Using Opinionated-Objective Terms to Improve Lexicon-Based Sentiment Analysis

OGUNLEYE, Bayode, BRUNSDON, Teresa, MASWERA, Tonderai, HIRSCH, Laurence <http://orcid.org/0000-0002-3589-9816> and GAUDOIN, Jotham

Using Opinionated-Objective Terms to Improve Lexicon Based Sentiment Analysis

Bayode Ogunleye [1], Teresa Brunsdon[2], Tonderai Maswera[3], Laurence Hirsch[3], Jotham Gaudoin[3]

[1] Department of Computing & Mathematics, University of Brighton, Brighton BN2 4GJ, United Kingdom
[2] Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom
[3] Department of Computing, Sheffield Hallam University, Sheffield S1 2NU, United Kingdom
* Corresponding Author Email: B.Ogunleye@brighton.ac.uk

Abstract. Sentiment analysis (SA) has received huge attention to understand customer perception, especially in the movie review (IMDB) domain. This is due to the availability of large, labelled dataset. This has enhanced the use and development of machine learning (ML) algorithms ranging from the traditional machine learning algorithms, deep learning algorithms to large language models. The ML models have shown great performances. However, the application of ML methods for SA is limited in service industry like banking, due to the unavailability of large training dataset. Thus, we consider the use of lexicon-based sentiment analysis appropriate. We employ 346,000 Nigeria bank customers' tweets to develop our corpus and thus, propose SentiLeye, a novel lexicon-based algorithm for sentiment analysis. Our algorithm incorporates corpus-based approach and external lexical resources for sentiment lexicon generation of Pidgin English language terms (a non-English under resourced language). Moreover, we demonstrate the use of verbs and adverbs that express opinion on service experience to improve the performance of lexicon-based sentiment analysis. Results show that SentiLeye outperforms popular off-the-shelf sentiment lexicons with macro F1 score of 76%. We conclude that results from domain specific algorithms such as SentiLeye evidence that general purpose lexicons cannot replace them.

Keywords: Sentiment Analysis, Sentiment Classification, Lexicon, Banking Industry, Pidgin English.

1 Introduction

The banking industry play significant role in every nation's economy (Anastasiou & Katsafados, 2023; Alsmadi et al. 2019; De Jesus & Da Nóbrega Besarria, 2023). Similarly, customers are important to the banks for profitability and stability. Thus, understanding customers' perspectives is vital. To this end, surveys, interviews, and polls can be conducted to understand customers' perceptions. However, surveys are labour intensive, expensive, and limited to pre-defined variables. The social web has become an alternative data source for academia and industry since 2004 (Martin-Domingo et al. 2019). According to Agnihotri et al. (2021), the social web serves as a communication platform for both banks and their customers. This was more obvious during the Covid-19 pandemic. The banks experienced a significant increase in digital transactions since the Covid-19 national lockdown (Jha & Shah, 2021; Ogunleye, 2021). It also led to increased usage of social media, by customers to express their feelings and experience. This has increased the need for banks to mine their customers perceptions using social media data. This can be helped using sentiment analysis methods.

Sentiment analysis (SA) classifies words or phrases into sentiment categorises such as positive and negative. There are two common SA approaches: lexicon-based and machine learning (ML). The latter uses supervised learning algorithms such as Bi-directional Long-Short Term Memory (Alnawas, 2022), and Support Vector Machine (Ogunleye, 2021) to predict sentiment polarities. In contrast, the lexicon-based approach uses dictionaries to map words according to their semantic orientation into sentiment categories. A lexicon is a dictionary built for a domain of interest, such as sentiment analysis. The lexicon-based approach performs well across different domains (Huang et al. 2020; Turner et al. 2020). However, their performance varies across domains. In general, the lexicon-based gives easily interpreted results, but the supervised machine learning approach is more accurate. Recently, the use of large language models (LLMs) like BERT have helped achieve up to 93% accuracy (Sun et al. 2019). However, ML approaches tend to be black box models that are impossible for humans to interpret. They usually also depend on large, annotated datasets which are mostly available in English. Unfortunately, there are languages like Pidgin English that are low resourced. In addition, domains like the banking lack sufficient labelled data (Ogunleye et al. 2023; Martinis et al. 2022). The authors in Du et al. (2023) added that literature in financial sentiment analysis is limited due to a lack of high quality large financial datasets because the domain is highly professional. This hinders the use of state-of-the-art (SOTA) ML approaches in this context. Thus, validates the use of lexicon-based sentiment analysis approach in a low resourced context (Fehle et al. 2021).

