negative	card, charge, fee	Bank charges	54.48%
na sure bank even for weekend no wahala customer care dey there	sure, wahala, care	Positive	customer, even, customer_care	Customer service experience	51.2%

Table 7.5: Example of Sentiment -Topic classification

From table 7.5 above, the first example shows unhappy customer willing to close account based on the fee charged on the account (maintenance fee). This study profiled the tweet as bank charges because the tweet is related to card maintenance fee charged on the account. The sentiment features identified are *"close"* (negative sentiment word), "*charging*" (negative sentiment word) and "*interested*" (positive sentiment word). It is worth noting that there is a term *"no"* which negates the positive word *"interested"*. The topic model identified terms like "*card*", *"charge"*, *"fee*" which are terms associated to bank charges as shown in section 7.3.5. Thus, the tweet equates to be negative sentiment tweet of bank charges. Similarly, the second example relates to tweet from customers recommending a bank as a good bank to another customer based on good service experience as the bank customer care offers good service on a weekend. The tweet consists of sentiment terms like "*sure*" (positive), *"wahala"* (negative), and *"care"* (positive). Also, there is a term "no" which negates the negative word *"wahala"*. The topic model identified terms like card, charge, fee which are terms associated to customer service experience as shown in section 7.3.1. Thus, the tweet equates to be positive sentiment tweet of customer service experience. The aggregate of the sentiment polarities is visualised in the plot produced in figure 7.5 below.

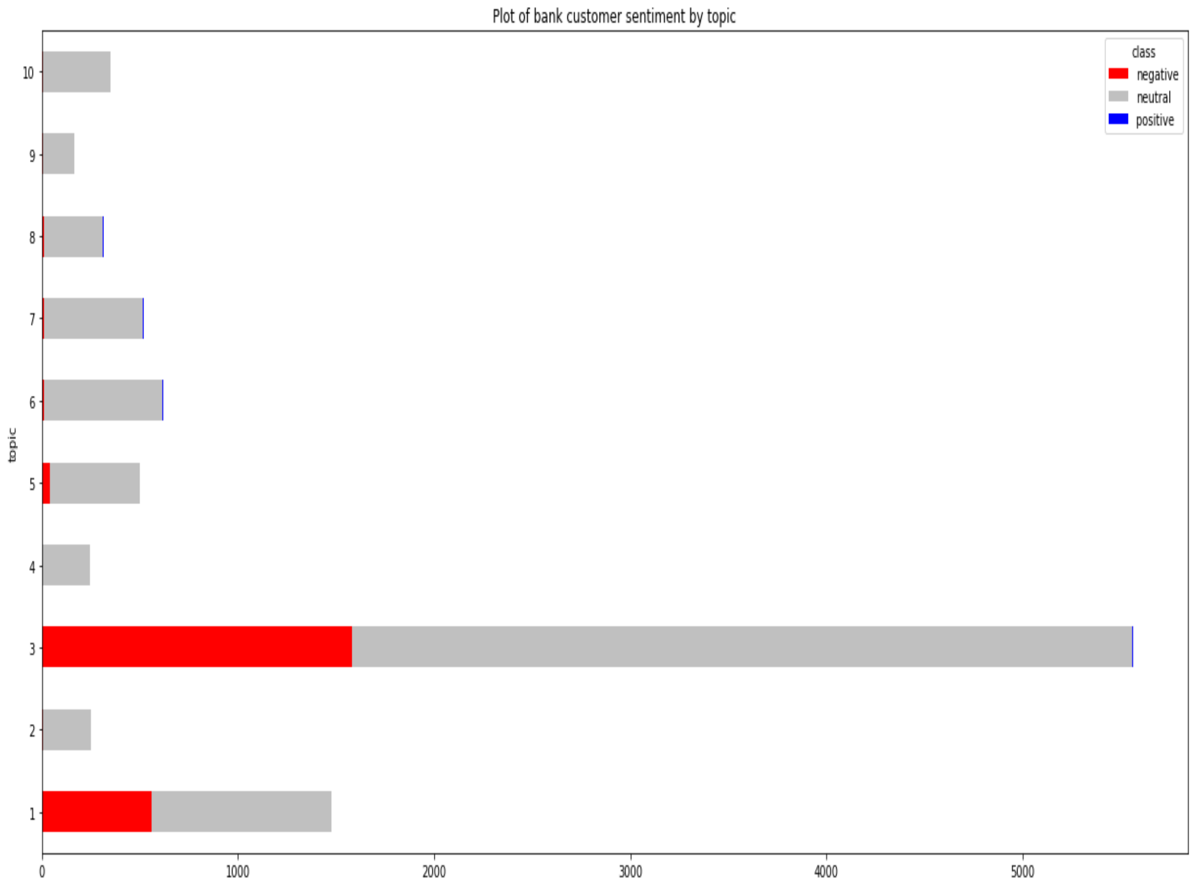


Figure 7.5: TSBS Visualization

Figure 7.5 above shows the plot of topics against the customers' sentiment. The plot is used to demonstrate the TSBS framework. The negative class (in red) indicates the negative sentiment (subjective) Tweets, while the positive class (in blue) indicates positive sentiment (subjective) Tweets. The neutral class (in grey) indicates neutral sentiment (objective) tweets. This plot helps in uncovering the hidden pattern within the topics. In general, the plot shows objective Tweets (neutral) dominate the topics. This suggests bank customers use social network site (SNS) to request information concerning their bank service or transaction. It could be argued that the low positive (sentiment) tweets are due to human behaviour towards the service industries. It is not frequent for customers to praise their banks on SNS for good service experienced.

Topic 3 (transaction problem) is the most tweeted topic. This shows transaction problem is the most significant topic in this study. More than half of the tweets in topic 3 were expressed as objective statement. This indicates customers try to inform the banks of their transaction problem without using any subjective statement. While more than a quarter of the tweets in topic 3 were negative. This suggests there were high number of customers that expressed their frustration about their transaction. Thus, the banks need to be proactive to transaction problems encountered by their customers. A very tiny proportion of positive tweets were also observed in Topic 3. This might be a situation where customers got their problem resolved and thus, praised the bank for their prompt resolution. Topic 1 (customer service experience) was the second amongst the topics with topmost tweets. The high number of neutral tweets in topic 1 suggest customers tweeted a lot to request information or advice about their account or transaction. While more than a quarter of the tweets were negative. This suggests there were a lot of problems surrounding customer service experience too and thus, some customers are unhappy. This result can be explained with the fact that, there was evidence from the topic 3 terms that suggest customers used the customer care unit to access banking services. However, customers' experience was bad and thus, customer used SNS to express their disappointments. For example, *"get joy forcing customer drag their foot their annoying attitude customer photocopy form".* The example illustrates a scenario of staff bad attitude to customer. It was also observed that some topics consist of tiny proportion of negative tweets. For example, topic 5 (bank charges), topic 6 (bank events), topic 7 (contact/complaint form), and topic 8 (mobile banking). This indicates customers are unsatisfied with some charges on their account and the mobile banking. However, some of these topics showed tiny proportion of positive tweets such as topic 6 (bank events), topic 7 (contact/complaint form), and topic 8 (mobile banking). While other topics such as topic 2,4,9, and 10 were mainly neutral.

With respect to the first research question (RQ1), it can be seen from the TSBS plot that topic 3 (transaction problem) is the topmost tweeted topic. The banks need to find ways to resolve transaction problems promptly. This might be a case of assigning more staff to solve transaction problems both online and bank branches. There is need for banks to adopt a system that can automatically reverse unsuccessful transaction, especially the electronic transactions. Topic 1(customer service experience) was also a major concern. The TSBS plot showed both topics 1 & 3 are concerning due to high numbers of negative tweets within these topics. The banks need to find a way to enhance customers' experience. This is because prior studies found out customers' experience influences their attitude to banking in Nigeria (Inegbedion et al. 2019). One of the tweets observed in topic 1 highlighted bad staff attitude. To resolve this issue, the banks need to provide regular staff training towards enhancing customers' experience. A significant proportion of objective tweets was observed towards Topic 8 (mobile banking), Topic 9 (Internet service performance) and Topic 10 (Credit card payment). It is worth knowing that terms in these topics are popular banking channels, for example, ATM, USSD, and internet banking. These terms correlate with keywords from the exploratory data analysis discussed in section 6.1 of the previous chapter. The automated teller machine (ATM) is the most used form of electronic banking in Nigeria (Oyelami et al. 2020; Mokhlis et al. 2018). Recently, the USSD was added to the banking channels for customers. The growing usage of these electronic banking services suggests this is an area that the bank can use as a competitive advantage. Lastly, the plot also showed several other topics with tiny proportion of negative tweets such as topic 5 (bank charges), topic 6 (bank events), topic 7 (contact/complaint form), and topic 8 (mobile banking). These topics with negative sentiment provide the banks with up-to-date aspects of banking that needs improvement.

7.7 Chapter Summary

In summary, this study utilised topic modelling techniques for aspect extraction because the statistical techniques can detect both explicit and implicit topics within the document. This chapter presents the comparison of the unsupervised topic models namely, latent semantic indexing (LSI), latent Dirichlet allocation (LDA), and hierarchical Dirichlet process (HDP). The coherence score and topic quality were used as evaluation metrics, and thus, LDA was ascertained as the best performed topic model in this context. This is because LDA produced competitive coherence score (0.3919) and top-quality terms which are interpretable. To understand what part of the banking product or services customers tweeted about, the unsupervised model was used to classify the customers' tweets into 10-topics. The topics were profiled as customer service experience, investment bank, transaction problem, internet service performance, bank charges, bank events, contact/complaint form, mobile banking, scam alert/advice, and credit card payment. Out of the 10-topics, transaction problem and customer service experience are the most significant topics.

The sentiment model in the previous chapter was integrated into the pipeline to demonstrate the Topic-Sentiment Banking System (TSBS) framework proposed. The TSBS was used to demonstrate the customers' sentiment towards the topics identified. The result presented showed the aspects of banking that need urgent attention. For example, transaction problem, customer service experience and mobile banking. Another interesting finding is that customers are unhappy with bank charges. To conclude, the TSBS demonstrated how beneficiaries such as the commercial banks can develop an automated system to detect topic and sentiment of bank customers' tweets. This will help in identifying customers experiencing bad service and monitor usage or adoptability of other banking channels. In the next chapter, this study will examine the generalisability and acceptance of the research outcome/framework.

CHAPTER 8: Research Validation

8.0 Introduction

This chapter discusses the validation of the Topic Sentiment Banking System (TSBS) framework developed for commercial banks in Nigeria. TSBS was developed based on natural language processing (NLP) techniques such as sentiment analysis and topic modelling which were implemented to investigate customers' attitude towards banking. The framework was designed to help the banks automate their system to classify customers' online contents into banking aspects (topics) and sentiment polarities (negative, neutral, and positive). This was needed to give the banks competitive advantage, monitor their product and service, reduce customer churn, enhance customer experience and profitability.

The aim of the validation process is to determine the reliability, relevance, and usefulness of TSBS to the banks. This will help understand if the research findings and recommendations are reliable and relevant. McBurney & White (2007) stated validity provides an indication that shows the extent to which a research conclusion corresponds with reality. Validation is an important part of research and its result because it evidences the relevance, confidence assurance, value, and reliability of the study in the organisational field (Hayashi et al. 2019). In summary, validation provides a strong background to which the research findings can be generalized. The next section discusses the concept of validation and subsequently the validation processes.

8.1 The Concept of Validation

Validation is an integral part of framework development processes. Kennedy et al. (2005) highlighted the importance of validation as a key part of framework development processes as it enhances the confidence in the framework and thus makes it more valuable. Golafshani (2003) added validation provides the means to understand if research conducted truly measures what it was intended to measure and how truthful are the research findings. The concept of validation has several meanings in different research stages especially the conceptual, methodological, and empirical domain (Brinberg and McGrath, 1985). In the conceptual domain, Landry et al. (1983) defined validation as accessing the extent of relevance of the assumptions and theories underlying the conceptual framework. They further stated conceptual validation generally aim to provide answers to questions like,

- Is investigation of problem conducted from the appropriate perspective?

- Is the perspective susceptible to getting the right solutions?

To conclude, conceptual validation is established by accessing the effectiveness, testability and adaptability of the concepts used (Apulu, 2012; Daub et al. 2021). The methodological validation entails accessing the efficiency, rigour, unbiases and explicitness. While the empirical validation ensures the research is relevant and useful in practical. Brinberg and McGrath (1992) summarised the importance of integrating the three domains as value, relevance, and generalisability. However, Winter (2000) argued that validation is not a single, fixed or universe concept but rather a contingent construct, inescapably grounded in the process and intentions of research methodologies and projects.

Validity can be described using two main components, which are, external and internal validation. Internal validation accesses the relationship between the treatment and research result. Williams (2020) defined internal validity as the identification of a causal effect via comparison with a valid counterfactual. While external validation accesses the extent which the research result can be generalised (Egbu, 2007; Yuilmaz, 2013). External validity is the extrapolation of research findings beyond the study sample to another population (Williams, 2020; Findley et al. 2021). The bank decision makers are concerned about the applicability of the research findings to the bank. In this context, the validation should demonstrate the degree to which the research result can be generalised to the Nigerian banking industry. Both (internal and external) techniques have been successfully employed in previous studies (Xiao 2002; Ankrah 2007; Ikpe 2009; Apulu 2012; Daramola et al. 2014; Supakkul et al. 2020).

In summary, the concept of validity is to examine that the research findings are true, relevant, and accurate not only from the researchers' perspective but also from the domain expert's perspective (Creswell & Miller, 2000). This will help understand the quality of the research in terms of credibility. The degree to which the research findings can be trusted and generalised depends on the process of validation adopted. In this study the external and internal validation were used to validate the research findings which are described in the section below.

8.2 Validation Approach

The two main components (external and internal) of validation were employed to validate the research findings. The approach aims to understand how reliable, relevance, and useful the research findings are to the Nigeria banking sector. To achieve this, this study will validate the TSBS framework and the research findings with the domain experts (bank staff) using survey questionnaire and interview. A non-probability (convenience) sampling approach is utilised to select participants of interest. Convenience sampling is a non-probability sampling technique used to select certain members of a target population that meets certain practical criteria such as easy accessibility, geographical proximity, expedience, availability at a given time, or the willingness to participate (Özdemir et al. 2011; Farrokhi & Mahmoudi-Hamidabad 2012; Emerson 2015; Sedgwick 2013; Suen et al. 2014; Brewis 2014; Etikan et al. 2016; Baltes & Ralph 2020; Andrade 2021). The sampling technique was chosen because it helps direct the survey to the target population of interest for this study. The technique is commonly used (Krupnikov et al. 2021) and has been employed successfully especially where randomised sampling is difficult to achieve (Etikan et al. 2016; Burke et al. 2020; Daud et al. 2021). Bosnjak et al. (2016) explained in their study that probabilistic sampling is rare when survey to be conducted is of interest to a specific domain. This study utilised the online questionnaire software because it is better compared to other techniques to access the target population. Other option considered was postal surveys. However, this is difficult to use due to time and cost constraints. This leaves the online questionnaire as an appropriate approach.

The problem of limited information and restrictive nature of survey questionnaires are overcome by providing appropriate information and thorough design of the questionnaire. To get a good representation of the target population, participants from different commercial banks were targeted. The criteria of selection of participants are primarily based on experience and role in the banking industry. The participants must have minimum of 2 years' experience in the bank and must be directly involved with customer queries/experience. The online questionnaire link that includes the TSBS framework and brief introduction of the application was sent to the participants. To ensure good response rate, follow up calls were made to participants to support their understanding with the implementation.

8.3 Questionnaire: Discussion of Validation Result

It is worth acknowledging that the sample used for validation is small. However, the research outcome received positive feedback which is detailed in this section. The online questionnaire was filled by 94 participants who are Nigeria bank staff. Out of the 94 responses, 5 were deleted because the respondents did not consent to the information collected for the purposes of this research study and/or unwilling to participate. This brings the questionnaire to 89 responses. The data was analysed using Microsoft Excel to determine the frequency (in percentage) to which the domain experts at least agree/disagree to the research outcome. The result of the descriptive analysis showed respondents were made up of 60% male, 39% female and 1% who prefer not to say their gender. The result showed about 85% of the respondent use social media more than twice a day and over 50% of the respondent stated they use the network mostly to access information. Interestingly, this study utilised Twitter dataset which is one of the most used social networks according to the respondents. Other most used social networks are Instagram, Facebook, and LinkedIn. In general, the result showed domain experts agree that the Topic Sentiment Banking System (TSBS) has huge potential benefits to the banking industry and the research findings were acceptable. The detail of the result is thus shown in Table 8.0 – 8.4 below.

To what extent do you agree with this?	Listening to customers' social voice is RELEVANT to understand their attitude and expectation.	Listening to customers' social voice is USEFUL to understand their attitude. This can help in making business decisions.	Monitoring customers' attitude towards bank product and service is BENEFICIAL.
Strongly Agree	49.4%	47.2%	59.6%
Agree	43.8%	44.9%	32.6%
Neutral	3.4%	2.2%	1.1%
Disagree	1.1%	2.2%	1.1%
Strongly Disagree	2.2%	3.4%	5.6%

Table 8.0 TSBS validation result 1

From table 8.0, at least 92% of the respondents agree that listening to customers perception using social media data is useful and relevant. In summary, in terms of usefulness and relevance, the domain experts agree that the concept of using social media data to understand bank customers attitude is a useful and relevant concept. Similarly, at least 92% of the respondent agree that the concept of monitoring customers' attitude towards bank product and service is beneficial.

To what extent do you agree with this?	Sentiment analysis helps classify happy and unhappy customer. Thus, marketing campaigns can be utilised to offer personalised product/service discounts to the customers.	Insights from sentiment analysis system can uncover areas that need improvement. Thus, such system is beneficial to the banks.	The usage of social network is growing. An automated system that organises and classifies your customers' message/queries into topics and sentiment positivity/negativity is considered beneficial.
Strongly Agree	35.2%	38.6%	38.6%
Agree	51.1%	50.0%	46.6%
Neutral	9.1%	5.7%	8.0%
Disagree	2.3%	3.4%	4.5%
Strongly Disagree	2.3%	2.3%	2.3%

Table 8.1: TSBS validation result 2

From table 8.1, majority of the respondent agree that the concept of TSBS is beneficial. Specifically, at least 86% agree that sentiment analysis is a useful concept to help classify happy and unhappy customers. Also, 88% of the respondent agree that insights generated from such system can uncover area that needs improvement, and thus, is beneficial to the banking industry. Similarly, at least 85% of respondent agree that automated system that organises and classifies customers' message/queries into topics and sentiment positivity/negativity is considered beneficial.

	13). Which one of these topics do you consider the MOST talked about theme (by customers) from your experience?
Bank events	1.16%
Contact/complaint form	4.65%
Mobile banking	2.33%
Bank charges	**31.40%**
Transaction problem	26.74%
Credit card payment	3.49%
Customer service experience	25.58%
Scam alert/advice	4.65%
Investment Bank	0%
Internet service performance	0%

Table 8.2 Topic by Frequency

At least 25% of the respondents indicate that customer service experience, transaction problem and bank charges are the most talked about topic in banking. Specifically, the result shows 31.4% of respondent chose bank charges, 26.74% of respondent chose transaction problem and 25.58% of respondent considered customer service experience as the most talked about topic from their experience. The validation result aligns with the research findings in chapter 7, that showed transaction problem and customer service experience are the most talked about topics. However, it is worth noting that the validation result affirms bank charges as the most important topic from the domain experts experience.

To what extent do you agree with this?	Transaction problems are a major challenge to bank customers.	Some customers are unhappy with BANK CHARGES (like interbank transfer, card maintenance & SMS debit alert fees).	A significant proportion of customers are unhappy with customer service experience	Our research suggests the use of mobile banking, USSD and credit card is increasing. Knowing this can give a bank a competitive advantage.
Strongly Agree	47.2%	62.9%	33.7%	56.2%
Agree	36.0%	28.1%	46.1%	38.2%
Neutral	9.0%	3.4%	10.1%	1.1%
Disagree	3.4%	0%	7.9%	0%
Strongly Disagree	4.5%	5.6%	2.2%	4.5%

Table 8.3 Validation of research findings

From table 8.3, the result from the respondents suggests the research findings are valid. 83% of the respondents at least agree that transaction problems are a major concern in the Nigeria banking sector. About 91% of respondents at least agree that some customers are unhappy with bank charges such as card maintenance fee. Also, 79% of respondent at least agree that a significant proportion of customers are unhappy with customer service experience. Lastly, over 94% at least agree that the customer usage of mobile banking, USSD and credit card is increasing and thus, can be used as competitive advantage. In summary, a high proportion of the bank experts at least agree with the insights generated from the TSBS in the previous chapter. However, it is worth noting that table 8.2 and 8.3 suggests the bank experts suggest bank charge is the most talked about topic and some customers are unhappy with the bank charges.

To what extent do you agree with this?	18). A topic sentiment classification system is URGENTLY needed in banks to help in providing customer centric bank services.	19). A topic sentiment classification system should be a COMPULSORY system for the banks to help in providing customer centric banking service.
Strongly Agree	28.4%	24.7%
Agree	53.4%	44.7%
Neutral	17.0%	25.9%
Disagree	0%	3.5%
Strongly Disagree	1.1%	1.2%

Table 8.4: TSBS validation result 3

From table 8.4, the result from the respondents suggests the TSBS framework is useful and relevant. Specifically, 81% of the respondents at least agree that a topic-sentiment classification system is urgently needed to help banks in providing customer centric bank services. While 69% at least agree that such system should be made compulsory for the banks. To conclude, majority of the respondents at least agree that the research outcome and TSBS framework are positive contributions to help banks in providing customer centric bank services. The concept of the TSBS framework were considered useful, beneficial, and relevant. In addition, some of the respondents provided further comments which are also positive about the research. These comments are shown as follow.

- Sentimental classification system can harness the increasing banking data.

- That bank should try to study and evaluate customers sentiments to improve on service rendering.
- Sentiment classification system can come in handy, and it will be a useful tool to drive client centricity using sentiment and behavioural data obtained from customers.
- A larger percentage of customers in Nigeria, the mass market is social media frenzy but not internet banking frenzy; so the USSD/phone banking caught up a very long time ago in Nigeria, I mean like 13 to 15 years ago
- I think a sentiment classification system (especially if automated) is a necessary development to attain the customer centric bank services goal but then I do not think it compulsory as different banks already have methods by which they track customer satisfaction. But then again it is very necessary.
- The topic of the survey is very relevant to today's banking system

8.4 Interview: Discussion of Validation Result

In addition to validating the TSBS outcome, this study conducted interview via Zoom to get the opinion of the domain IT experts on the reliability, relevance, and usefulness of the TSBS framework. This study constructed five open-ended questions which can be found in appendix E for this purpose. This was adopted because open-ended questions allow respondent to express their views spontaneously without the influence of the researcher. A structured interview was utilised because this allows the interviewer to ask respondents the same set of questions and thus record their responses (Saunder et al. 2007). The criteria of selection of participants are primarily based on experience and their role in the bank. The criteria employed is that participants must have minimum of 5 years' experience in the bank's IT department. There were three bank IT staff that accepted to be interviewed. It is worth acknowledging that the sample is small. However, the framework received positive feedback. The three participants consist of 2 male and 1 female in the age range of 35 – 47 years and they are active users of social media. During the interview, the TSBS framework was presented, demonstrated, and discussed. Afterwards, the set of questions were presented to the respondents and their responses are summarised as follows.

Question 1: How would you describe your experience with the TSBS whilst demonstrating?

Participant 1 - In general, the framework demonstrates the application of human intelligence. The system is transparent and not complex. The system demonstrates thorough processes and there was comparison of methods for each component before proposing the method with optimal performance.

Participant 2 - The framework provided clarity on the requirements of the system. The system is easy to understand and navigate around the modules including the resources. The system appears easy to deploy. The framework can be delivered using popular programming language. Thus, reduces development cost and time. The system demonstrates thorough processes and the performance achieved is good. The appropriateness of techniques adopted by the system were well discussed. Clarity on top level functionalities were provided in the framework. A common challenge with framework is how they struggle to demonstrate connection and communication. However, the TSBS was able to demonstrate the connection at all stages from data source to database to modelling and the prediction stage.

Participant 3 - The framework is suitable for the bank to understand their customers. The framework is adaptable which is the most important attribute. Another important thing is, it is easy to integrate with existing customer experience survey channels. The scope of the input is not limited to tweet but can work well with our current approaches of understanding customer perception. For example, reviews, satisfaction rating questionnaire and customer service data. This increases the chances of the system been adopted. An interesting attribute of the framework is the flexibility. The framework provides a flexible architecture to accommodate changes in frequency threshold, number of terms and topics. The system components are easy to manage because they contain modules and resources that are flexible to modify and query. Whilst demonstrating the system, insights were coming up within on the benefits of the systems as it could revolutionise and complement strategies and decision-making processes in terms of product development and customer experience. All phases of the analytics are clear, thorough, and transparent.

Question 2: To what extent is this system beneficial to the banks?

Participant 1 - The system is beneficial because it helps to understand how customers use our services, most importantly, the active aspects of our services.

Participant 2 - The framework is beneficial because it provides components that produces fine-grained outputs which can be applied to analyse customer emails, surveys, chat data, and queries. The system highlights the significance of the topics. For example, a case of identifying the most mentioned topic in bank chat box data.

Participant 3 - The banks already have the Know Your Customer (KYC) units that intend to know more about their current and intended customers. The KYC analyst analyse customer behaviour and make suggestions in the areas of product development and customer satisfaction. The framework is beneficial to them to understand customers' behaviour towards the bank service. Thus, make recommendations to support marketing and customer experience.

Question 3: What areas do you think the TSBS would be useful?

Participant 1 - Rather than just understanding if a customer is satisfied or not satisfied, it is more useful that the system can help identify the customer issues in terms of aspect.

Participant 2 - The monthly customer experience survey identifies on a scale of 0 – 10 how customers are satisfied with services provided by the banks using pre-defined variables. However, the framework is more useful in providing insights into customers attitude based on undefined aspects of the bank services.

Participant 3 – The current system of investigating customer experience relies on In-app reviews, data from the customer fulfilment centres, emails, customer feedback, and rating. The TSBS framework is useful to incorporate all these channels of measuring customer experience to provide better insight into meeting customer on-going needs.

Question 4: Do you consider the framework relevant to what is needed in your bank?

Participant 1 - The framework is relevant and similar to the current customer experience and satisfaction survey used by the customer experience and value management (CEVM) team. However, the framework demonstrates new approach to understand customers. Interestingly, the system can accommodate the existing survey approach. The framework consists of modules in Python language. This can conveniently communicate and access our current code which makes the development and deployment easier.

Participant 2 - The highlighted topics reflect how the bank customers get our services and their feeling establish their experience. This system provides insight about this, which is relevant to bank need. In my opinion, the framework is acceptable

Participant 3 - Customer perception, experience and behaviour understanding are essential to banks. There are several schemes that already exist to achieve this. For example, the customer experience survey. However, this framework provides an automated approach that can analyse data from diverse sources which evidences the relevance of your system.

Question 5: In your own words, can you describe the framework in terms of adoption, development, and deployment.

Participant 1 - The importance of using the machine to perform like human cannot be undermined. However, the issue is that the quality of the training dataset in this case relies strongly on the annotators. The diverse in Nigeria language makes this more difficult. Whilst adopting this system, a careful approach needs to be taken in the labelling phase to generate reliable resource. Otherwise, this framework can be developed and deployed easily. If available, I am highly likely to use it.

Participant 2 - The banks are aware of systems like this. However, it is not yet available. This framework is easy to integrate with the current survey system because the system is transparent at all stages and the development is easy due to the friendly tools applied. The framework's openness and careful application of the tools aligns with the end goal to understand customer behaviour. Therefore, the banks are more likely to adopt the framework.

Participant 3 - The system requires a lot of human effort to label required input data. However, the framework requirements are not expensive to acquire and with that in comparison to the benefits of the system, the banks are more likely to adopt this system.

8.5 Chapter Summary

In this chapter, this study examined research validity, which is to determine if the research findings and outcome are true, relevant, and accurate not only from the researchers' perspective but also from the domain expert's perspective. This will help understand the research credibility and understand if the research outcome can be trusted and generalised. To achieve this, the external and internal validation were employed. Egbu (2007) stated internal validation seeks to highlight the strength of the model/framework and external validation seeks to highlight the relevance and usefulness of the framework. Apulu (2012) added internal validity examines the credibility of the inferences made from the data and external validity examines the generalisability. During the experimentation, the TSBS framework was developed using several NLP approaches such as sentiment analysis and topic modelling. A benchmark comparison of the techniques was conducted, and the models were tested with different set to the training set.

The internal validation was based on the model performance, which was reported in terms of evaluation metrics such as accuracy, precision, recall and f1-score. The detail was provided in section 6.1 for sentiment classification and section 7.2 and 7.5 for topic modelling. In addition, some of the findings have also been published in the literature as discussed in section 7.6. With respect to external validation, the result from the analysis of the participants' responses suggests the TSBS framework is useful, beneficial, and relevant. Most of the respondents agreed that the framework is needed to build customer centric service in the banking sector. To conclude, the result suggests the research findings are valid and can be generalised across the Nigeria banking sector. In the next chapter, this study will therefore conclude on the entire study, evaluate the research objective and questions. This will help understand if the aim of the research study has been achieved. Thus discuss the limitations and recommendations.

CHAPTER 9: Conclusion & Future Work

9.0 Conclusion

The concept of investigating customer attitude is to understand the feelings, emotion, and belief of customers towards a product or service. There are few studies discussed in chapter 2 that investigated customers' attitude in the banking industry. However, these studies were found to have used pre-defined questionnaires and interviews for their investigation. There are obvious biases of these data collection tools as sometimes, towards what the author wants. The question is, *"what if customers' feelings are out of scope of the questionnaires or interviews?"*. This is where the advantage lies with data from social network sites (SNS) as customers express their feelings freely. The exponential growth and usage of SNS as discussed in section 1.2.2 validates the need to utilise the user generated content (UGC) to inform decision-making processes and policies. Unfortunately, there is no study found to have investigated customers' attitude with SNS data.

Since customers' feeling or behaviour towards a service is based on emotional or sentiment connection made. Sentiment analysis is deemed appropriate to measure the attitude and understand customers' emotional or sentiment connection towards bank service. SNS such as Twitter is a popular social network across the globe. However, analysing this type of dataset is complex. In the context of this study, it is more difficult because it involves the natural language processing of Pidgin English, English, and Nigeria local words. The lexicon-based approach relies strongly on words for sentiment analysis. It is labour intensive and time consuming to produce a robust lexicon to perform this task. Alternatively, the supervised learning models perform well. However, they also rely on annotated set for training. The major concern is that there is limited lexical resource and no validated model to suit this purpose. To address these problems and fill the gap identified, objectives formulated were as follows.

- To review literature on bank customers' attitude, service experience, sentiment classification and aspect extraction for sentiment analysis. This will provide background knowledge to this study especially on the methodologies applicable in this context.
- To propose Topic-Sentiment Banking System (TSBS), a context-informed framework to understand customers' attitude towards up-to-date undefined aspects of banking. This will help detail approaches towards answering RQ1.

- To develop a novel lexicon algorithm that captures the tone and semantic orientation of banking and Pidgin English terms. This will help establish one of the components needed for the TSBS and thus answer RQ2.
- To conduct a benchmark comparison of sentiment classification and topic modelling techniques at tweet-level. This will help establish other components needed for the TSBS and thus answer RQ1, RQ3 and RQ4.

The objectives will thus help to achieve the aim of the study and answer the following research questions (RQ):

- RQ1. To what extent can automated sentiment analysis (SA) systems help generate actionable insights on customers' attitude towards bank product and service?
- RQ2. What are the effects (in terms of performance) of local & context-based terms in lexicon-based sentiment analysis?
- RQ3. Can machine learning models perform as an off-the-shelf sentiment classification method in the banking domain?
- RQ4. What are the major causes of misclassification?

9.1 Evaluation of Research Objectives & Questions

In this section, the objectives as stated above will be reviewed and assessed to check if the research question has been answered. Subsequent sections 9.1.1 to 9.1.4 thus discuss the evaluation of objective 1 to 4.

9.1.1 Objective 1

> 1). To review literature on bank customers' attitude, service experience, sentiment classification and aspect extraction for sentiment analysis. This will provide background knowledge to this study especially on the methodologies applicable in this context.

To have adequate background knowledge to this study, literature in key areas such as banking, customer attitude, service experience, sentiment classification and aspect extraction for sentiment analysis were reviewed. The knowledge gained from this uncovered, the current issues, available resources, techniques, and solutions already proposed in the field of aspect extraction and sentiment analysis. For example, in chapter 2, it was revealed that technological advancement has increased competitiveness in the banking industry. This huge competition led to decreased bank profitability. Thus, the need for banks to build a better service culture which is customer focused is imperative. Customer attitude, defined as positive or negative predisposition towards a subject is an important concept in marketing research (Grossman & Till, 1998). Literature reviewed in this regard, showed the appropriateness to use sentiment analysis (SA) system to investigate customers' attitude.

In chapter 3, SA systems were reviewed, research work has been done at various levels of sentiment classification (document, sentence, and word). The most notable thing is that studies differ in terms of techniques, type of dataset and domain in which they have been conducted. Furthermore, studies also differ in text data pre-processing techniques. For example, dealing with stop-words, stemming, lemmatization, negation, and expansion of malformed words. Literature suggests there are two main approaches applied to SA namely, machine learning and lexical based approach. Unfortunately, both techniques have been understudied in the banking context. As illustrated in figure 3.0 of chapter 3, there is no lexicon or machine learning model validated in the bank domain. This implies best approach in terms of performance is yet to be known. This led to the development of sentiment lexicon "SentiLeye" to suit the Nigeria bank context in terms of lexicon-based approach. Despite the good performance of supervised models reported in previous studies. Literature suggests there is no annotated dataset available in the banking context to compare models. Unfortunately, supervised models suffer significant performance loss when domain boundaries are crossed (Mudinas et al. 2018). This thus evidence why there is lack of SA research work in the banking context. A simple way to use supervised model for SA application is to have the domain-specific annotated dataset available for training purpose. In summary, there is need to create annotated set to train supervised learning models to validate the best performed SA classification model in bank domain.

Similarly, literature suggest comparative study of topic modelling techniques for aspect extraction is limited. This motivated this study to consider a comprehensive comparison of the statistical topic models in this context. This method is preferred because statistical topic models have the advantage to extract both implicit and explicit aspects. Another interesting thing is that previous studies focused on using aspects extracted to improve the performance of classifiers by performing aspect-based sentiment analysis. There were very few studies found to have used the aspects extracted to investigate the customers' sentiment. To conclude, literature reviewed was not only used to unveil the research gap, but also used to adopt models compared for both sentiment analysis and aspect extraction.

9.1.2 Objective 2

> 2). To propose Topic-Sentiment Banking System (TSBS), a context-informed framework to understand customers' attitude towards up-to-date undefined aspects of banking. This will help detail approaches towards answering the research questions.

Chapter 4 focused on the development of sentiment analysis approaches and thus, proposed TSBS framework for this purpose. The framework comprises of three main components namely, lexicon-based approach, machine learning approach for the sentiment classification task and topic modelling for the aspect extraction task and the techniques applied were discussed.