Based on this background, we focus on the lexicon-based SA. Unfortunately, only few studies have applied lexicon-based sentiment analysis in the banking context. Studies like Wu et al. (2023), Li et al. (2021) and Bos & Frasincar (2022) stated that limited literature in financial sentiment analysis is due to the complexity and terminology of the domain, and this warrants a domain specific system. Thus, we propose a novel SentiLeye algorithm to suit this context. Our paper contributes to existing knowledge by comparing lexicon SA models to ascertain the best performing sentiment lexicon in the banking context. We demonstrate the use of opinionated-objective terms to improve sentiment lexicon performance. Most importantly, we prepared and made publicly available, benchmark lexical resources in Pidgin English and the banking context to improve and encourage research in this area. The remaining sections of the paper are organised as follows. Section 2 will provide a comprehensive review of the relevant literature, offering the necessary background knowledge for this study. Section 3 will discuss the methodologies employed. Section 4 will present the findings, along with a detailed discussion. Finally, Section 5 will present the conclusions drawn from the study and provide recommendations for further research.

2 Related Work

Sentiment analysis (SA) is a task that uses statistical learning, computational linguistic, natural language understanding and processing approaches for generating insight (Ogunleye, 2021). For example, Yousefinaghani et al. (2021) collected 4,552,652 tweets between 7th January 2020 and 3rd January 2021. They classified people's opinion on Covid 19 vaccination into positive, neutral, and negative and thus showed positive sentiment polarity was dominant towards the vaccination. Furthermore, SA has been applied to understand; patient review of healthcare service experience (Gui & He, 2021), corporate financial performance (Wang et al. 2023), public opinion and emotion for non-fungible tokens (Qian et al. 2022) and predict election outcomes (Amusa et al. 2016; O'Connor et al. 2010; Oyewola et al. 2023; Ringsquandl & Petkovic, 2013; Rizk et al. 2023). The concept of sentiment analysis has been extended to cyberbullying detection (Atoum, 2023), depression detection (Babu & Kanaga, 2022), fake news detection (Alonso et al. 2021; Iwendi et al. 2022), and recommender system (Bhattacharya et al. 2022; Choudhary et al. 2023). The lexicon-based sentiment analysis approaches have shown good performance across domains due to their general lexical knowledge (Huang et al. 2020). However, the traditional lexicons are not appropriate for social web text due to continuous use of informal words (Ogunleye, 2021). SentiStrength was developed to address this problem (Thelwall et al. 2012). Based on the consistent performance of SentiStrength, the lexicon has been extended to suit domains like software engineering (Islam & Zibran, 2018). Ahmad et al. (2017) showed SentiStrength outperformed other eleven lexicons using Twitter, BBC comments and DIGG comments datasets with performance accuracy of 92 - 95%. However, the problem of context independent in SentiStrength poses a challenge when dealing with words that have different meanings in different context (Islam & Zibran, 2018; Saif et al. 2016). Based on this, Saif et al. (2016) proposed SentiCircle. The lexicon was created to identify the contextual meaning of words, allowing the algorithm to adjust the sentiment strength and polarity of each word accordingly. Saif et al. (2016) conducted an evaluation of SentiCircle using datasets related to the Obama debate, health care reform, and STS-Gold Twitter. The study demonstrated that SentiCircle has the ability to recognize word patterns across different domains, resulting in higher accuracy compared to MPQA and SentiWordNet. However, SentiCircle was outperformed by SentiStrength in terms of F1-measure. According to Koto & Adriani (2015), Afinn is a reliable lexicon, the lexicon showed comparable or superior performance compared to other lexicons such as SentiWordNet, Opinion lexicon, and SentiStrength. Ribeiro et al. (2016) conducted a comparative study with 24 sentiment lexicons using 18 datasets from social networks, online reviews, and comments. Their findings indicated that SentiStrength and Afinn performed well, particularly with social network datasets. On the other hand, Zimbra et al. (2018) presented an argument in favour of NRC. They evaluated 28 sentiment analysis techniques, including NRC and SentiStrength, across 5 distinct datasets. The results of their experiment demonstrated that NRC outperformed SentiStrength in four domains in terms of accuracy. Qian et al. (2022) collected 290,282 English tweets of verified Twitter account holders of non-fungible token buyers and owners from 15th October 2021 to 20th June 2022. They used NRC lexicon to understand public opinion and emotion on non-fungible token (NFT).