Due to unavailable resources for SA in the Nigeria banking context, sentiment wordlist for the lexicon based approach was generated using both the corpus-based approach (tweets) and dictionary (WordNet) based approaches. The SentiLeye algorithm was thus developed which is one of the contributions of this chapter 4. Furthermore, in chapter 4, labelled Twitter dataset in the Nigeria banking context was created. This study employed the use of three human annotators to manually label tweets which were randomly selected from the bank customer dataset retrieved from Twitter. The annotated dataset contained 7,037 tweets and were split in ratio of 80:20 for training and testing set, respectively. The training set amounts to 5629 and test set amounts to 1408 tweets. The training set contained 4022 neutral, 1306 negative and 301 positive tweets. This training set were observed to have class imbalance problem and thus, employed the use of SMOTE (Lemaître et al. 2017) to balance the class distribution such that all classes have 4022 tweets. The supervised models employed were K-Nearest Neighbour, Naïve Bayes, Logistic Regression (Multinomial), Support Vector Machine, Random Forest, Gradient Boosting, Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory, And Bi-Directional Long Short-Term Memory. These classification models were selected based on their performance as reported from literature reviewed in chapter 3. Vector space models such as N-gram, TF-IDF and Word2Vec were used for feature representation of text. This helped to understand the influence of frequency-based vector models and the word embeddings on the classification models. The third component of the TSBS framework is the aspect extraction for sentiment analysis. Statistical topic modelling methods such as latent semantic indexing (LSI), latent Dirichlet allocation (LDA), and hierarchical Dirichlet process (HDP) were adopted since both explicit and implicit aspects are needed to form the banking themes in this context. In summary, the second objective was achieved by proposing topic-sentiment banking system (TSBS). The system integrates the sentiment model in chapter 6 into the topic model pipeline in chapter 7. Therefore, the TSBS was used to demonstrate the customers' sentiment towards the topics identified.

9.1.3 Objective 3

> 3). To develop a novel lexicon algorithm that captures the tone and semantic orientation of banking and Pidgin English terms. This will help establish one of the components needed for the TSBS and thus answer RQ2.

Since the performance of SA methods is not yet known in this context, it was necessary to build from scratch, lexical resources to address this gap identified. Thus, this study developed "SentiLeye" lexicon to detect bank specific terms with language such as Pidgin English and English terms put into consideration. Most importantly, the opinionated factual terms were considered during the development. This study benefitted from dictionaries such as WordNet, General Inquirer, and general-purpose lexicon Sentistrength to develop a robust lexicon classifier. Data were retrieved from Twitter for 3 months (from May 2019 to August 2019) to develop the bank domain specific corpus and subsequently retrieved their synonyms from the online English dictionaries. The most important part of the labelling was that the tone and semantic orientation of the terms were taking into consideration such that strong positive terms are labelled +2 to strong negative labelled -2. An intense manual labelling and inspection was used to annotate and validate the lexicon terms and result. This involved three main annotators. The manual annotation and validation were chosen because it has been shown to produce the best result compared to other methods (Boukes, 2020). During the experimentation, lexicons were updated to take the Pidgin and domain terms into consideration. it was found that local terms & context-based words significantly influence the performance of the lexicons (detailed in section 5.2). An explanation for this is that lexicon-based approach relies strongly on their wordlist, and thus, if word does not exist in the lexicon, the algorithm cannot capture the semantic orientation of such word. In summary, the main findings suggest:

- lexicons should consider domain specific terms because general purpose lexicons suffer in context. This is because they are developed from general lexical knowledge.

- lexicons should be updated with opinionated factual words. This helps in service industry where objective sentences can mean opinion which thus influences other people.
- there is need for lexical algorithms to pay attention to handling negation. It was observed during the experimentation that some popular lexicon did not pay attention to negation and thus, affected their performance when used in this context.
- there is need to create wordlist for non-English text.

The main contribution of chapter 5 was SentiLeye, a novel lexicon algorithm developed to capture Nigeria Pidgin English and English words in the banking domain. In addition, this study contributes by providing labelled dataset which is suitable to validate lexicons in Nigeria banking context. Lastly, this study provides sentiment Pidgin wordlist to encourage further research in Pidgin language processing.

9.1.4 Objective 4

> 4). To conduct a benchmark comparison of sentiment classification and topic modelling techniques at tweet-level. This will help establish other components needed for the TSBS and thus answer RQ1, RQ3 and RQ4.

In chapter 6, a thorough benchmark comparison of the supervised models was carried out. Beforehand, the exploratory data analysis (EDA) plots showed keywords like ATM, transaction, airtime, card, mobile app, reactivate account, internet banking, airtime transfer, credit card, branch banking, made payment, reverse money, USSD code, mobile banking, failed transaction, and account debited are the most frequent terms in the customers' tweets. Most of these terms are banking channels, while others are objective terms that indicate transaction type. The EDA suggest usage and popularity of the banking channels and transaction type respectively. Thus, the banks can explore these channels to gain competitive advantage. The evaluation of the classification models was done using metrics such as accuracy, precision, recall, and f1-measure. The comparative result presented showed support vector machine (SVM) was the best performed classifier with 82% accuracy and thus was used in implementing the sentiment analysis system. This result helped in answering the RQ3 to attest SVM as an off-the-shelf sentiment classification method in the banking context. In addition, SVM can serve as baseline model for future work. The sentiment classification test result was manually inspected to check for causes of misclassification. It was shown that the use of foreign language (Yoruba, Hausa, and Igbo), and opinionated-objective statement remains a problem. Thus, answered the RQ4. In general, predictions showed a high proportion of customers utilised the social media channel to request information while a concerning proportion are unhappy with their banking experience and a tiny proportion are happy.

In chapter 6, the sentiment classification result showed a significant proportion of the customers were unhappy. However, the analysis could not suggest to what aspect of banking the customers were unhappy at. This opens the need for aspect extraction for sentiment analysis in this context. This study employed the use of topic modelling technique to extract the aspects which customers tweeted about. The aspects were then used to investigate the customers' sentiment such that each tweet can be attributed to their aspect and sentiment polarity. In chapter 7, this study showed various means of extracting aspects, ranging from using noun frequency, syntactic relation, and dependencies, to statistical topic modelling. This study chose to use the statistical topic modelling approach because the techniques can detect both implicit and explicit aspects which is an advantage over other techniques. The part of speech (POS) tagging was used to detect nouns, and adverbs from the corpus with a threshold frequency. A comparison of these topic models (LSI, LDA, MALLET and HDP) was conducted to investigate the most performed model in this context. It was shown that the use of POS tagging and LDA produced a more interpretable terms within the topics with a reasonable coherence score. The findings in chapter 7 suggests customer service experience, investment bank, transaction problem, internet service performance, bank charges, bank events, contact/complaint form, mobile banking, scam alert/advice, and credit card payment are the key topics customers tweeted about. Some of these topics have been shown as the top tweeted terms during the EDA process in chapter 6. For example, credit card, and mobile banking. This highlights the importance of the topics.

To answer the RQ1, the result presented showed there are aspects of banking that need urgent attention. For example, customer service experience, transaction problem, and bank charges. This is because these topics were highly tweeted about, and significant proportion of the customers were unhappy with these topics. Some of these topics have been researched previously as a concerning aspects of banking. For example, customer service experience (Ugwuanyi et al. 2020), and bank charges (Haruna et al. 2018). Furthermore, there are some topics that customers demonstrated mixed feelings. This is a case where some customers are happy, while some are unhappy. For example, mobile banking, the significant of this aspect of banking was evidenced in the study of Inegbedion et al. (2019). The features of this banking channel that received positive tweets can thus be used to attract more customers while the negative feature can be improved. Overall, this study showed the significance of transaction problem. Transaction problem is the most significant topic in this study and the high percentage of the neutral and negative sentiment tweets suggests a lot of customers encountered transaction problems. While it was observed that some customers tweeted for a solution to their problem, others expressed their frustration to the bank service. Surprisingly, past studies did not pay attention to transaction problem explicitly. This will thus, be an interesting area for future research work. Lastly, in chapter 8, this study validated the research outcome using two main validation components namely, internal validation and external validation. The internal validity was achieved by the model performance metrics reported in section 6.1 for sentiment classification and section 7.2 and 7.5 for topic modelling and some of the findings have also been published in the literature as discussed in section 7.6. With respect to external validation, the result from the analysis of the participants' responses suggests the TSBS framework is useful, beneficial, and relevant. The result also showed the research findings are valid and can be generalised across the Nigeria banking sector.

9.2 Contributions

This study contributes to existing knowledge in many ways. The significance of the findings and contributions of this study can be summarised as follows.

I. The bank customers attitude had been investigated by previous studies using survey data and thus, the research findings are limited to pre-defined variables. In this modern age, social media usage is increasing exponentially. Therefore, this study contributes by providing up-to-date insights into customers' attitude towards undefined aspects of Nigeria banking using customers social media data.

II. In the field of sentiment analysis (SA), most studies found focused on dataset in official languages such as English, Spanish, and Chinese (Pontiki et al. 2016). This study contributes to the development of lexical resources for pidgin English including its variants within the three main ethnic groups in Nigeria.

III. Loughran & McDonald (2011) warned future researchers in the finance sector not to use any sentiment lexicon created and evaluated outside the domain for analysis. Krishnamoorthy (2018) also stated sentiment analysis tools accurately classify financial text better when performance measures of financial terms are considered. Yet, there is no available lexicon validated in the banking domain. Based on these assertions, this study contributes to fill this gap with the development of "SentiLeye" a novel domain-specific lexicon.

IV. Construction of standard annotated dataset (in Pidgin and/or English) in the banking context. This annotated dataset has been made publicly available.

V. Benchmark comparison of the sentiment classification models in the banking context.

VI. Comparison of the unsupervised statistical topic models in the banking context.

VII. Proposed TSBS framework to automatically uncover hidden topics and their sentiment.

9.3 Limitation & Future work

This study investigated customers attitude using social media data. The investigation was conducted successfully and thus, the report provides comprehensive knowledge of natural language processing, aspect extraction and sentiment analysis applied to the bank domain. However, the research work is limited. One of the limitations is that the dataset used does not represent the entire bank customers. This is because not all bank customers use social media. In addition, the models presented in this study are limited to the bank domain and thus needs to be retrained to use in a different domain. SentiLeye lexicon was developed specifically to suit bank domain and was not validated out of the domain. The lexicon was shown to perform better than general-purpose lexicons in chapter 5. This finding aligns with the study of Palmer et al. (2020) that showed domain specific lexicon outperforms general purpose lexicon in financial context. However, this limits SentiLeye to bank domain. Another limitation that will lead to future work is the word coverage in the lexicon. This should be ongoing and thus, be updated regularly due to the constantly changing words on social media. Generally, in this thesis, a major limitation of the classification models is the fact that the classification is at Tweet-level. It is worth noting that the assumption of Tweet-level classification is that words in a tweet share the same topic and/or sentiment. Unfortunately, this may not be the case all the time. Most importantly, the comparative study conducted were limited to the applied models. There are other models for future work such as Variational autoencoder (VAE) and Bidirectional encoder representations from transformer (BERT). These models need to be compared in this domain.

One of the major contributions of this study is the creation of lexical resources to suit Nigeria banking context. This puts into consideration the language, bank terms and social media data. This thus provides background knowledge and baseline performance for future work. Therefore, future work identified are as follows.

- In chapter 5, the Pidgin English terms used in the lexicon are of Nigeria pidgin English and should be extended to capture the variations in other countries like Ghana, Sierra Leone, and Cameroon.
- There is need to improve on the annotated dataset to enhance the performance of the classification models. This study utilised supervised machine and deep learning models. These models perform better with large training dataset as they can learn more. Therefore, the training set can be improved with Tweets that contains Nigeria local languages which is often mixed or used with the Nigeria Pidgin English.
- The machine learning model was shown to outperform the deep learning models. However, there are other classification models in both categories that were not used. Thus, future work will be to compare more sophisticated models such as BERT. In addition, the SentiLeye wordlist can also be used to enhance the training set or form a hybrid SA approach.
- In Chapter 7, unsupervised topic model was used for aspect extraction in this study. Alternatively, future work will be to annotate tweets to their topics, and thus, use supervised topic classification models for prediction.

References

Abbas, M., Memon, K. A., Jamali, A. A., Memon, S., & Ahmed, A. (2019). Multinomial Naive Bayes Classification Model for Sentiment Analysis. IJCSNS, 19(3), 62.

Abdul-Mageed, M., Diab, M. T., & Korayem, M. (2011). Subjectivity and sentiment analysis of modern standard Arabic. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2 (pp. 587-591). Association for Computational Linguistics.

Abercrombie, G. (2021). Topic-centric sentiment analysis of UK parliamentary debates (Doctoral dissertation, University of Manchester)

Adamu, H., Lutfi, S. L., Malim, N. H. A. H., Hassan, R., Di Vaio, A., & Mohamed, A. S. A. (2021). Framing Twitter Public Sentiment on Nigerian Government COVID-19 Palliatives Distribution Using Machine Learning. Sustainability, 13(6), 3497.

Adaramola, A. O., & Kolapo, F. T. (2019). Assessment of Bank Technology Machine and Mobile Banking as Market Strategies to Raising Performance of Banks in Nigeria. Journal of Economics and Behavioral Studies, 11(3 (J)), 108-115.

Addi, H. A., & Ezzahir, R. (2020). Sampling techniques for Arabic Sentiment Classification: A Comparative Study. In Proceedings of the 3rd International Conference on Networking, Information Systems & Security (pp. 1-6).

Adel, H., & Schütze, H. (2016). Exploring different dimensions of attention for uncertainty detection. arXiv preprint arXiv:1612.06549.

Aduaka, U., & Awolusi, O. D. (2020). Electronic Banking and Profitability in the Nigerian Banking Industry. Information Management and Business Review, 12(2 (I)), 20-37.

Afandi, M. A., & Muta'ali, A. (2019). Will Traditional Bank's Customers Switch to FinTech Lending? A Perspective of Push-Pull-Mooring Framework. In Proceedings of the 2019 11th International Conference on Information Management and Engineering (pp. 38-43).

Afolabi, I. T., Sowunmi, O. Y., & Adigun, T. (2019). Semantic Text Mining using Domain Ontology. In Roceedings of the World Congress on Engineering and Computer Science (pp. 1-6).

Agnihotri, D., Kulshreshtha, K., & Tripathi, V. (2021). Emergence of social media as new normal during COVID-19 pandemic: a study on innovative complaint handling procedures in the context of banking industry. International Journal of Innovation Science.

Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modeling? And how to fix it using search-based software engineering. Information and Software Technology, 98, 74-88.

Ahmad, M., Aftab, S., Muhammad, S. S., & Waheed, U. (2017). Tools and techniques for lexicon driven sentiment analysis: a review. Int. J. Multidiscip. Sci. Eng, 8(1), 17-23.

Aishatu, I. O., Lim, G. C., Olakojo, S. A., von Furstenberg, G. M., & Wang, L. F. (2017). Customer loyalty towards internet banking in Nigeria. Journal of Advances in Economics and Finance, 2(3), 149-159.

Ajisafe, R. A., & Ajide, F. M. (2014). Bank competition and economic growth: Evidence from Nigeria. Journal of Emerging Trends in Economics and Management Sciences, 5(5), 419-425.

Akhtar, N., Zubair, N., Kumar, A., & Ahmad, T. (2017). Aspect based sentiment-oriented summarization of hotel reviews. Procedia computer science, 115, 563-571.

Al Hamoud, A., Alwehaibi, A., Roy, K., & Bikdash, M. (2018). Classifying Political Tweets Using Naïve Bayes and Support Vector Machines. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 736-744). Springer, Cham.

Al Omari, M., Al-Hajj, M., Hammami, N., & Sabra, A. (2019). Sentiment classifier: logistic regression for Arabic services’ reviews in Lebanon. In 2019 International Conference on Computer and Information Sciences (ICCIS) (pp. 1-5). IEEE.

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2017). Arabic language sentiment analysis on health services. In 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR) (pp. 114-118). IEEE.

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. Front. Artif. Intell, 3, 42.

Alemneh, G. N., Rauber, A., & Atnafu, S. (2020). Corpus based Amharic sentiment lexicon generation. In Proceedings of the The Fourth Widening Natural Language Processing Workshop (pp. 1-3).

Ali, F., Kwak, D., Khan, P., El-Sappagh, S., Ali, A., Ullah, S., & Kwak, K. S. (2019). Transportation sentiment analysis using word embedding and ontology-based topic modeling. Knowledge-Based Systems, 174, 27-42.

Alsmadi, A. A., Sha'ban, M., & Al-Ibbini, O. A. (2019). The Relationship between E-banking Services and Bank Profit in Jordan for the Period of 2010-2015. In Proceedings of the 2019 5th International Conference on E-Business and Applications (pp. 70-74).

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. Journal of computational science, 27, 386-393.

Al-Smadi, M., Talafha, B., Al-Ayyoub, M., & Jararweh, Y. (2019). Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. International Journal of Machine Learning and Cybernetics, 10(8), 2163-2175.

Altintas, N. (2020). The impact of regulatory reforms on cost efficiency, ownership, and competition in the Turkish commercial banking sector (Doctoral dissertation, University of Reading).

Al-Wugayan, A. A. (2019). Relationship versus customer experience quality as determinants of relationship quality and relational outcomes for Kuwaiti retail banks. International Journal of Bank Marketing.

Alzamzami, F., Hoda, M., & El Saddik, A. (2020). Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation. IEEE Access.

Amaturo, E., & Punziano, G. (2017). Blurry Boundaries: Internet, Big-New Data, and Mixed-Method Approach. In Data Science and Social Research (pp. 35-55). Springer, Cham.

Amazon (2015). Amazon mechanical turk - welcome URL: https://www.mturk.com/

Amusa, L., Yahya, W. B., & Balogun, A. O. (2016). Data mining of Nigerians' sentiments on the administration of federal republic of Nigeria.

Andaleeb, S. S., Rashid, M., & Rahman, Q. A. (2016). A model of customer-centric banking practices for corporate clients in Bangladesh. International Journal of Bank Marketing, 34(4), 458-475

Anderson, A., Douglas, P. K., Kerr, W. T., Haynes, V. S., Yuille, A. L., Xie, J., & Cohen, M. S. (2014). Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. NeuroImage, 102, 207-219.

Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. Indian Journal of Psychological Medicine, 43(1), 86-88.

Anis, S., Saad, S., & Aref, M. (2020, October). Sentiment analysis of hotel reviews using machine learning techniques. In International Conference on Advanced Intelligent Systems and Informatics (pp. 227-234). Springer, Cham.

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems, 108, 110-124.

Aquino, A. M., & Chavez, E. P. (2018). Analysis on the use of Latent Semantic Indexing (LSI) for document classification and retrieval system of PNP files. In MATEC Web of Conferences (Vol. 189, p. 03009). EDP Sciences.

Aramburu, I. A., & Pescador, I. G. (2019). The effects of corporate social responsibility on customer loyalty: The mediating effect of reputation in cooperative banks versus commercial banks in the Basque country. Journal of Business Ethics, 154(3), 701-719.

Asghar, M. Z., Kundi, F. M., Ahmad, S., Khan, A., & Khan, F. (2018). T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme. Expert Systems, 35(1), e12233.

Asghar, M. Z., Sattar, A., Khan, A., Ali, A., Masud Kundi, F., & Ahmad, S. (2019). Creating sentiment lexicon for sentiment analysis in Urdu: The case of a resource-poor language. Expert Systems, e12397.

Asghari, M., Sierra-Sosa, D., & Elmaghraby, A. S. (2020). A topic modeling framework for spatio-temporal information management. Information Processing & Management, 102340.

Ashrafpour, N., Esfahlan, H. N., Aali, S., & Taghizadeh, H. (2021). The prerequisites and consequences of customers' online experience regarding the moderating role of brand congruity: evidence from an Iranian bank. Journal of Islamic Marketing.

Asuncion, A., Welling, M., Smyth, P., & Teh, Y. W. (2012). On smoothing and inference for topic models. arXiv preprint arXiv:1205.2662.

Ateeq, M., Habib, H., Umer, A., & Rehman, M. U. (2014). C++ or Python? Which One to Begin with: A Learner's Perspective. In 2014 International Conference on Teaching and Learning in Computing and Engineering (pp. 64-69). IEEE.

Athanasiou, V., & Maragoudakis, M. (2017). A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: a case study for modern Greek. Algorithms, 10(1), 34.

Axhiu, M., & Aliu, A. (2020) Aspect-term Extraction from Albanian Reviews with Topic Modeling Techniques.

Ayora, V., Horita, F., & Kamienski, C. (2021). Profiling Online Social Network Platforms: Twitter vs. Instagram. In Proceedings of the 54th Hawaii International Conference on System Sciences (p. 2792).

Aziz, N. A. A., Wen, L. J., Azman, H., & Sufian, A. (2020). Mobile Payment Adoption And Customer" s Attitude. European Journal of Molecular & Clinical Medicine, 7(3), 2771-2902.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. MIS quarterly, 40(4).

Balage Filho, P., & Pardo, T. (2013). NILC_USP: A hybrid system for sentiment analysis in twitter messages. In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (Vol. 2, pp. 568-572).

Balakrishnan, V., Selvanayagam, P. K., & Yin, L. P. (2020). Sentiment and Emotion Analyses for Malaysian Mobile Digital Payment Applications. In Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis (pp. 67-71).

Baltes, S., & Ralph, P. (2020). Sampling in software engineering research: A critical review and guidelines. arXiv preprint arXiv:2002.07764.

Banea, C., Mihalcea, R., & Wiebe, J. (2008). A Bootstrapping Method for Building Subjectivity Lexicons for Languages with Scarce Resources. In LREC (Vol. 8, pp. 2-764).

Bapat, D. (2021). Exploring the relationship between lifestyle, digital financial element and digital financial services experience. International Journal of Bank Marketing.

Barclays (2015), accessed at How do I connect my Twitter ID to my Pingit account? | Barclays on 29/12/2020

Barua, A., Thomas, S. W., & Hassan, A. E. (2014). What are developers talking about? an analysis of topics and trends in stack overflow. Empirical Software Engineering, 19(3), 619-654.

Barua, S., Islam, M. M., Yao, X., & Murase, K. (2012). MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning. IEEE Transactions on knowledge and data engineering, 26(2), 405-425.

Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. Procedia Engineering, 53(7), 453-462.

Batra, R., & Daudpota, S. M. (2018). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-5). IEEE.

Bensoltane, R., & Zaki, T. (2022). Towards Arabic aspect-based sentiment analysis: a transfer learning-based approach. Social Network Analysis and Mining, 12(1), 1-16.

Berry, M. W., & Browne, M. (2005). Email surveillance using non-negative matrix factorization. Computational & Mathematical Organization Theory, 11(3), 249-264.

Bertalan, V. G., & Ruiz, E. E. S. (2019). Using topic modeling to find main discussion topics in brazilian political websites. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (pp. 245-248).

Bhargava, M. G., & Rao, D. R. (2018). Sentimental analysis on social media data using R programming. International Journal of Engineering & Technology, 7(2.31), 80-84.

Bhattacharya, S., Srinivasan, P., & Polgreen, P. (2014). Engagement with health agencies on twitter. PLoS One, 9(11), e112235.

Binkley, D., Heinz, D., Lawrie, D., & Overfelt, J. (2014). Understanding LDA in source code analysis. In Proceedings of the 22nd international conference on program comprehension (pp. 26-36).

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analysing text with the natural language toolkit. " O'Reilly Media, Inc.".

Boillot, M. A. (2012). U.S. Patent No. 8,312,479. Washington, DC: U.S. Patent and Trademark Office.

Bouchlaghem, R., Elkhelifi, A., & Faiz, R. (2016). A machine learning approach for classifying sentiments in Arabic tweets. In Proceedings of the 6th international conference on web intelligence, mining and semantics (pp. 1-6).

Boukes, M. (2020) The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms.

Boukes, M., van de Velde, B., Araujo, T., & Vliegenthart, R. (2020). What's the tone? Easy doesn't do it: Analyzing performance and agreement between off-the-shelf sentiment analysis tools. Communication Methods and Measures, 14(2), 83-104.

Bowes, D., Hall, T., & Gray, D. (2012). Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix. In Proceedings of the 8th international conference on predictive models in software engineering (pp. 109-118).

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Brewis, J. (2014). The ethics of researching friends: On convenience sampling in qualitative management and organization studies. British Journal of Management, 25(4), 849-862.

Britz, D., Goldie, A., Luong, M. T., & Le, Q. (2017). Massive exploration of neural machine translation architectures. arXiv preprint arXiv:1703.03906.

Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. (2009). Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem.

In Pacific-Asia conference on knowledge discovery and data mining (pp. 475-482). Springer, Berlin, Heidelberg.

Cao, B., Wu, J., Wang, S., Gao, H., Fan, J., Deng, S., & Liu, X. (2021). Unsupervised Derivation of Keyword Summary for Short Texts. ACM Transactions on Internet Technology (TOIT), 21(2), 1-23.

Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing & Applications, 10(2), 184-192

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., & Zdeborová, L. (2019). Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), 045002.

Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. Applied Soft Computing, 50, 135-141.

Cecchetti, S. G., Schoenholtz, K. L., and Fackler, J. (2011), Money, Banking and Financial Markets. Global Edition (3rd), McGraw-Hill/Irwin, New York.

Chahal, H., & Dutta, K. (2015). Measurement and impact of customer experience in banking sector. Decision, 42(1), 57-70.

Chauhan, S., Akhtar, A., & Gupta, A. (2022). Customer experience in digital banking: a review and future research directions. International Journal of Quality and Service Sciences.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

Chen, F., Yuan, Z., & Huang, Y. (2020). Multi-source data fusion for aspect-level sentiment classification. Knowledge-Based Systems, 187, 104831.

Chen, P., Sun, Z., Bing, L., & Yang, W. (2017). Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 452-461).

Cheng, X., Xu, W., Wang, T., & Chu, W. (2018) Variational Semi-supervised Aspect-term Sentiment Analysis via Transformer. arXiv preprint arXiv:1810.10437.

Chien, J. T. (2017). Bayesian nonparametric learning for hierarchical and sparse topics. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(2), 422-435.

Chiluwa, I. (2016). West African English in digital discourse. Covenant Journal of Language Studies, 1(1).

Chollet, F. (2015). keras.

Chory, R. N., Nasrun, M., & Setianingsih, C. (2018). Sentiment Analysis on User Satisfaction Level of Mobile Data Services Using Support Vector Machine (SVM) Algorithm. In 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS) (pp. 194-200). IEEE.

Clark, J. L., Algoe, S. B., & Green, M. C. (2018). Social network sites and well-being: the role of social connection. Current Directions in Psychological Science, 27(1), 32-37.

Clematide, S., Klenner, M., Montoyo, A., Martínez-Barco, P., Balahur, A., & Boldrini, E. (2010). Evaluation and extension of a polarity lexicon for German.

Colón-Ruiz, C., & Segura-Bedmar, I. (2020). Comparing deep learning architectures for sentiment analysis on drug reviews. Journal of Biomedical Informatics, 110, 103539.

Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological), 20(2), 215-232.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. Behavior research methods, 49(3), 803-821.

Cruz-Roa, A., Díaz, G., Romero, E., & González, F. A. (2011). Automatic annotation of histopathological images using a latent topic model based on non-negative matrix factorization. Journal of pathology informatics, 2.

Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Information Processing & Management, 57(2), 102034.

Dai, L., Chen, H., & Li, X. (2011). Improving sentiment classification using feature highlighting and feature bagging. In 2011 IEEE 11th International Conference on Data Mining Workshops (pp. 61-66). IEEE.

Dandala, B., Joopudi, V., & Devarakonda, M. (2019). Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks. Drug safety, 42(1), 135-146.

Daniel, B. K. (2018). Reimaging research methodology as data science. Big data and cognitive computing, 2(1), 4.

Dao, B., Nguyen, T., Venkatesh, S., & Phung, D. (2017). Latent sentiment topic modelling and nonparametric discovery of online mental health-related communities. International Journal of Data Science and Analytics, 4(3), 209-231.

Daramola, O., Olajide, F., Adewumi, A. O., & Ayo, C. K. (2014). An experimental validation of public cloud mobile banking. Research Journal of Applied Sciences, Engineering and Technology, 7(24), 5304-5314.

Daub, O., Cunningham, B. J., Bagatto, M. P., Johnson, A. M., Kwok, E. Y., Smyth, R. E., & Cardy, J. O. (2021). Adopting a Conceptual Validity Framework for Testing in Speech-Language Pathology. American Journal of Speech-Language Pathology, 1-15.

Daud, N. I. M., Aziz, N. A. S., Mohd, F., & Sabri, N. A. A. (2021). The Influence of E-service Quality On Marketplace Platform Towards Customer Loyalty. Journal of Entrepreneurship and Business, 9(1), 1-11.

Dauda, S. Y., & Lee, J. (2015). Technology adoption: A conjoint analysis of consumers′ preference on future online banking services. Information Systems, 53, 1-15.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391-407.

Dermouche, M., Velcin, J., Khouas, L., & Loudcher, S. (2014). A joint model for topic-sentiment evolution over time. In 2014 IEEE International Conference on Data Mining (pp. 773-778). IEEE.

Deshmukh, A., Hegde, G., Lathi, R., & Govikarn, S. (2012). A literature survey on latent semantic indexing. International Journal of Engineering Inventions, 1(4), 2278-7461.

Devi, D. N., Kumar, C. K., & Prasad, S. (2016). A feature-based approach for sentiment analysis by using support vector machine. In 2016 IEEE 6th International Conference on Advanced Computing (IACC) (pp. 3-8). IEEE.

Devi, M. D., & Saharia, N. (2020). Exploiting Topic Modelling to Classify Sentiment from Lyrics. In International Conference on Machine Learning, Image Processing, Network Security and Data Sciences (pp. 411-423). Springer, Singapore.

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. Procedia Computer Science, 87, 44-49.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dewangan, J. K., Sharaff, A., & Pandey, S. (2020). Improving Topic Coherence Using Parsimonious Language Model and Latent Semantic Indexing. In ICDSMLA 2019 (pp. 823-830). Springer, Singapore.

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982.

Dey, S., Wasif, S., Tonmoy, D. S., Sultana, S., Sarkar, J., & Dey, M. (2020). A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews. In 2020 International Conference on Contemporary Computing and Applications (IC3A) (pp. 217-220). IEEE.

Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. Journal of Consumer Marketing, 34(6), 480-488.

Ding, Z., Qi, Y., & Lin, D. (2021, March). Albert-based sentiment analysis of movie review. In 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) (pp. 1243-1246). IEEE.

Dohaiha, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2018). Deep learning for aspect-based sentiment analysis: a comparative review. Expert Systems With Applications.

Eguchi, K., & Lavrenko, V. (2006). Sentiment retrieval using generative models. In Proceedings of the 2006 conference on empirical methods in natural language processing (pp. 345-354).

Eke, V. I., & Singhry, H. B. (2020). An Assessment of Bank Customers' intention to use internet banking: The Role of Service Quality. International Journal of Progressive Sciences and Technologies, 23(1), 424-434.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In machine learning in radiation oncology (pp. 3-11). Springer, Cham.

eMarketer. (July 17, 2017). Number of social media users worldwide from 2010 to 2021 (in billions) [Chart]. In Statista. Retrieved July 17, 2019, from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Emerson, R. W. (2015). Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? Journal of Visual Impairment & Blindness, 109(2), 164-168.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In LREC (Vol. 6, pp. 417-422).

Ethnologue (2020) https://www.ethnologue.com/guides/countries-most-languages

Ethnologue (2021) https://www.ethnologue.com/guides/how-many-languages

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics, 5(1), 1-4.

Evers, J. C. (2018). Current issues in qualitative data analysis software (QDAS): A user and developer perspective. The Qualitative Report, 23(13), 61-73.

Fallahi, A., & Jafari, S. (2011). An expert system for detection of breast cancer using data preprocessing and bayesian network. International Journal of Advanced Science and Technology, 34, 65-70.

Farrokhi, F., & Mahmoudi-Hamidabad, A. (2012). Rethinking convenience sampling: Defining quality criteria. Theory & Practice in Language Studies, 2(4).

Feng, G., Guo, J., Jing, B. Y., & Sun, T. (2015). Feature subset selection using naive Bayes for text classification. Pattern Recognition Letters, 65, 109-115.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. Journal of artificial intelligence research, 61, 863-905.

Fernando, J., Khodra, M. L., & Septiandri, A. A. (2019, September). Aspect and opinion terms extraction using double embeddings and attention mechanism for indonesian hotel reviews. In 2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA) (pp. 1-6). IEEE.

Findley, M. G., Kikuta, K., & Denly, M. (2021). External Validity. Annual Review of Political Science, 24, 365-393.

Fithriasari, K., Jannah, S. Z., & Reyhana, Z. (2020). Deep learning for social media sentiment analysis. MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics, 99-111.

Fu, X., Wei, Y., Xu, F., Wang, T., Lu, Y., Li, J., & Huang, J. Z. (2019). Semi-supervised Aspect-level Sentiment Classification Model based on Variational Autoencoder. Knowledge-Based Systems, 171, 81-92.

Gabarron, E., Dorronzoro, E., Rivera-Romero, O., & Wynn, R. (2019). Diabetes on Twitter: a sentiment analysis. Journal of diabetes science and technology, 13(3), 439-444.

Gandhi, H., & Attar, V. (2020). Extracting aspect terms using CRF and bi-LSTM models. Procedia Computer Science, 167, 2486-2495.

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering, 2(4), 42-47.

García, S., & Herrera, F. (2009). Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. Evolutionary computation, 17(3), 275-306.

García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowledge-Based Systems, 98, 1-29.

Garg, R., Rahman, Z., & Qureshi, M. N. (2014). Measuring customer experience in banks: scale development and validation. Journal of Modelling in Management.

Garg, S. B., & Subrahmanyam, V. V. (2021). A Survey on Various Negation Handling Techniques in Sentiment Analysis. Smart and Sustainable Intelligent Systems, 259-280.

Garzaro, D. M., Varotto, L. F., & de Carvalho Pedro, S. (2020). Internet and mobile banking: the role of engagement and experience on satisfaction and loyalty. International Journal of Bank Marketing.

Gautam, G., & Yadav, D. (2014). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh International Conference on Contemporary Computing (IC3) (pp. 437-442). IEEE.

Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of support vector machine algorithm in big data background. Mathematical Problems in Engineering, 2021.

Ghosh, D., & Guha, R. (2013). What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. Cartography and geographic information science, 40(2), 90-102.

Ghosh, S., Hazra, A., & Raj, A. (2020). A Comparative Study of Different Classification Techniques for Sentiment Analysis. International Journal of Synthetic Emotions (IJSE), 11(1), 49-57.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009.

Goel, A., Gautam, J., & Kumar, S. (2016). Real time sentiment analysis of tweets using Naive Bayes. In 2016 2nd International Conference on Next Generation Computing Technologies (NGCT) (pp. 257-261). IEEE.

Gohil, S., Vuik, S., & Darzi, A. (2018). Sentiment analysis of health care tweets: review of the methods used. JMIR public health and surveillance, 4(2), e43.

Gong, J., & Kim, H. (2017). RHSBoost: Improving classification performance in imbalance data. Computational Statistics & Data Analysis, 111, 1-13.

Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. Nature Reviews Molecular Cell Biology, 23(1), 40-55.

Guo, C., Yan, X., & Li, Y. (2020). Prediction of Student Attitude towards Blended Learning Based on Sentiment Analysis. In Proceedings of the 2020 9th International Conference on Educational and Information Technology (pp. 228-233).

Gupta, I., & Joshi, N. (2021). A Review on Negation Role in Twitter Sentiment Analysis. International Journal of Healthcare Information Systems and Informatics (IJHISI), 16(4), 1-19.

Habibabadi, S. K., & Haghighi, P. D. (2019). Topic Modelling for Identification of Vaccine Reactions in Twitter. In Proceedings of the Australasian Computer Science Week Multiconference (pp. 1-10).