Authors in Sohangir et al. (2018) applied VADER in financial social media dataset (StockTwits) and achieved performance with accuracy of 94% and F1 score of 79%. However, Todd et al. (2019) pointed out that previous studies have often employed general-purpose lexicons like Afinn and LIWC in domains where their validity has not been established, without considering the classification accuracy. They argued that it is necessary to re-evaluate words or adapt lexicons for specific domains to ensure optimal lexicon performance. Dąbrowski et al. (2023) added that there is need to consider algorithm efficiency before deployment. The use of general-purpose sentiment lexicon in the financial domain is more difficult and can be misleading. This is because terms like "*tax" "cost" and "liability"* are scored negative in general purpose lexicons like General Inquirer (GI). However, these terms are neutral in the financial context. Studies like Correa et al. (2021), Ogunleye (2021), Krishnamoorthy (2018), and Loughran & McDonald (2011) expressed their concerns over lexicons not validated in the financial context. They went further to demonstrate the mediocre performance of GI with financial text. Loughran & McDonald (2011) proposed LM lexicon validated in the financial context. Their lexicon comprises of 354 positive, 2355 negative, 297 uncertainty, 904 litigious, 19 strong modal, 27 weak modal and 184 constraining words. The LM lexicon has been successfully used in several studies (Das, 2014; Das et al. 2022; Gandhi et al. 2019; Kearney and Liu, 2014; Li et al. 2021; Ogunleye, 2021). Correa et al. (2021) developed a domain specific dictionary to classify financial stability documents. Their lexicon consists of 391words which was labelled manually by two human annotators into 96 positive and 295 negative words. Thus, performed SA using the lexicon on financial stability report from 2005 to 2017 of thirty countries. Li et al. (2021) developed CFDSL a novel Chinese financial domain sentiment lexicon for financial distress prediction. Man et al. (2019) conducted a survey of literature on financial sentiment analysis and found out there is no sentiment lexicon validated to suit financial social media text and thus concluded this problem is due to limited lexical resources in the domain. Consoli et al. (2022) developed SentiBigNomics lexicon to suit economic and financial context. They made use of Amazon AWS SageMaker service to recruit human annotators for labelling their lexicon wordlist and thus showed their approach outperformed baseline lexicons such as SentiWordNet.

In the banking domain, Mittal & Agrawal, (2022) performed sentiment analysis on 32,217 online reviews of Indian bank customers collected from bankbazaar.com using NRC lexicon. However, they did not present the performance evaluation result of NRC before using the result as a constraint in their regression model. Authors in Afolabi et al. (2019) employed SentiWordNet and domain ontology concepts to classify tweets from bank customers. However, it is important to note that their wordlist was limited in terms of coverage. Additionally, their approach was sub-optimal as they attempted to expand slangs and bank-related words without considering the contextual nuances. This led to potential errors, as they overlooked the specific meanings of certain words. For instance, the term *"atm"* was expanded to *"at-the-moment"* instead of recognizing it as *"automated teller machine"*. The review of existing literature indicates that sentiment analysis studies in the Nigerian banking context are limited due to the lack of resources in terms of bank domain and Pidgin English lexical resources. African countries, which account for 32% of the world's total living languages, face challenges in building lexicons in their native languages due to limited lexical resources available (Kaity & Balakrishnan, 2019). Catelli et al. (2022) added that there is need to develop non-English sentiment lexicons to suit the low-resourced language. Dashtipour et al. (2022) enhanced their lexicon PerSent using 1000 idiomatic expression to improve on classifying Persian language text. Schmidt et al (2021) developed SentText lexicon to classify German language text which was validated in digital humanity context. Fehle et al. (2021) evaluated German language sentiment lexicons and validated SentiWS and SentiMerge as the best performed lexicons with accuracy and f1-score around 67%. Catelli et al. (2023) applied SentIta, an Italian lexicon to 353,217 Italian tweets to detect people's opinion and attitude towards Covid-19 vaccine. De-Melo (2022) developed SentiLexBR in Portuguese language. Machova et al. (2020) developed sentiment lexicon in Slovak language. Unfortunately, there is no sentiment lexicon validated for Pidgin English language.