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17, 26-32.

Hai, Z., Cong, G., Chang, K., Cheng, P., & Miao, C. (2017). Analysing sentiments in one go: a supervised joint topic modelling approach. IEEE Transactions on Knowledge and Data Engineering, 29(6), 1172-1185.

Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In International conference on intelligent computing (pp. 878-887). Springer, Berlin, Heidelberg.

Hasson, S. G., Piorkowski, J., & McCulloh, I. (2019). Social media as a main source of customer feedback: alternative to customer satisfaction surveys. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 829-832).

Hayashi Jr, P., Abib, G., & Hoppen, N. (2019). Validity in qualitative research: A processual approach. The Qualitative Report, 24(1), 98-112.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence) (pp. 1322-1328). IEEE.

He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2018). Exploiting document knowledge for aspect-level sentiment classification. arXiv preprint arXiv:1806.04346.

Hernando, A., Bobadilla, J., & Ortega, F. (2016). A non-negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model. Knowledge-Based Systems, 97, 188-202.

Hesse-Biber, S., & Johnson, R. B. (2013). Coming at things differently: Future directions of possible engagement with mixed methods research.

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. Computers & Education, 145, 103724.

Hingmire, S., Chougule, S., Palshikar, G. K., & Chakraborti, S. (2013). Document classification by topic labeling. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 877-880).

Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-based sentiment analysis using BERT. In Proceedings of the 22nd nordic conference on computational linguistics (pp. 187-196).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 50-57).

Hong, J., Nam, A., & Cai, A. (2019). Multi-Class Text Sentiment Analysis.

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in twitter. In Proceedings of the first workshop on social media analytics (pp. 80-88).

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177).

Hu, S., Liang, Y., Ma, L., & He, Y. (2009). MSMOTE: Improving classification performance when training data is imbalanced. In 2009 second international workshop on computer science and engineering (Vol. 2, pp. 13-17). IEEE.

Huang et al. (2019) used the contextual word representation from the pre-trained language model BERT with fine-tuning method to solve out of domain ABSA problem.

Huang, B., & Carley, K. M. (2019). Syntax-aware aspect level sentiment classification with graph attention networks. arXiv preprint arXiv:1909.02606.

Huang, C., Trabelsi, A., & Zaïane, O. R. (2019). ANA at SemEval-2019 Task 3: Contextual Emotion detection in Conversations through hierarchical LSTMs and BERT. arXiv preprint arXiv:1904.00132.

Huang, M., Xie, H., Rao, Y., Feng, J., & Wang, F. L. (2020). Sentiment strength detection with a context-dependent lexicon-based convolutional neural network. Information Sciences, 520, 389-399.

Huang, M., Xie, H., Rao, Y., Liu, Y., Poon, L. K., & Wang, F. L. (2020). Lexicon-Based Sentiment Convolutional Neural Networks for Online Review Analysis. IEEE Transactions on Affective Computing.

Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & operations research, 32(10), 2513-2522.

Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., ... & Zhang, J. (2021, April). Lstm based sentiment analysis for cryptocurrency prediction. In International Conference on Database Systems for Advanced Applications (pp. 617-621). Springer, Cham.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4), 330-338.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

ICICI (2013), accessed at ICICI Bank Pockets on Facebook | Feature on 29/12/2020

Inegbedion, H. E. (2018). Factors that influence customers' attitude toward electronic banking in Nigeria. Journal of Internet Commerce, 17(4), 325-338.

Inegbedion, H., Inegbedion, E. E., Osifo, S. J., Eze, S. C., Ayeni, A., & Akintimehin, O. (2019). Exposure to and usage of e-banking channels. Journal of Science and Technology Policy Management.

Internet World Stats. (2020). Most common languages used on the internet as of January 2020, by share of internet users [Graph]. In Statista. Retrieved October 01, 2020, from https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/

Iqbal, F., Hashmi, J. M., Fung, B. C., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. (2019). A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction. IEEE Access, 7, 14637-14652.

Isacenkova, J., Thonnard, O., Costin, A., Francillon, A., & Balzarotti, D. (2014). Inside the scam jungle: A closer look at 419 scam email operations. EURASIP Journal on Information Security, 2014(1), 4.

Islam, M. R., & Zibran, M. F. (2018). SentiStrength-SE: Exploiting domain specificity for improved sentiment analysis in software engineering text. Journal of Systems and Software, 145, 125-146.

Istia, S. S., & Purnomo, H. D. (2018). Sentiment analysis of law enforcement performance using support vector machine and K-nearest neighbor. In 2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE) (pp. 84-89). IEEE.

Jadav, B. M., & Vaghela, V. B. (2016). Sentiment analysis using support vector machine based on feature selection and semantic analysis. International Journal of Computer Applications, 146(13).

Jain, A. P., & Katkar, V. D. (2015). Sentiment analysis of Twitter data using data mining. In 2015 International Conference on Information Processing (ICIP) (pp. 807-810). IEEE.

Jakob, N., & Gurevych, I. (2010). Extracting opinion targets in a single and cross-domain setting with conditional random fields. In Proceedings of the 2010 conference on empirical methods in natural language processing (pp. 1035-1045).

Jangid, H., Singhal, S., Shah, R. R., & Zimmermann, R. (2018). Aspect-based financial sentiment analysis using deep learning. In Companion Proceedings of the The Web Conference 2018 (pp. 1961-1966).

Jayal, A., Lauria, S., Tucker, A., & Swift, S. (2011). Python for teaching introductory programming: A quantitative evaluation. Innovation in Teaching and Learning in Information and Computer Sciences, 10(1), 86-90.

Jiang, N., Tian, F., Li, J., Yuan, X., & Zheng, J. (2020). MAN: mutual attention neural networks model for aspect-level sentiment classification in SIoT. IEEE Internet of Things Journal, 7(4), 2901-2913.

Jin, W., Ho, H. H., & Srihari, R. K. (2009). OpinionMiner: a novel machine learning system for web opinion mining and extraction. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1195-1204).

Jo, Y., & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 815-824).

Johnson, R., & Zhang, T. (2016). Supervised and semi-supervised text categorization using LSTM for region embeddings. arXiv preprint arXiv:1602.02373.

Johnston, R., & Kong, X. (2011). The customer experience: a road-map for improvement. Managing Service Quality: An International Journal, 21(1), 5-24.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Kaity, M., & Balakrishnan, V. (2019). An automatic non-English sentiment lexicon builder using unannotated corpus. The Journal of Supercomputing, 1-26.

Kaji, N., & Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of HTML documents. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL) (pp. 1075-1083).

Kalaivani, P., & Shunmuganathan, K. L. (2013). Sentiment classification of movie reviews by supervised machine learning approaches. Indian Journal of Computer Science and Engineering, 4(4), 285-292.

Karami, A., & Elkouri, A. (2019). Political popularity analysis in social media. In International Conference on Information (pp. 456-465). Springer, Cham.

Kasri, M., Birjali, M., & Beni-Hssane, A. (2019). A comparison of features extraction methods for Arabic sentiment analysis. In Proceedings of the 4th International Conference on Big Data and Internet of Things (pp. 1-6). National Conference on Computer Science and Application Engineering (pp. 1-5).

Kassab, L., Kryshchenko, A., Lyu, H., Molitor, D., Needell, D., & Rebrova, E. (2020). On Nonnegative Matrix and Tensor Decompositions for COVID-19 Twitter Dynamics. arXiv preprint arXiv:2010.01600.

Kastrati, Z., Arifaj, B., Lubishtani, A., Gashi, F., & Nishliu, E. (2020). Aspect-Based Opinion Mining of Students' Reviews on Online Courses. In Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence (pp. 510-514).

Kastrati, Z., Kurti, A., & Imran, A. S. (2020). WET: Word embedding-topic distribution vectors for MOOC video lectures dataset. Data in brief, 28, 105090.

Kaur, P. C., Ghorpade, T., & Mane, V. (2016). Topic extraction and sentiment classification by using latent dirichlet markov allocation and SentiWordNet. In Proceedings of the International Conference on Advances in Information Communication Technology & Computing (pp. 1-6).

Kennedy, R. C., Xiang, X., Madey, G. R., & Cosimano, T. F. (2005). Verification and validation of scientific and economic models. In Proc. Agent (pp. 177-192).

Khan, F. H., Qamar, U., & Bashir, S. (2017). A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. Knowledge and information Systems, 51(3), 851-872.

Khan, Z., Iltaf, N., Afzal, H., & Abbas, H. (2020). Enriching non-negative matrix factorization with contextual embeddings for recommender systems. Neurocomputing, 380, 246-258.

Khanaferov, D., Luc, C., & Wang, T. (2014). Social network data mining using natural language processing and density based clustering. In 2014 IEEE International Conference on Semantic Computing (pp. 250-251). IEEE

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971.

Kherwa, P., & Bansal, P. (2020). A Comparative Empirical Evaluation of Topic Modeling Techniques. In International Conference on Innovative Computing and Communications (pp. 289-297). Springer, Singapore.

Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. Journal of Information Science, 44(4), 491-511.

Kim, S., Zhang, J., Chen, Z., Oh, A. H., & Liu, S. (2013). A hierarchical aspect-sentiment model for online reviews. In AAAI.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Ko, N., Jeong, B., Choi, S., & Yoon, J. (2017). Identifying product opportunities using social media mining: application of topic modeling and chance discovery theory. IEEE Access, 6, 1680-1693.

Komulainen, H., & Saraniemi, S. (2019). Customer centricity in mobile banking: a customer experience perspective. International Journal of Bank Marketing.

Koto, F., & Adriani, M. (2015). A comparative study on twitter sentiment analysis: Which features are good?. In International Conference on Applications of Natural Language to Information Systems (pp. 453-457). Springer, Cham.

Koulouri, T., Lauria, S., & Macredie, R. D. (2014). Teaching introductory programming: A quantitative evaluation of different approaches. ACM Transactions on Computing Education (TOCE), 14(4), 1-28.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150.

KPMG (2018). Payment Service Banks: The 'Challenger' Banks in Nigeria. https://assets.kpmg/content/dam/kpmg/ng/pdf/advisory/Payment-Service-banks.pdf

Krouska, A., Troussas, C., & Virvou, M. (2017). Comparative Evaluation of Algorithms for Sentiment Analysis over Social Networking Services. J. UCS, 23(8), 755-768.

Krupnikov, Y., Nam, H. H., & Style, H. (2021). Convenience samples in political science experiments. Advances in Experimental Political Science, 165.

Kudo, T., & Matsumoto, Y. (2001). Chunking with support vector machines. In Second Meeting of the North American Chapter of the Association for Computational Linguistics.

Kulkarni, A., Mandhane, M., Likhitkar, M., Kshirsagar, G., & Joshi, R. (2021). L3cubemahasent: A marathi tweet-based sentiment analysis dataset. arXiv preprint arXiv:2103.11408.

Kumar, B. S., & Ravi, V. (2017). One-class text document classification with OCSVM and LSI. In Artificial Intelligence and Evolutionary Computations in Engineering Systems (pp. 597-606). Springer, Singapore.

Kumar, P., Mokha, A. K., & Pattnaik, S. C. (2021). Electronic customer relationship management (E-CRM), customer experience and customer satisfaction: evidence from the banking industry. Benchmarking: An International Journal.

Kumari, S. S., & Babu, G. A. (2016). Sentiment on social interactions using linear and non-linear clustering. In 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB) (pp. 177-181). IEEE.

Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-based sentiment analysis in the social web. Journal of Basic and Applied Scientific Research, 4(6), 238-48.

Kušen, E., Cascavilla, G., Figl, K., Conti, M., & Strembeck, M. (2017). Identifying emotions in social media: comparison of word-emotion lexicons. 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), (pp. 132-137). IEEE.

Landry, M., Malouin, J. L., & Oral, M. (1983). Model validation in operations research. European Journal of Operational Research, 14(3), 207-220.

Lasser, J., Manik, D., Silbersdorff, A., Säfken, B., & Kneib, T. (2021). Introductory data science across disciplines, using Python, case studies, and industry consulting projects. Teaching Statistics, 43, S190-S200.

Lau, J. H., Cook, P., & Baldwin, T. (2013). unimelb: Topic modelling-based word sense induction for web snippet clustering. In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (pp. 217-221).

Lau, J. H., Cook, P., McCarthy, D., Newman, D., & Baldwin, T. (2012). Word sense induction for novel sense detection. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 591-601).

Law, D., Gruss, R., & Abrahams, A. S. (2017). Automated defect discovery for dishwasher appliances from online consumer reviews. Expert Systems with Applications, 67, 84-94.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), 788-791.

Lee, K. D., Han, K., & Myaeng, S. H. (2016). Capturing word choice patterns with LDA for fake review detection in sentiment analysis. In Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics (pp. 1-7).

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research, 18(1), 559-563.

Li, C. (2020). The Effect of Lexicon Based and Machine Learning Based Sentiment Analysis on the Accuracy of a Stock Market Prediction System.

Li, D., Zamani, S., Zhang, J., & Li, P. (2019). Integration of Knowledge Graph Embedding into Topic Modeling with Hierarchical Dirichlet Process. In Proceedings of the 2019

Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 940-950).

Li, F., Han, C., Huang, M., Zhu, X., Xia, Y., Zhang, S., & Yu, H. (2010). Structure-aware review mining and summarization. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010) (pp. 653-661).

Li, F., Huang, M., & Zhu, X. (2010). Sentiment analysis with global topics and local dependency. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 24, No. 1, pp. 1371-1376).

Li, H., Mukherjee, A., Si, J., & Liu, B. (2015). Extracting verb expressions implying negative opinions. In Twenty-Ninth AAAI Conference on Artificial Intelligence.

Li, T., Zhang, Y., & Sindhwani, V. (2009). A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (pp. 244-252).

Li, W., Shao, W., Ji, S., & Cambria, E. (2022). BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. Neurocomputing, 467, 73-82.

Li, X., Bing, L., Zhang, W., & Lam, W. (2019). Exploiting BERT for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:1910.00883.

Li, X., Bing, L., Zhang, W., & Lam, W. (2019). Exploiting BERT for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:1910.00883.

Liao, W., Zeng, B., Yin, X., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Applied Intelligence, 51(6), 3522-3533.

Lin, C., & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. In Proceedings of the 18th ACM conference on Information and knowledge management (pp. 375-384).

Lin, C., He, Y., Everson, R., & Ruger, S. (2011). Weakly supervised joint sentiment-topic detection from text. IEEE Transactions on Knowledge and Data engineering, 24(6), 1134-1145.

Lin, S., Yang, S., Ma, M., & Huang, J. (2018). Value co-creation on social media: Examining the relationship between brand engagement and display advertising effectiveness for Chinese hotels. International Journal of Contemporary Hospitality Management.

Lin, X. M., Ho, C. H., Xia, L. T., & Zhao, R. Y. (2021). Sentiment analysis of low-carbon travel APP user comments based on deep learning. Sustainable Energy Technologies and Assessments, 44, 101014.

Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, 2(2010), 627-666.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Liu, B. (2015). Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge university press.

Liu, B. (2020). Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge university press.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415-463). Springer, Boston, MA.

Liu, Q., Zhang, H., Zeng, Y., Huang, Z., & Wu, Z. (2018). Content attention model for aspect-based sentiment analysis. In Proceedings of the 2018 World Wide Web Conference (pp. 1023-1032).

Liu, R. (2020). Stock Selection Strategy Based on Support Vector Machine. In 2020 The 3rd International Conference on Machine Learning and Machine Intelligence (pp. 10-13).

Liu, S., Lee, K., & Lee, I. (2020). Document-level multi-topic sentiment classification of email data with BiLSTM and data augmentation. Knowledge-Based Systems, 105918.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Longadge, R., & Dongre, S. (2013). Class imbalance problem in data mining review. arXiv preprint arXiv:1305.1707.

Loureiro, S. M. C., & Ferreira, E. S. (2017). How customer experience attributes influence customer satisfaction and online bank credibility. In 10th Annual Conference of the EuroMed-Academy-of-Business (pp. 838-856). Euromed press.

Luo, M., Nie, F., Chang, X., Yang, Y., Hauptmann, A., & Zheng, Q. (2017). Probabilistic non-negative matrix factorization and its robust extensions for topic modeling. In Thirty-first AAAI conference on artificial intelligence.

Lutz, M. (2001). Programming python. " O'Reilly Media, Inc.".

Lv, Y. H., Liu, Q., & Jiang, C. B. (2021, March). Sentiment analysis measure of topic mapping in LDA-LSTM combination model. In IOP Conference Series: Earth and Environmental Science (Vol. 702, No. 1, p. 012015). IOP Publishing.

M'sik, B., & Casablanca, B. M. (2020). Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus. International Journal, 9(4).

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. arXiv preprint arXiv:1709.00893.

Ma, L., & Fan, S. (2017). CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. BMC bioinformatics, 18(1), 169.

Ma, T., Qu, D., Ma, R., Feng, W., & Li, K. (2016). Online topic evolution modeling based on hierarchical Dirichlet Process. In 2016 IEEE First International Conference on Data Science in Cyberspace (DSC) (pp. 400-405). IEEE.

Ma, X., Zeng, J., Peng, L., Fortino, G., & Zhang, Y. (2019). Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis. Future Generation Computer Systems, 93, 304-311.

Ma, Y., Peng, H., & Cambria, E. (2018). Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In Thirty-Second AAAI Conference on Artificial Intelligence.

Ma, Y., Peng, H., & Cambria, E. (2018, April). Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

Machová, K., Mikula, M., Gao, X., & Mach, M. (2020). Lexicon-Based Sentiment Analysis Using Particle Swarm Optimization. Electronics, 9(8), 1317.

Madhavaram, S., & Appan, R. (2010). The potential implications of web-based marketing communications for consumers' implicit and explicit brand attitudes: A call for research. Psychology & Marketing, 27(2), 186-202.

Mai, L., & Le, B. (2020). Joint sentence and aspect-level sentiment analysis of product comments. Annals of Operations Research, 1-21.