In service industries like the banking, there are terms which are used to demonstrate customer experience on a service. This term could be used to express positive or negative experience or operation of the banking services. It is common with the customers' tweet to use the opinionated-objective sentences to demonstrate their frustration as some customers want to be polite. Liu (2015) concurred that the issue of detecting factual words that mean sentiment in a specific context remains unresolved. For instance, *"dem don debit my account since last month oooo till now, no refund @EFCC @CBN"*. This tweet demonstrates a customer's tweet directed at the bank, expressing dissatisfaction with the poor service they encountered. In this instance, the customer tagged the Economic and Financial Crime Commission (EFCC) in Nigeria and the Central Bank of Nigeria (CBN) out of frustration, aiming to escalate the issue. It is worth noting that opinionated-objective sentences like this are common in customer discussion. Text data do not necessarily need to contain adjectives like good or bad before expressing an opinion or sentiment. Yadav et al. (2020) emphasized more on this and thus proposed noun-verb approach from the background that adjectives are not good sentiment indicators of financial data. Therefore, there is need to consider opinionated objective terms explicitly for lexicon development.

There are two common methods for creating sentiment lexicons, namely, manual, and automatic approaches. The manual approach involves gathering and labelling of sentiment terms manually. This approach has been used successfully by several studies. For example, SentiBigNomics (Consoli et al. 2022), MPQA (Wilson et al. 2005), and LIWC (Pennebaker et al. 2001). However, it is labour intensive and time-consuming process, and often result to limited coverage. On the other hand, the automatic approach involves using a set of seed words to generate synonyms and autonyms with the help of online dictionaries or corpus. This approach has been adopted by several studies. For example, De-Melo (2022) developed SentiLexBR using automatic approach. Machova et al. (2020) developed big and small size lexicons using automatic approach. They generated their wordlist from lexicons like SentiStrength, Hu and Liu Opinion lexicon, SenticNet, Sentiment140, and Afinn and thus used particle swarm optimisation to label the words in their lexicon. Their lexicon consists of 598 positive, 772 negative, 41 intensifiers and 19 negation words in the big lexicon and 85 positive, 15 negative words in their small lexicon. They showed their lexicon approach outperformed existing lexicon in 4720 reviews from the electronic, book, movie, and politics domain. Yadav et al. (2020) developed their lexicon using automatic approach. They generated seed words from corpus of text data collected for 5 months from online stock market (Reuters.com & moneycontrol.com). They labelled the words using PMI-IR (point wise mutual information – information retrieval) measure and their lexicon showed superior performance with 79% accuracy and f1-score 86%. Numerous studies (Choi & Cardie, 2009; Hatzivassiloglou & McKeown, 1997; Kanayama & Nasukawa, 2006) have utilized a corpus-based automatic approach to generate sentiment lexicons. This approach is advantageous as it facilitates the detection of syntactic relations between opinions and targets. However, it has certain limitations. Firstly, it often leads to the creation of domain-specific lexicons, which may not be applicable to other domains. Additionally, it lacks pre-processing tools that can adequately support languages other than the one used in the corpus. Al-Twairesh et al. (2016) used PMI to label words generated from corpus and thus proposed AraSenTi (131,342 Arabic words in AraSenTi-Trans and 93,961 Arabic words in AraSenTi-PMI), an Arabic lexicon for sentiment analysis. Farah & Kakisim (2023) used n-gram to enhance the performance of their lexicon by preserving the word order of text in the feature space. Thus, proposed their framework for lexicon enhancement.