Maldonado, S., Weber, R., & Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using support vector machines. Information sciences, 286, 228-246.

Martin-Domingo, L., Martín, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ) Journal of Air Transport Management

Mbama, C. I., & Ezepue, P. O. (2018). Digital banking, customer experience and bank financial performance: UK customers' perceptions. International Journal of Bank Marketing, 36(2), 230-255.

Mbama, C. I., & Ezepue, P. O. (2018). Digital banking, customer experience and bank financial performance: UK customers' perceptions. International Journal of Bank Marketing.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. Retrieved from http://mallet.cs.umass.edu/

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. In AAAI-98 workshop on learning for text categorization (Vol. 752, No. 1, pp. 41-48).

McKinney, W. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51-56).

Mehmood, Y., & Balakrishnan, V. (2020). An enhanced lexicon-based approach for sentiment analysis: a case study on illegal immigration. Online Information Review.

Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C. (2007). Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of the 16th international conference on World Wide Web (pp. 171-180).

Melo, P. F., Dalip, D. H., Junior, M. M., Gonçalves, M. A., & Benevenuto, F. (2019). 10SENT: A stable sentiment analysis method based on the combination of off-the-shelf approaches. Journal of the Association for Information Science and Technology.

Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.

Milu, S. A., Emon, M. I. S., Ahmed, S. S., Alam, M. n J., Mahtab, S. S., Bhuiyan, J. A., & Hasan, M. (2020). Sentiment Analysis of Bengali Reviews for Data and Knowledge Engineering: A Bengali Language Processing Approach. In International Conference on Communication, Computing and Electronics Systems (pp. 81-91). Springer, Singapore.

Minaee, S., Azimi, E., & Abdolrashidi, A. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206.

Misra, S., & Li, H. (2019). Noninvasive fracture characterization based on the classification of sonic wave travel times. Machine Learning for Subsurface Characterization, 243.

Miyato, T., Dai, A. M., & Goodfellow, I. (2016). Adversarial training methods for semi-supervised text classification. arXiv preprint arXiv:1605.07725.

Moghaddam, S., & Ester, M. (2011). ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 665-674).

Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (pp. 26-34). Association for Computational Linguistics.

Mokha, A. K., & Kumar, P. (2022). Examining the Interconnections Between E-CRM, Customer Experience, Customer Satisfaction and Customer Loyalty: A Mediation Approach. Journal of Electronic Commerce in Organizations (JECO), 20(1), 1-21.

Moliner-Tena, M. A., Monferrer-Tirado, D., & Estrada-Guillén, M. (2019). Customer engagement, non-transactional behaviors and experience in services: A study in the bank sector. International Journal of Bank Marketing.

Moniruzzaman, M., Bagirov, A., & Gondal, I. (2020). Partial Undersampling of Imbalanced Data for Cyber Threats Detection. In Proceedings of the Australasian Computer Science Week Multiconference (pp. 1-4).

Mostafa, L. (2020). Machine Learning-Based Sentiment Analysis for Analyzing the Travelers Reviews on Egyptian Hotels. In Joint European-US Workshop on Applications of Invariance in Computer Vision (pp. 405-413). Springer, Cham.

Moussa, M. E., Mohamed, E. H., & Haggag, M. H. (2020). A generic lexicon-based framework for sentiment analysis. International Journal of Computers and Applications, 42(5), 463-473.

Mudinas, A., Zhang, D., & Levene, M. (2018). Bootstrap domain-specific sentiment classifiers from unlabeled corpora. Transactions of the Association for Computational Linguistics, 6, 269-285.

Muhammad, A., Wiratunga, N., Lothian, R., & Glassey, R. (2013). Domain-Based Lexicon Enhancement for Sentiment Analysis. In SMA@ BCS-SGAI (pp. 7-18)

Mukherjee, P., Badr, Y., Doppalapudi, S., Srinivasan, S. M., Sangwan, R. S., & Sharma, R. (2021). Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection. Procedia Computer Science, 185, 370-379.

Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In Proceedings of the 2004 conference on empirical methods in natural language processing (pp. 412-418).

Munikar, M., Shakya, S., & Shrestha, A. (2019). Fine-grained sentiment classification using bert. In 2019 Artificial Intelligence for Transforming Business and Society (AITB) (Vol. 1, pp. 1-5). IEEE.

Munir, S., Wasi, S., & Jami, S. I. (2019). A Comparison of Topic Modelling Approaches for Urdu Text. Indian Journal of Science and Technology, 12, 45.

Mwanajimba, Z. N. (2019). Attitude toward self-service technologies among retail customers of commercial banks in Tanzania: The case of National Micro-finance Bank (Doctoral dissertation, The University of Dodoma).

Nakamura, M., Kajiwara, Y., Otsuka, A., & Kimura, H. (2013). Lvq-smote–learning vector quantization based synthetic minority over–sampling technique for biomedical data. BioData mining, 6(1), 16.

Nandal, N., Tanwar, R., & Pruthi, J. (2020). Machine learning based aspect level sentiment analysis for Amazon products. Spatial Information Research, 1-7.

Nguyen, T. H., & Shirai, K. (2015). Topic modeling based sentiment analysis on social media for stock market prediction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1354-1364).

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.

Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

Nurifan, F., Sarno, R., & Sungkono, K. R. (2019). Aspect based sentiment analysis for restaurant reviews using hybrid elmo-wikipedia and hybrid expanded opinion lexicon-senticircle. International Journal of Intelligent Engineering and Systems, 12(6), 47-58.

Ogbuju, E., Oladipo, F., Yemi-Petters, V., Abdumalik, R., Olowolafe, T., & Aliyu, A. (2020). Sentiment Analysis of the Nigerian Nationwide Lockdown Due to COVID19 Outbreak. Available at SSRN 3665975.

Ogueji, K., & Ahia, O. (2019). Pidginunmt: Unsupervised neural machine translation from west african pidgin to english. arXiv preprint arXiv:1912.03444.

Olokoyo, F. O. (2013). Bank reforms and perfomance of banks in Nigeria. International Journal of Education and Research, 1(9).

Olokoyo, O. F. (2018). The effect of bank deregulation on bank performance in Nigeria. IJAME.

Olsson, F., & Handmark, G. (2020). Sentiment Analysis and Aspect Extraction for Business Intelligence. LU-CS-EX.

Orife, I. (2020). Towards Neural Machine Translation for Edoid Languages. arXiv preprint arXiv:2003.10704.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in human behaviour, 31, 527-541.

Oskamp, S., & Schultz, P. W. (2005). Attitudes and opinions. Psychology Press.

Osuna, E., Freund, R., & Girosit, F. (1997, June). Training support vector machines: an application to face detection. In Proceedings of IEEE computer society conference on computer vision and pattern recognition (pp. 130-136). IEEE.

Oyelami, L. O., Adebiyi, S. O., & Adekunle, B. S. (2020). Electronic payment adoption and consumers' spending growth: empirical evidence from Nigeria. Future Business Journal, 6(1), 1-14.

Oyewusi, W. F., Adekanmbi, O., & Akinsande, O. (2020). Semantic Enrichment of Nigerian Pidgin English for Contextual Sentiment Classification. arXiv preprint arXiv:2003.12450.

Oyeyemi, S. O., Gabarron, E., & Wynn, R. (2014). Ebola, Twitter, and misinformation: a dangerous combination?. Bmj, 349, g6178.

Özdemir, R. S., Louis, K. O. S., & Topbaş, S. (2011). Public attitudes toward stuttering in Turkey: Probability versus convenience sampling. Journal of fluency disorders, 36(4), 262-267.

Ozyurt, B., & Akcayol, M. A. (2021). A new topic modelling based approach for aspect extraction in aspect-based sentiment analysis: SS-LDA. Expert Systems with Applications, 168, 114231.

Palmer, M., Roeder, J., & Muntermann, J. (2020). Towards Automated Analysis of Financial Analyst Communication: The induction of a Domain-Specific Sentiment Dictionary. In Proceedings of the 28th European Conference on Information Systems (ECIS).

Pandey, U., & Chakravarty, S. (2010). A survey on text classification techniques for e-mail filtering. In 2010 Second International Conference on Machine Learning and Computing (pp. 32-36). IEEE.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational Linguistics (p. 271). Association for Computational Linguistics.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.

Panichella, A., Dit, B., Oliveto, R., Di Penta, M., Poshynanyk, D., & De Lucia, A. (2013). How to effectively use topic models for software engineering tasks? an approach based on genetic algorithms. In 2013 35th International Conference on Software Engineering (ICSE) (pp. 522-531). IEEE.

Parikh, R., & Movassate, M. (2009). Sentiment analysis of user-generated twitter updates using various classification techniques. CS224N Final Report, 118.

Park, S., Lee, J., & Kim, K. (2019). Semi-supervised distributed representations of documents for sentiment analysis. Neural Networks, 119, 139-150.

Patel, R., & Passi, K. (2020). Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. IoT, 1(2), 218-239.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Perez, F., Granger, B. E., & Hunter, J. D. (2010). Python: an ecosystem for scientific computing. Computing in Science & Engineering, 13(2), 13-21.

Perez-Rosas, V., Banea, C., & Mihalcea, R. (2012). Learning Sentiment Lexicons in Spanish. In LREC (Vol. 12, p. 73).

Perkins, J. (2010) Python Text Processing with NLTK 2.0 Cookbook; Packt Publishing Ltd.: Birmingham, UK.

Polignano, M., Basile, P., de Gemmis, M., & Semeraro, G. (2019). A comparison of word-embeddings in emotion detection from text using bilstm, cnn and self-attention. In Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (pp. 63-68).

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., & Hoste, V. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30).

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems, 108, 42-49.

Poria, S., Chaturvedi, I., Cambria, E., & Bisio, F. (2016, July). Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In 2016 international joint conference on neural networks (IJCNN) (pp. 4465-4473). IEEE.

Poushter, J., Bishop, C., & Chwe, H. (2018), Social media use continues to rise in developing countries but plateaus across developed ones. Pew Research Center, 22.

Pudaruth, S. (2021). Customer experiences and a conceptual model proposed for retail banking experiences in bank settings of Mauritius. Academy of marketing studies journal, 25(1), 1-13.

Purnamasari, P. D., Taqiyuddin, M., & Ratna, A. A. P. (2017). Performance comparison of text-based sentiment analysis using recurrent neural network and convolutional neural network. In Proceedings of the 3rd International Conference on Communication and Information Processing (pp. 19-23).

Qi, B., Costin, A., & Jia, M. (2020). A framework with efficient extraction and analysis of Twitter data for evaluating public opinions on transportation services. Travel Behaviour and Society, 21, 10-23

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. European Journal of Education Studies.

Rahman, M. M., & Wang, H. (2016). Hidden topic sentiment model. In Proceedings of the 25th International Conference on World Wide Web (pp. 155-165).

Ramaswamy, s., Khande, r., Patil, y. S., & Kalkar, p. (2021). Adapting digital banking services during covid-19 pandemic: a study on customer experiences in selected district of maharashtra state, India. Journal of electrical engineering and technology (ijeet), 12(5), 8-17.

Ramentol, E., Caballero, Y., Bello, R., & Herrera, F. (2012). SMOTE-RSB*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. Knowledge and information systems, 33(2), 245-265.

Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., & Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. Applied Soft Computing, 74, 634-642.

Rasjid, Z. E., & Setiawan, R. (2017). Performance comparison and optimization of text document classification using k-NN and naïve bayes classification techniques. Procedia computer science, 116, 107-112.

Rasool, A., Tao, R., Kashif, K., Khan, W., Agbedanu, P., & Choudhry, N. (2020). Statistic Solution for Machine Learning to Analyze Heart Disease Data. In Proceedings of the 2020 12th International Conference on Machine Learning and Computing (pp. 134-139).

Ray, P., & Chakrabarti, A. (2020). A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Applied Computing and Informatics.

RBC (2013), accessed at RBC first North American Financial Institution to bring person-to-person electronic money transfers to Facebook Messenger - RBC on 29/12/2020

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks.

Řehůřek, R., & Sojka, P. (2011). Gensim—statistical semantics in python. Retrieved from genism. org.

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In Proceedings of the 20th international conference on machine learning (ICML-03) (pp. 616-623).

Robinson, D. (2017). The incredible growth of Python. Stack Overflow—Sep, 6.

Romānova, I., & Kudinska, M. (2016). Banking and Fintech: a challenge or opportunity?. In Contemporary issues in finance: Current challenges from across Europe. Emerald Group Publishing Limited.

Ruder, S., Ghaffari, P., & Breslin, J. G. (2016). A hierarchical model of reviews for aspect-based sentiment analysis. arXiv preprint arXiv:1609.02745.

Sagala, T. W., Saputri, M. S., Mahendra, R., & Budi, I. (2020). Stock Price Movement Prediction Using Technical Analysis and Sentiment Analysis. In Proceedings of the 2020 2nd Asia Pacific Information Technology Conference (pp. 123-127).

Said, A. M., & Muqrashi, A. S. A. (2020). Hotel Reviews Analysis based on Sentiment Classification: Oman Case Study. In Proceedings of the 2020 6th International Conference on Computer and Technology Applications (pp. 144-148).

Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5-19.

Sak, H., Senior, A. W., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling.

Sakhaei, F., Afshari, A., & Esmaili, E. (2014). The impact of service quality on customer satisfaction in Internet banking. Journal of mathematics and computer science, 9(1), 33-40.

Salehin, S. S., Miah, R., & Islam, M. S. (2020). A Comparative Sentiment Analysis On Bengali Facebook Posts. In Proceedings of the International Conference on Computing Advancements (pp. 1-8).

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613-620.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers," IBM Journal 3, 211-229. Reprinted (with minor additions and corrections) in Computers and Thought, edited by Feigenbaum and Feldman, McGrawHill, 1963

Samuel, A. L. (1967). Some studies in machine learning using the game of checkers. II—Recent progress. IBM Journal of research and development, 11(6), 601-617.

Santoso, H. A., Rachmawanto, E. H., Nugraha, A., Nugroho, A. A., & Basuki, R. S. (2020). Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization. Telkomnika, 18(2), 799-806.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 1-21.

Sazzed, S. (2020). Development of Sentiment Lexicon in Bengali utilizing Corpus and Cross-lingual Resources. In 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 237-244). IEEE.

Sedgwick, P. (2013). Convenience sampling. Bmj, 347.

Seo, S., Kim, C., Kim, H., Mo, K., & Kang, P. (2020). Comparative study of deep learning-based sentiment classification. IEEE Access, 8, 6861-6875.

Shaikh, A. A., & Karjaluoto, H. (2015). Mobile banking adoption: A literature review. Telematics and informatics, 32(1), 129-142.

Shaikh, A.A., Alharthi, M.D. and Alamoudi, H.O. (2020), "Examining key drivers of consumer experience with (non-financial) digital services- an exploratory study", Journal of Retailing and Consumer Services, Vol. 55 No. 2020, p. 102073.

Sharma, A., & Dey, S. (2012). A comparative study of feature selection and machine learning techniques for sentiment analysis. In Proceedings of the 2012 ACM research in applied computation symposium (pp. 1-7).

Shi, T., Kang, K., Choo, J., & Reddy, C. K. (2018). Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In Proceedings of the 2018 World Wide Web Conference (pp. 1105-1114).

Shijia, E., Yang, L., Zhang, M., & Xiang, Y. (2018). Aspect-based financial sentiment analysis with deep neural networks. In Companion Proceedings of the The Web Conference 2018 (pp. 1951-1954). International World Wide Web Conferences Steering Committee.

Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2019). Sentence Level Sentiment Identification and Calculation from News Articles Using Machine Learning Techniques. In Computing, Communication and Signal Processing (pp. 371-376). Springer, Singapore.

Shoukry, A., & Rafea, A. (2012). Sentence-level Arabic sentiment analysis. In 2012 International Conference on Collaboration Technologies and Systems (CTS) (pp. 546-550). IEEE.

Shrestha, D., Wenan, T., Rajkarnikar, N., & Jeong, S. R. (2020). Consumers attitude towards Internet banking services in an underdeveloped country: A case of Pokhara, Nepal. Journal of Internet Computing and Services, 21(5), 75-85.

Shrivastava, S., Jeyanthi, P. M., & Singh, S. (2020). Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting. Cogent Economics & Finance, 8(1), 1729569.

Siddiqua, U. A., Ahsan, T., & Chy, A. N. (2016). Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis. In 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET) (pp. 1-4). IEEE.

Silva, C., & Ribeiro, B. (2003). The importance of stop word removal on recall values in text categorization. In Proceedings of the International Joint Conference on Neural Networks, 2003. (Vol. 3, pp. 1661-1666). IEEE.

Singh, P. K. (2021). Deep Learning Approach for Negation Handling in Sentiment Analysis. IEEE Access.

Singh, R., & Goel, V. (2019). Various Machine Learning Algorithms for Twitter Sentiment Analysis. In Information and Communication Technology for Competitive Strategies (pp. 763-772). Springer, Singapore.

Sohangir, S., Petty, N., & Wang, D. (2018). Financial Sentiment Lexicon Analysis. In 2018 IEEE 12th International Conference on Semantic Computing (ICSC) (pp. 286-289). IEEE.

Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. Journal of Big Data, 5(1), 3.

Song, M., Park, H., & Shin, K. S. (2019). Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. Information Processing & Management, 56(3), 637-653.

Srijith, P. K., Hepple, M., Bontcheva, K., & Preotiuc-Pietro, D. (2017). Sub-story detection in Twitter with hierarchical Dirichlet processes. Information Processing & Management, 53(4), 989-1003.

Statista. (2020). Number of social network users worldwide from 2017 to 2025 (in millions) [Chart]. In Statista. Retrieved September 24, 2020, from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Statista. (2021). Number of social network users in Nigeria from 2017 to 2026 (in millions) [Chart]. In Statista. Retrieved October 23, 2021, from https://www.statista.com/statistics/972907/number-of-social-network-users-in-nigeria/

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. International journal of information management, 39, 156-168.