In summary, our literature review findings suggest that to develop an efficient sentiment lexicon, it is crucial to identify terms specific to the banking domain, opinionated-objective terms, and Pidgin English terms in order to validate the sentiment lexicon in this particular context. This aligns with the findings of Farah & Kakisim (2023) and Zhou et al. (2014), which highlight that a high-quality lexicon leads to improved performance in sentiment classification. Consequently, this study emphasizes the importance of a high-quality lexicon that is tailored to the Nigerian banking context. Table 1 below presents a summary of off-the-shelf lexicons found in the existing literature. In the next session, this study details the method employed to develop SentiLeye, our novel lexicon to suit this context.

Table 1. Description of popular sentiment lexicons (Ogunleye, 2021)

Lexicon	Author	Total	No of Positive	No of Negative	Scale	Labelling method	General/Domain Purpose
Afinn	Nielsen (2011)	2477	878	1598	-5 to +5	Manual	General
LM	Loughran & McDonald (2011)	3917	354	2355	-1 to +1	Manual	Domain specific
Bing	Hu & Liu (2004)	6789	2006	4783	-1 to +1	Machine Learning	General
SentiWordNet	Baccianella et al. (2010)	20093	8898	11029	-1 to +1	Machine Learning	General
SO-CAL	Taboada et al. (2011)	5971	2438	3530	-1 to +1	Amazon Mechanical Turk	General
NRC	Mohammad & Turney (2013)	5555 (emotion exclusive)	2312	3324	0 or 1	Amazon Mechanical Turk	General
WKWSCI	Khoo & Johnkhan (2018)	10221	3121	7100	-3 to +3	Manual	General

VADER	Hutto & Gilbert (2014)	7500	____	_____	-4 to +4	Amazon Mechanical Turk	General
SentiStrength	Thelwall et al. (2012)	58119	____	_____	-5 to +5	Manual	General

3 Methodology

We propose the use of SentiLeye algorithm for lexicon-based sentiment analysis. We compare our approach to popular off-the-shelf lexicons such as Afinn, SentiStrength, NRC, Bing, SentiWordNet, VADER and TextBlob. This will help evaluate how the general-purpose lexicons perform in this particular context. We use SentiStrength, SentiWordNet and VADER as the baseline models. This is because these lexicons have achieved good performance and have been successfully adapted to several domains, especially, in domains with little or no lexical resources available (Consoli et al. 2022; Denecke & Reichenpfader, 2023; Ghosh et al. 2023; Gouthami, & Hegde, 2023; Long et al. 2023; Martinis et al. 2022; Ogunleye, 2021; Veltmeijer & Gerritsen, 2023). We use Python programming language in this study. This is because Python is popular in data science community and has sophisticated libraries for analytics.

3.1 Dataset

We utilised the data collected in the study of Ogunleye (2021) and is publicly accessible in the Kaggle repository (https://www.kaggle.com/batoog/datasets?scroll=true). Specifically, we focused on the 'text' column of the dataset for our research purposes. The dataset comprises tweets directed at the Twitter handles of 18 commercial banks in Nigeria. Ogunleye (2021) collected a total of 959,000 tweets from Nigerian bank customers over a period of nine months, spanning from May 12, 2019, to February 2020. For the development of our lexicon, we use a total of 346,000 Nigerian bank customers’ tweets collected from 12 May 2019 to 25 August 2019 (three months). For evaluation purpose, we randomly selected 1000 Nigerian bank customers’ tweets collected from 25 August 2019 to February 2020 in the same study. Unfortunately, we were unable to obtain another domain specific dataset to suit the context of this study. This could have extended the evaluation process of our algorithm. A common challenge of social media data is the text shortness (Ogunleye et al. 2023). However, it is more challenging in this case due to difference in phraseology as the tweets were in Pidgin English and English language.