Su, J., Yu, S., & Luo, D. (2020). Enhancing aspect-based sentiment analysis with capsule network. IEEE Access, 8, 100551-100561.

Suen, L. J. W., Huang, H. M., & Lee, H. H. (2014). A comparison of convenience sampling and purposive sampling. Hu Li Za Zhi, 61(3), 105.

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588.

Sun, D. L., & Fevotte, C. (2014). Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence. In 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 6201-6205). IEEE.

Sun, H., Jiang, T., & Dai, Y. (2019). Sentiment analysis of commodity reviews based on multilayer LSTM network. In Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing (pp. 1-5).

Sundermeyer, M., Schlüter, R., & Ney, H. (2012). LSTM neural networks for language modeling. In Thirteenth annual conference of the international speech communication association.

Supakkul, S., Ahn, R., Junior, R. G., Villarreal, D., Zhao, L., Hill, T., & Chung, L. (2020) Validating goal-oriented hypotheses of business problems using machine learning: An exploratory study of customer churn. In International Conference on Big Data (pp. 144-158). Springer, Cham.

Suresh, H. (2016). An unsupervised fuzzy clustering method for twitter sentiment analysis. In 2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS) (pp. 80-85). IEEE.

Sweidan, A. H., El-Bendary, N., & Al-Feel, H. (2021). Sentence-level aspect-based sentiment analysis for classifying adverse drug reactions (ADRs) using hybrid ontology-XLNet transfer learning. IEEE Access, 9, 90828-90846.

Syed, A. Z., Aslam, M., & Martinez-Enriquez, A. M. (2010). Lexicon based sentiment analysis of Urdu text using SentiUnits. In Mexican International Conference on Artificial Intelligence (pp. 32-43). Springer, Berlin, Heidelberg.

Symeonidis, S., Effrosynidis, D., & Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. Expert Systems with Applications, 110, 298-310.

Szmigin, I., & Piacentini, M. (2018). Consumer behaviour. Oxford University Press.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.

Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. Expert Systems with applications, 34(4), 2622-2629.

Tang, D., Qin, B., & Liu, T. (2015). Document modelling with gated recurrent neural network for sentiment classification. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432).

Tang, D., Qin, B., & Liu, T. (2016). Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900.

Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: a transfer learning based approach. Journal of Big Data, 7(1), 1-26.

Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: a transfer learning based approach. Journal of Big Data, 7(1), 1-26.

Tao, J., Thakker, U., Dasika, G., & Beu, J. (2019). Skipping RNN State Updates without Retraining the Original Model. In Proceedings of the 1st Workshop on Machine Learning on Edge in Sensor Systems (pp. 31-36).

Tashakkori, A., & Teddlie, C. (Eds.). (2010). Sage handbook of mixed methods in social & behavioral research. sage.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2005). Sharing clusters among related groups: Hierarchical Dirichlet processes. In Advances in neural information processing systems (pp. 1385-1392).

Thakor, A. V. (2020). Fintech and banking: What do we know?. Journal of Financial Intermediation, 41, 100833.

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In Cyberemotions (pp. 119-134). Springer, Cham.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163-173

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544-2558.

Thongtan, T., & Phienthrakul, T. (2019). Sentiment classification using document embeddings trained with cosine similarity. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop (pp. 407-414).

Titov, I., & McDonald, R. (2008). A joint model of text and aspect ratings for sentiment summarization. In proceedings of ACL-08: HLT (pp. 308-316).

Todd, K., Lapointe, A. P., & Broglio, S. P. (2019). Sentiment Analysis of Journal Articles and News Articles Pertaining to CTE. Archives of Clinical Neuropsychology, 34(5), 738-738.

Tripathi, G., & Naganna, S. (2015). Feature selection and classification approach for sentiment analysis. Machine Learning and Applications: An International Journal, 2(2), 1-16.

Troussas, C., & Virvou, M. (2020). Advances in Social Networking-based Learning: Machine Learning-based User Modelling and Sentiment Analysis (Vol. 181). Springer Nature.

Troya, A., Gopalakrishna Pillai, R., Rodriguez Rivero, D. C., Genc, D. Z., Kayal, D. S., & Araci, D. (2021, December). Aspect-Based Sentiment Analysis of Social Media Data With Pre-Trained Language Models. In 2021 5th International Conference on Natural Language Processing and Information Retrieval (NLPIR) (pp. 8-17).

Turner, Z., Labille, K., & Gauch, S. (2020). Lexicon-Based Sentiment Analysis for Stock Movement Prediction. International Journal of Mechanical and Industrial Engineering, 14(5), 185-191.

Twitter. (April 23, 2019). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions) [Chart]. In Statista. Retrieved July 16, 2019, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Udenze, S., Onwuliri, E. C., & Ugoala, B. (2020). Awareness and use of Whatsapp for banking and financial services: a study of social media users in north-central Nigeria. Nnamdi Azikiwe University Journal of Communication and Media Studies, 1(1).

Udofot, I., & MBARACHI, C. S. (2016). Social media English in Nigeria. Research Journal of English Language and Literature, 4(2), 775-784.

Ugwuanyi, C. C., Uduji, J. I., & Oraedu, C. (2020). Customer experience with self-service technologies in the banking sector: evidence from Nigeria. International journal of business and systems research.

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy, 21(6), 589.

Van Rossum, G. (2007). Python Programming Language. In USENIX annual technical conference (Vol. 41, p. 36).

Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.

Vo, A. D., Nguyen, Q. P., & Ock, C. Y. (2018). Automatic knowledge extraction for aspect-based sentiment analysis of customer reviews. In Proceedings of the 10th International Conference on Computer Modeling and Simulation (pp. 110-113).

Vu, N. T., Adel, H., Gupta, P., & Schütze, H. (2016). Combining recurrent and convolutional neural networks for relation classification. arXiv preprint arXiv:1605.07333.

Wahid, J. A., Shi, L., Gao, Y., Yang, B., Wei, L., Tao, Y., ... & Yagoub, I. (2022). Topic2Labels: A framework to annotate and classify the social media data through LDA topics and deep learning models for crisis response. Expert Systems with Applications, 116562.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv preprint arXiv:1804.07461

Wang, C. C., Day, M. Y., Chen, C. C., & Liou, J. W. (2018). Detecting spamming reviews using long short-term memory recurrent neural network framework. In Proceedings of the 2nd International Conference on E-commerce, E-Business and E-Government (pp. 16-20).

Wang, C., Paisley, J., & Blei, D. (2011). Online variational inference for the hierarchical Dirichlet process. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (pp. 752-760). JMLR Workshop and Conference Proceedings.

Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: a rating regression approach. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 783-792).

Wang, J., & Li, C. (2020). Research on User Satisfaction of Video Education Application Based on Reviews. In 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC) (pp. 340-343). IEEE.

Wang, R., Li, Z., Cao, J., & Chen, T. (2019). Chinese Text Feature Extraction and Classification Based on Deep Learning. In Proceedings of the 3rd Inter

Wang, S., Chen, Z., & Liu, B. (2016). Mining aspect-specific opinion using a holistic lifelong topic model. In Proceedings of the 25th international conference on world wide web (pp. 167-176).

Wang, S., Lv, G., Mazumder, S., & Liu, B. (2020). Detecting Domain Polarity-Changes of Words in a Sentiment Lexicon. arXiv preprint arXiv:2004.14357.

Wang, T., Yuan, C., & Wang, C. (2020). Does Applying Deep Learning in Financial Sentiment Analysis Lead to Better Classification Performance?

Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 606-615).

Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. International Journal of Science and Research (IJSR), 5(4), 819-821.

Wen, Y., Xu, A., Liu, W., & Chen, L. (2018). A wide residual network for sentiment classification. In Proceedings of the 2018 2Nd International Conference on Deep Learning Technologies (pp. 7-11).

Whitty, M. T. (2020). Is there a scam for everyone? Psychologically profiling cyberscam victims. European Journal on Criminal Policy and Research, 1-11.

Widiastuti, N. I. (2019). Convolution Neural Network for Text Mining and Natural Language Processing. In IOP Conference Series: Materials Science and Engineering (Vol. 662, No. 5, p. 052010). IOP Publishing.

Williams, M. J. (2020). External validity and policy adaptation: From impact evaluation to policy design. The World Bank Research Observer, 35(2), 158-191.

Winatmoko, Y. A., Septiandri, A. A., & Sutiono, A. P. (2019). Aspect and opinion term extraction for hotel reviews using transfer learning and auxiliary labels. arXiv preprint arXiv:1909.11879.

Woo, S. E., O'Boyle, E. H., & Spector, P. E. (2017). Best practices in developing, conducting, and evaluating inductive research.

Wu, C., Wu, F., Wu, S., Yuan, Z., Liu, J., & Huang, Y. (2019). Semi-supervised dimensional sentiment analysis with variational autoencoder. Knowledge-Based Systems, 165, 30-39.

Wunderlich, F., & Memmert, D. (2020). Innovative Approaches in Sports Science—Lexicon-Based Sentiment Analysis as a Tool to Analyze Sports-Related Twitter Communication. Applied Sciences, 10(2), 431.

Xia, Y., Tang, N., Hussain, A., & Cambria, E. (2015, April). Discriminative bi-term topic model for headline-based social news clustering. In The twenty-eighth international flairs conference.

Xiang, B., & Zhou, L. (2014). Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 434-439).

Xie, Q., Dai, Z., Hovy, E., Luong, M. T., & Le, Q. V. (2019). Unsupervised data augmentation for consistency training. arXiv preprint arXiv:1904.12848.

Xie, X., Ge, S., Hu, F., Xie, M., & Jiang, N. (2019). An improved algorithm for sentiment analysis based on maximum entropy. Soft Computing, 23(2), 599-611.

Xing, F., Malandri, L., Zhang, Y., & Cambria, E. (2020). Financial Sentiment Analysis: An Investigation into Common Mistakes and Silver Bullets. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 978-987).

Xu, G., Meng, X., & Wang, H. (2010). Build Chinese emotion lexicons using a graph-based algorithm and multiple resources. In Proceedings of the 23rd international conference on computational linguistics (pp. 1209-1217). Association for Computational Linguistics.

Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232.

Xu, L., Liu, J., Wang, L., & Yin, C. (2017). Aspect based sentiment analysis for online reviews. In Advances in Computer Science and Ubiquitous Computing (pp. 475-480). Springer, Singapore.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013, May). A biterm topic model for short texts. In Proceedings of the 22nd international conference on World Wide Web (pp. 1445-1456).

Yang, J., Yang, R., Lu, H., Wang, C., & Xie, J. (2019). Multi-Entity Aspect-Based Sentiment Analysis with Context, Entity, Aspect Memory and Dependency Information. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 18(4), 1-22.

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE Access, 8, 23522-23530.

Yang, Z., Kotov, A., Mohan, A., & Lu, S. (2015). Parametric and non-parametric user-aware sentiment topic models. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 413-422).

Yanuar, M. R., & Shiramatsu, S. (2020, February). Aspect extraction for tourist spot review in indonesian language using bert. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 298-302). IEEE.

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. European journal of education, 48(2), 311-325.

Yiran, Y., & Srivastava, S. (2019). Aspect-based Sentiment Analysis on mobile phone reviews with LDA. In Proceedings of the 2019 4th International Conference on Machine Learning Technologies (pp. 101-105).

Yiran, Y., & Srivastava, S. (2019, June). Aspect-based Sentiment Analysis on mobile phone reviews with LDA. In Proceedings of the 2019 4th International Conference on Machine Learning Technologies (pp. 101-105).

Yu, T., & Nwet, K. T. (2020). Comparing SVM and KNN Algorithms for Myanmar News Sentiment Analysis System. In Proceedings of 2020 the 6th International Conference on Computing and Data Engineering (pp. 65-69).

Yuan, L., Wang, J., Yu, L. C., & Zhang, X. (2020). Graph Attention Network with Memory Fusion for Aspect-level Sentiment Analysis. In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (pp. 27-36).

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2018). A survey of sentiment analysis in social media. Knowledge and Information Systems, 1-47.

Zhang, D., & Wang, D. (2015). Relation classification via recurrent neural network. arXiv preprint arXiv:1508.01006.

Zhang, H., Huating, S., & Wu, X. (2020). Topic model for graph mining based on hierarchical Dirichlet process. Statistical Theory and Related Fields, 4(1), 66-77.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

Zhang, Y., Zhang, Z., Miao, D., & Wang, J. (2019). Three-way enhanced convolutional neural networks for sentence-level sentiment classification. Information Sciences, 477, 55-64.

Zhao, J., Feng, Q., Wu, P., Warner, J. L., Denny, J. C., & Wei, W. Q. (2019). Using topic modelling via non-negative matrix factorization to identify relationships between genetic variants and disease phenotypes: A case study of Lipoprotein (a)(LPA). PloS one, 14(2), e0212112.

Zhao, P., Hou, L., & Wu, O. (2020). Modelling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. Knowledge-Based Systems, 193, 105443.

Zhou, C., Sun, C., Liu, Z., & Lau, F. (2015). A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630.

Zhou, J., Huang, J. X., Hu, Q. V., & He, L. (2020). Is position important? deep multi-task learning for aspect-based sentiment analysis.

Zhou, S., Chen, Q., & Wang, X. (2014). Fuzzy deep belief networks for semi-supervised sentiment classification. Neurocomputing, 131, 312-322.

Zhu, B., Zheng, X., Liu, H., Li, J., & Wang, P. (2020). Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. Chaos, Solitons & Fractals, 140, 110123.

Zhu, L., Pergola, G., Gui, L., Zhou, D., & He, Y. (2021). Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection. arXiv preprint arXiv:2106.01071.

Zhuang, H., Guo, F., Zhang, C., Liu, L., & Han, J. (2020). Joint Aspect-Sentiment Analysis with Minimal User Guidance. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1241-1250).

Appendix A: Research Ethics

A- 1: Twitter Data (SHU Approval)

Sheffield Hallam University

BAYODE OGUNLEYE <b4010127@my.shu.ac.uk>

Converis - Ethics Review - Approval

converis@shu.ac.uk <converis@shu.ac.uk> 3 August 2018 at 15:33
To: "Ogunleye, Bayode" <Bayode.O.Ogunleye@student.shu.ac.uk>

Dear Bayode

Title of Ethics Review: STATISTICAL ANALYSIS OF CUSTOMER CHURN BEHAVIOUR IN NIGERIA BANKING SECTOR

Ethic Review ID: ER7216241

The University has reviewed your ethics application named above and can confirm that the project has been approved.

You are expected to deliver the project in accordance with the University's research ethics and integrity policies and procedures: https://www.shu.ac.uk/research/ethics-integrity-and-practice.

As the Principal Investigator you are responsible for monitoring the project on an ongoing basis and ensuring that the approved documentation is used. The project may be audited by the University during or after its lifetime.

Should any changes to the delivery of the project be required, you are required to submit an amendment for review.

Wishing you success you with your study

Ethics Approval for Twitter.pdf

A- 2: Twitter Data (Twitter Approval)

Account Application Approved

From: Twitter Developer Accounts (developer-accounts@twitter.com)
To: hirisecontract@yahoo.com
Date: Monday, May 20, 2019, 08:23 PM GMT+1

Developer

Account Application Approved

Your Twitter developer account application has been approved!

Thanks for applying for access. We've completed our review of your application, and are excited to share that your request has been approved.

Sign in to your developer account to get started.

Thanks for building on Twitter!

developer.twitter.com | @twitterdev

Twitter, Inc. 1355 Market Street, San Francisco, CA 94103

PDF
Adobe Acrobat Document

PDF
Adobe Acrobat Document

A- 3: Survey Data for Validation

Sheffield Hallam University

BAYODE OGUNLEYE <b4010127@my.shu.ac.uk>

Converis - Ethics Review - Approval with Advisory Amendments

converis@shu.ac.uk <converis@shu.ac.uk> 19 August 2021 at 21:22
To: "Ogunleye, Bayode" <Bayode.O.Ogunleye@student.shu.ac.uk>

Status change comment
DO NOT WRITE ANYTHING IN THIS NOTES BOX AS IT CAN BE SEEN BY ALL OTHER USERS. Proceed to select the workflow status and click Done.

- Dear Bayode

Title of Ethics Review: Statistical learning approaches to sentiment analysis in Nigeria banking context
Ethic Review ID: ER35822832

The University has reviewed your ethics application named above and can confirm that the project has been approved.

The following advisory amendments were suggested, which you may wish to address:

The application needs some clarification, please see below.

You stated in P3 that "The plan is to store the data in an external hard drive which can be easily destroyed after use." and in the DMP you stated that the data are to be stored in the institutional repository.

Please amend your application to the recommended SHU Q drive to store research data.

Sentiment analysis will use customer voice, presumably from social media? Please clearly state how you plan to acquire such data, or how you plan to get access to such data.

For validation purposes, a survey will be conducted with bank staff, so you need to be clear about the criteria for inclusion/exclusion, how many people you are targeting, what would be an acceptable level of replies, and you also need to include the information for participants document with the legal statement on data collection and anonymity (mind you that if you target employees of a particular bank, their identity could be correlated to how they answer the questions, etc). Therefore, please clarify how you see this survey be conducted and you may also wish to submit a draft version of sample survey questions if relevant.

Alternatively to the validation information above, you could sate that you will acquire customer voice data now and state that the bank staff survey will be designed after you had the opportunity to analyse customer data and the survey will be subject of a further ethics application at a later stage of research. Therefore, it is imperative that you are clear about acquisition, access and processing of customer data.

If this is a second resubmission, the Lead reviewers comments will appear below:

Thanks for the clarification on data collection and analysis.

We recommend approval of the edited application subject to the following conditions:

1) On the survey, the address provided as (https://www.shu.ac.uk/

research/quality/ethics-and-integrity) is invalid. Please provide the correct link.

2) Also on the survey, please include the legal statement below:
"The University undertakes research as part of its function for the community under its legal status. Data protection allows us to use personal data for research with appropriate safeguards in place under the legal basis of public tasks that are in the public interest. A full statement of your rights can be found at:https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research
All University research is reviewed to ensure that participants are treated appropriately and their rights respected. This study was approved by STA-FREC Faculty Research Ethics Committee."

You are expected to deliver the project in accordance with the University's research ethics and integrity policies and procedureshttps://www.shu.ac.uk/research/ethics-integrity-and-practice.

As the Principal Investigator you are responsible for monitoring the project on an ongoing basis and ensuring that the approve documentation is used. The project may be audited by the University during or after its lifetime.

Should any changes to the delivery of the project be required, you are required to submit an amendment for review.

If you have a query regarding your application, please contact your Faculty Ethics Administrator in the first instance.

HWB - hwbethics@shu.ac.uk
STA - STAfrec@shu.ac.uk
SBS - sbsethics@shu.ac.uk
SSH - SSH-ResearchEthics@shu.ac.uk

Wishing you success with your study

Kind regards,
Ethics Research Support

*** This is an automatically generated email, please do not reply ***

Ethics Approval for Questionnaire.pdf

Appendix B: Questionnaire

Introduction

Dear Sir/Madam,

VALIDATION OF TOPIC SENTIMENT BANKING SYSTEM (TSBS)

I am a PhD student at Sheffield Hallam University (SHU), United Kingdom and I am currently conducting research on statistical learning approaches to sentiment analysis in Nigeria banking context. The aim is to understand customers' attitude towards banking using customers' social voice. To achieve this, sentiment analysis and topic modelling techniques were employed.

"Sentiment analysis (SA) is a tool that analyses peoples' opinion, evaluation, attitude and emotions towards tangible or intangible issues like product, services or topics. SA is useful to identify critical issues in real-time, for example, an angry customer about to churn (switch service provider). While Topic modelling also known as aspect extraction involves the process of extracting topics in the context of bank product and services."

The study proposed Topic Sentiment Banking System (TSBS) to illustrate the development processes which the banks can automate. The TSBS is a classification system that classifies customers' message/queries into topics (for example, ATM) and polarity (positive or negative). It is important to know your views regarding the TSBS research outcome in terms of relevance and usefulness. This will help establish the relevance of the research findings and recommendations.

I would like to thank you in advance for your valued and kind consideration. At SHU, confidentiality and anonymity are guaranteed as all the information gathered will conform to the University's Ethical procedure (https://www.shu.ac.uk/research/excellence/ethics-and-integrity/policies). The University undertakes research as part of its function for the community under its legal status. Data protection allows us to use personal data for research with appropriate safeguards in place under the legal basis of public tasks that are in the public interest. A full statement of your rights can be found at https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research. All University research is reviewed to ensure that participants are treated appropriately and their rights respected. This study was approved by STA-FREC Faculty Research Ethics Committee. If you would like to receive further information about the research, please feel free to contact me (b4010127@my.shu.ac.uk).

Bayode Ogunleye
Industry and Innovation Research Institute
Harmer 2418
Sheffield Hallam University
Howard Street | Sheffield | S1 1WB
United Kingdom
Email: b4010127@my.shu.ac.uk

*Required

1. I have read the Introduction for this study and have had details of the study explained to me. I understand that I am free to withdraw from the study without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher. I consent to the information collected for the purposes of this research study, once anonymised (so that I cannot be identified), to be used for any other research purposes. *

   Mark only one oval.

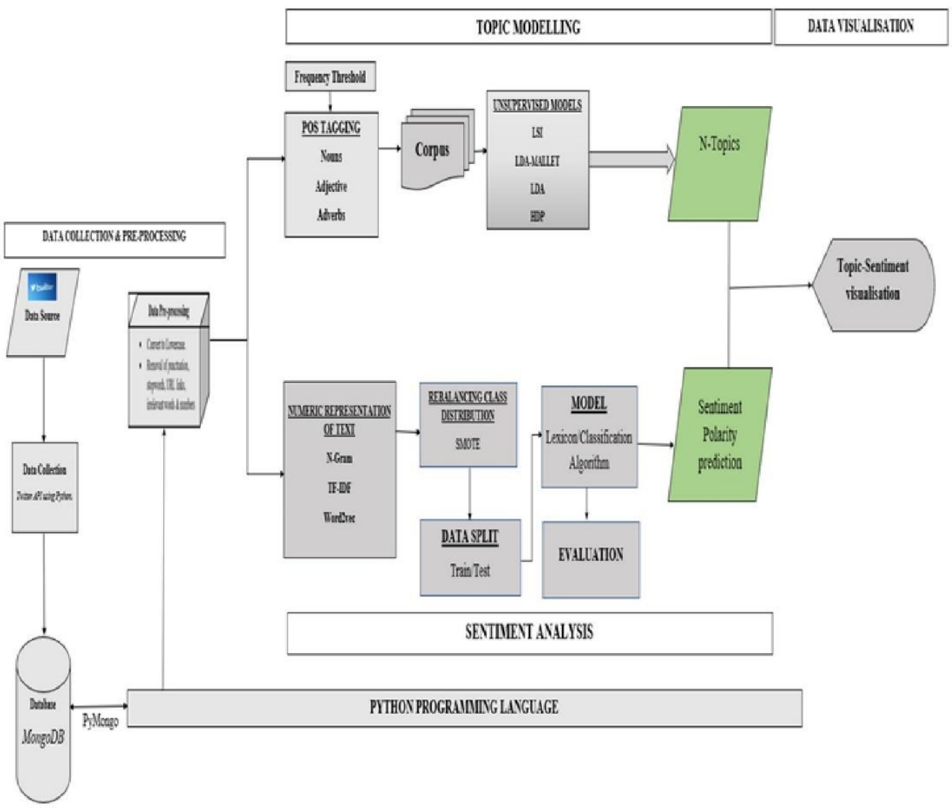
   - No
   - Yes

2. Please indicate your willingness to participate in this exercise by clicking on your preferred option below. I assure you that the data and information provided will remain strictly confidential and will be used only for the purpose of this research. *

   Mark only one oval.

   - No, I am not willing to participate
   - Yes, I am willing to participate

PROPOSED TSBS FRAMEWORK



RESEARCH FEEDBACK FORM: Part I

Please provide response to the questions.

3. 1). Do you use social media ?

Mark only one oval.

- ◯ No
- ◯ Yes

4. 2). How often do you use social media?

Mark only one oval.

- More than twice a day
- Once daily
- More than twice weekly
- Once weekly
- More than twice in a Month
- Once monthly
- Other: ______

5. 3). Which of the social media do you use?

Tick all that apply.

- [ ] Facebook
- [ ] Twitter
- [ ] Linkedln
- [ ] Instagram

Other: [ ] ______

6. 4). For what purpose do you use social media?

Mark only one oval.

- Work
- Connect with family and friends
- News
- Access information
- Other: ______

7. 5). What is your gender?

Mark only one oval.

- Male
- Female
- Prefer not to say

RESEARCH FEEDBACK FORM: Part II

Please provide responses on how valid the research claims and findings are with regards to your banking experience.

8. 6). Listening to customers' social voice is RELEVANT to understand their attitude and expectation. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

9. 7). Listening to customers' social voice is USEFUL to understand their attitude. This can help in making business decisions. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

10. 8). Monitoring customers' attitude towards bank product and service is beneficial. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

11. 9). Sentiment analysis helps classify happy and unhappy customer. Thus, marketing campaigns can be utilised to offer personalised product/service discounts to the customers. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

12. 10). What people think about your competitors can help you re-evaluate your priorities. Therefore, an automated system that provides valuable insight into your competitors analysis is relevant to your need. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

13. 11). Insights from sentiment analysis system can uncover areas that need improvement. Thus, such system is beneficial to the banks. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

14. 12). The usage of social network is growing. An automated system that organises and classifies your customers' message/queries into topics and sentiment positivity/negativity is considered beneficial. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

15. 13). Which one of these topics do you consider the MOST talked about theme (by customers) from your experience?

Mark only one oval.

- Bank events
- Contact/complaint form
- Mobile banking
- Bank charges
- Transaction problem
- Credit card payment
- Customer service experience
- Internet service performance
- Scam alert/advice
- Investment Bank

16. 14). Transaction problems are a major challenge to bank customers. To what extent do you agree?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

17. 15). Some customers are unhappy with BANK CHARGES (like interbank transfer, card maintenance & SMS debit alert fees). To what extent do you agree with this?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

18. 16). A significant proportion of customers are unhappy with customer service experience. To what extent do you agree with this?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

19. 17). Our research suggest the use of mobile banking, USSD and credit card is increasing. Knowing this can give a bank a competitive advantage. To what extent do you agree with this?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

20. 18). A topic sentiment classification system is URGENTLY needed in banks to help in providing customer centric bank services. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

21. 19). A topic sentiment classification system should be a COMPULSORY system for the banks to help in providing customer centric banking service. To what extent do you agree/disagree with this?

Mark only one oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

22. Please provide any additional note.

Appendix C: Interview Questions

- How would you describe your experience with the TSBS whilst demonstrating?
- To what extent is this system beneficial to the banks?
- What areas do you think the TSBS would be useful?
- Do you consider the framework relevant to what is needed in your bank?
- In your own words, can you describe the framework in terms of adoption, development, and deployment.

Appendix D - 1: Cover letter for Interview (TSBS framework Validation)

Sheffield Hallam University

Industry & Innovation Research Institute,
Sheffield Hallam University,
Howard Street,
Sheffield.
S1 1WB
12th May 2022.

Dear Sir/Madam,

VALIDATION OF TSBS FRAMEWORK

I am a PhD student at Sheffield Hallam University, United Kingdom and I undertook research that investigated customers' attitude towards banking using customers' social voice. I used sentiment classification and topic modelling techniques. Thus, proposed Topic Sentiment Banking System (TSBS) to illustrate the start-end processes involved in the development which can be automated. The TSBS is a classification system that classifies customers' message/queries into topics (for example, ATM) and polarity (positive or negative).

I am writing to request for your co-operation as I would be grateful if you could participate in an interview to seek your opinion about the TSBS framework. The purpose of the interview is to establish the usefulness and relevance of the system. At SHU, confidentiality and anonymity are guaranteed as all the information gathered will conform to the University's Ethical procedure (https://www.shu.ac.uk/research/excellence/ethics-and-integrity/policies). The University undertakes research as part of its function for the community under its legal status. Data protection allows us to use personal data for research with appropriate safeguards in place under the legal basis of public tasks that are in the public interest. A full statement of your rights can be found at https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research. All University research is reviewed to ensure that participants are treated appropriately, and their rights respected. This study was approved by STA-FREC Faculty Research Ethics Committee.

If you would like to receive further information about the research, please feel free to contact me (b4010127@my.shu.ac.uk). Kindly indicate your willingness to participate in this exercise by signing and returning the declaration in the next page.

Yours Faithfully,

Bayode Ogunleye
(PhD Student)

Please return form by email to b4010127@my.shu.ac.uk

Sheffield Hallam University

Declaration:

I wish to be interviewed and I understand that any information I provide will remain strictly confidential and only for the purpose of this research. In addition, I understand that the interview will be conducted via Zoom.

Signature ______________________________

Preferred date/time of interview ______________________________

Please return form by email to b4010127@my.shu.ac.uk

Appendix D - 2: Summary of Feedback (TSBS framework Validation)

INTERVIEW FEEDBACK External Inbox x

[redacted]

to B.Ogunleye@student.shu.ac.uk

10:52 (1 hour ago)

Dear @ B.Ogunleye@student.shu.ac.uk,

Below is my interview feedback:

In general, the system demonstrates the application of human intelligence with the framework is transparent and not complex.
The framework demonstrates thorough processes and there was comparison of methods for each component before proposing the method with optimal performance.

The system is beneficial because it helps to understand how customers use our services, most importantly, the active aspects of our services. Rather than just understanding a customer is unhappy or happy, it is quite useful that the system can help identify the customer issues in terms of aspect. The framework is relevant and like the current customer experience and satisfaction survey used by the customer experience and value management (CEVM) team.

However, the framework demonstrates new approach to understand customers. Interestingly, the system can accommodate the existing survey approach. The framework consists of modules in Python language.

This can conveniently communicate and access our current code which makes the development and deployment easier. Generally, a slight difficulty is that the system requires a lot of human coding. For example, human labelling of tweets. This is labour intensive.

Thank you.

Figure: 10.1: Participant 1 interview feedback

INTERVIEW FEEDBACK External Inbox x

[redacted]

to B.Ogunleye@student.shu.ac.uk

12:54 (8 hours ago)

The framework is suitable for the bank to understand their customers. The framework is adaptable which is the most positive thing. Another important thing is it is easy to integrate with existing measures of customer experience survey. The scope of the input is not limited to tweet but can work well with our current approaches of understanding customer perception. For example, reviews, satisfaction rating questionnaire and customer service data. This increases the chances of the system been adopted. An interesting aspect is the flexibility to be able to choose number of topics as desired. The system components are easy to manage because they contain modules and resources that are flexible to modify and query. The framework provides a flexible architecture to accommodate changes in frequency threshold, number of terms and topics. Whilst demonstrating the system, insights were coming up within on the benefits of the systems as it could revolutionise and complement strategies and decision-making processes in terms of product development and customer experience. Another attractive thing is the lexical resources already created for bank use. For example, at the pre-processing phase, there are list that just needs additional words to be added for pruning of unwanted words. All phases of the analytics are clear, thorough, and transparent. The framework adopted a popular language thus the framework appears cheap to adopt.

Thank you.

Figure: 10.2: Participant 2 interview feedback

INTERVIEW FEEDBACK External Inbox x

[redacted] 12:01 (9 hours ago)

to B.OGUNLEYE@STUDENT.SHU.AC.UK

The framework provided clarity on the requirements of the system. The system is easy to understand and navigating around the modules including the resources appears easy to deploy. The system demonstrates thorough processes. The performance achieved is good and the appropriateness of techniques adopted by the system were well discussed. The framework can be delivered in any popular programming language that can be integrated with databases that supports unstructured data. Thus, reduces development cost and time. Clarity on top level functionalities were provided in the framework. A common challenge with framework is how they struggle to demonstrate connection and communication. However, the TSBS was able to demonstrate the connection at all stages from data source to database to data modelling and prediction stage. The banks can benefit from the system components because the framework can be applied to analyse customer emails, feedback, reviews, surveys, chat data, and queries. For example, a case of identifying the most significant topic in chat box data. Another useful part of the system is that it highlights the significance of the topics. The highlighted significant topics aids customer behavior understanding which is relevant to bank need. In my opinion, the framework is acceptable.

Thank you.

Figure: 10.3: Participant 3 interview feedback

Appendix E: Links to Dataset

DATASET Links:

https://www.kaggle.com/batoog/labelled-pidgin-english-bank-tweets-1000

https://www.kaggle.com/batoog/labelled-pidgin-english-wordlist-for-research

https://www.kaggle.com/batoog/labelledpidginenglish7037tweets-for-research

https://www.kaggle.com/batoog/dirtybanktweets-for-research

Appendix F: Python Code

PDF
SentiLeye algorithm code.pdf

PDF
Machine learning code.pdf

Deep learning codes.pdf

TM codes.pdf

```
# Dictionary of English Contractions
contractions_dict = { "ain't": "are not","'s":" is","aren't": "are not",
                     "can't": "cannot","can't've": "cannot have",
                     "'cause": "because","could've": "could have","couldn't": "could not",
                     "couldn't've": "could not have", "didn't": "did not","doesn't": "does not",
                     "don't": "do not","hadn't": "had not","hadn't've": "had not have",
                     "hasn't": "has not","haven't": "have not","he'd": "he would",
                     "he'd've": "he would have","he'll": "he will", "he'll've": "he will have",
                     "how'd": "how did","how'd'y": "how do you","how'll": "how will",
                     "I'd": "I would", "I'd've": "I would have","I'll": "I will",
                     "I'll've": "I will have","I'm": "I am","I've": "I have", "isn't": "is not",
                     "it'd": "it would","it'd've": "it would have","it'll": "it will",
                     "it'll've": "it will have", "let's": "let us","ma'am": "madam",
                     "mayn't": "may not","might've": "might have","mightn't": "might not",
                     "mightn't've": "might not have","must've": "must have","mustn't": "must not",
                     "mustn't've": "must not have", "needn't": "need not",
                     "needn't've": "need not have","o'clock": "of the clock","oughtn't": "ought not",
                     "oughtn't've": "ought not have","shan't": "shall not","sha'n't": "shall not",
                     "shan't've": "shall not have","she'd": "she would","she'd've": "she would have",
                     "she'll": "she will", "she'll've": "she will have","should've": "should have",
                     "shouldn't": "should not", "shouldn't've": "should not have","so've": "so have",
                     "that'd": "that would","that'd've": "that would have", "there'd": "there would",
                     "there'd've": "there would have", "they'd": "they would",
                     "they'd've": "they would have","they'll": "they will",
                     "they'll've": "they will have", "they're": "they are","they've": "they have",
                     "to've": "to have","wasn't": "was not","we'd": "we would",
                     "we'd've": "we would have","we'll": "we will","we'll've": "we will have",
                     "we're": "we are","we've": "we have", "weren't": "were not","what'll": "what will",
                     "what'll've": "what will have","what're": "what are", "what've": "what have",
                     "when've": "when have","where'd": "where did", "where've": "where have",
                     "who'll": "who will","who'll've": "who will have","who've": "who have",
                     "why've": "why have","will've": "will have","won't": "will not",
```

Figure: 10.4: Screenshot of contraction list

Appendix G: Screenshots of Classification Model Performance

PHD ML result.pdf

Appendix H: Screenshot of Nigeria Commercial Banks' Handle

@accessbank_help	@myaccessbank	@Sterling_Bankng	@UNIONBANK_NG
@AskCiti	@diamondbankhelp	@UBACares	@UnityBankPlc
@ecobank_nigeria	@fidelitybankplc	@wemabank	@ZenithBank
@FirstBankngr	@keystonebankng	@PolarisBankLtd	@StanbicIBTC
@MyFCMB	@gtbank_help		

Table 10.1: Screenshot of Nigeria commercial banks' handle