3.2 Data Pre-processing

Data pre-processing techniques such as removal of numbers was performed. However, care was taken in this context. This is because in Nigeria language context some numbers have interpretations. For example, the number ‘419’ means ‘scam’. Thus, the text data was looped and changed to that effect before removing other numbers. Subsequently, we removed the URLs and stopwords. The natural language toolkit (NLTK) library was used for stopwords removal. The text data was converted to lowercase and stemming was applied. Table 2 below shows example of input raw text against the cleaned text.

Table 2. Examples of raw tweets and cleaned tweets.

Tweets	Cleaned Tweets
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 of your faulty machines. How can you be sending same message multiple times?	address team charge faulty machines send message multiple times

3.3 Lexicon Generation

Following the pre-processing stage, we applied the corpus based automatic generating technique to extract bank domain and Pidgin terms. Ogueji & Ahia (2019) crawled Pidgin English sentences and words from the BBC Pidgin English news website to develop their lexical resources. We used their lexical resource because the news website captured variations of Pidgin. Their corpus consists of 56695 sentences and 32925 words. Subjective Pidgin English words were manually retrieved and checked for errors. This was then added to the bank domain and Pidgin English terms. The natural language toolkit (NLTK) library was used for part of speech (POS) tagging. The POS tagging was applied to identify the adjectives (subjective), adverbs and verbs (opinionated objective) and thus created 1100 seed words. These words were manually labelled (at a range of strong positive as +2 to strong negative -2) by three native speakers of Pidgin English and English language. Finally, the seed words used were expanded through an automatic sentiment generation technique using WordNet, an English lexical resource developed by Miller (1995). This technique involved retrieving synonyms and antonyms for the seed words. The synonyms were labelled based on their corresponding root word in the corpus, while the antonyms were labelled as opposites. Additionally, the study benefited from incorporating SentiStrength emoticons and slang wordlist (Thelwall et al. 2012), due to the sentiment lexicon performance with social media data. A manual inspection was done to remove any errors. We employed manual annotation because it produces reliable result compared to other methods (Boukes, 2020). The lexicon algorithm can be summarised as follow.

- Check for word matches between the sentence/document and the SentiLeye package wordlist.
- Detect the presence of negating words. If found, reverse the result of the functioning words. In this study, the negating term will affect the next two words or prior words (using index) to the functioning word.
- Detect booster words that can either positively or negatively impact the value of the functioning words. Similarly, the next two words (using index) will be affected.
- Add up the values associated with the words to obtain a cumulative score.

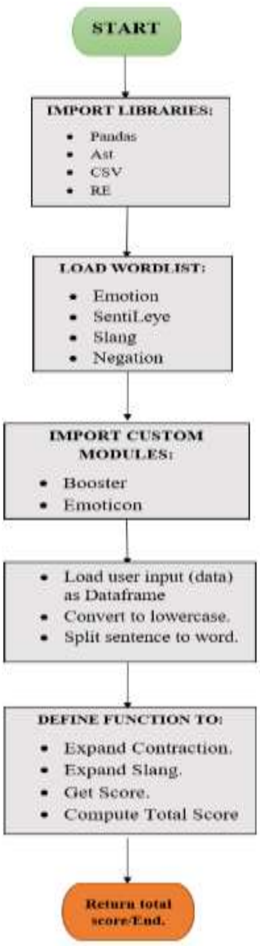

Fig 1. An illustration of SentiLeye algorithm

Figure 1 above presents the flowchart of our algorithm and Figure 2 below presents the Pseudocode of the algorithm.

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.

Fig 2. Pseudocode for SentiLeye Algorithm

3.4 Corpus Statistics

Table 4 below presents the lexicon statistics and Table 5 presents the distribution of the sentiment polarity labels. It is worth mentioning that the slang, negation, and booster words are not scored as they are list of words performing specific function when identified in a document.

Table 4. SentiLeye Statistics

Language	Count
Pidgin English	352
English	3734
Slang	51
Emotion	185
Emoticon	116
Booster	22

Table 5. Distribution of sentiment polarity labels

Sentiment Polarity	Label	Count
Strong Positive	2	1482
Positive	1	1649
Negative	-1	891
Strong Negative	-2	353

3.5 Evaluation Approach & Metrics

We randomly selected 1000 tweets and employed three human annotators for labelling of the tweets into positive, negative, and neutral sentiment polarities. The third annotator was employed in case of disagreement between the first two annotators. We evaluated the sentiment lexicons against the human annotated tweets and thus, the predictions of the lexicons were compared with the ground truth. Table 6 below presents the statistics of the annotated tweets. It is worth mentioning that the negative labelled tweets were high because the entire tweets collected contains high proportion of negative tweets.

Table 6. Human annotated tweet statistics

Sentiment Polarity	Count
Positive	42
Negative	708
Neutral	250

Figure 3 below presents the confusion matrix values true positive, true negative, false positive and false negative and this was used to calculate the accuracy, precision, recall and the F1-score.

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

Fig 3. Confusion matrix (Ogunleye, 2021)

Where:

- True positive (TP): correctly predicted as positive and is indeed positive.
- True negative (TN): correctly predicted as negative and is indeed negative.
- False positive (FP - type I error): incorrectly predicted as positive when it is actually negative.
- False negative (FN - type II error): incorrectly predicted as negative when it is actually positive.

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

$$\frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}} \tag{1}$$

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

$$\frac{\mathbf{TP}}{\mathbf{TP + FP}} \tag{2}$$

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

$$\frac{\mathbf{TP}}{\mathbf{TP + FN}} \tag{3}$$

F1 measure provides the balance between precision and recall:

$$\frac{\mathbf{2PR}}{\mathbf{P + R}} \tag{4}$$

4 Results

We present our comparative result of the sentiment lexicons in Table 4 below and thus critically discuss our findings during the experimentation.

Table 7. Evaluation result of sentiment lexicons.

Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SentiLeye	**77**	85	78	**76**
Afinn	38	71	37	45
TextBlob	41	71	40	44
VADER	35	74	34	43
SentiStrength	61	69	64	64
NRC	32	69	31	39
Bing	50	71	51	55
SentiWordNet	39	66	37	46

From Table 7 above, we showed that our sentiment lexicon algorithm achieved an accuracy of 77% and weighted F1 score of 76% which is superior in terms of performance to other lexicons. This performance is attributed to the ability of SentiLeye being able to detect domain-dependant terms, opinionated-objective terms, Pidgin English words, and negation. SentiStrength seconded our approach in terms of performance, the lexicon achieved an accuracy of 61% and weighted F1 score of 64%. This makes sense as SentiStrength has been shown to be a top performing lexicon especially with social media text (Thelwall et al. 2012; Ahmad et al. 2017). SentiStrength captured more neutral tweets than SentiLeye because SentiStrength is a general-purpose sentiment lexicon with more terms and coverage. The Bing lexicon also showed satisfactory performance which is due to the good coverage of social media terms, malformed words, and opinionated-objective words like failed in their lexicon. However, Bing struggled with domain-dependant words like deduct, debited. Other lexicons performed poorly due to the word coverage and labels. The most notable problem was the handling of negation. Some of the lexicons compared did not put negation into consideration. For example, tweets like, "*not able to send dm please rectify issue someone made transaction to my account since yesterday received alert no longer funny referred twitter send dm yet no one responded"* has been misclassified by NRC, Afinn, VADER as positive, and Textblob as neutral. Afinn is a popular sentiment lexicon, however, the Afinn algorithm does not treat negation. Figure 4 below shows an example of *"not good"* in the Afinn web sentiment classifier. Afinn reported *"not good"* as a positive class with a score of 3. Unfortunately, this was one of the problems encountered in the analysis.

Fig 4. Afinn Sentiment classifier

In order to gain a comprehensive understanding of the performance of these lexicons in terms of their predicted classes, a plot was generated to illustrate the correct predictions. Figure 5 below demonstrates the correct classification of positive, negative, and neutral classes by SentiLeye algorithms.

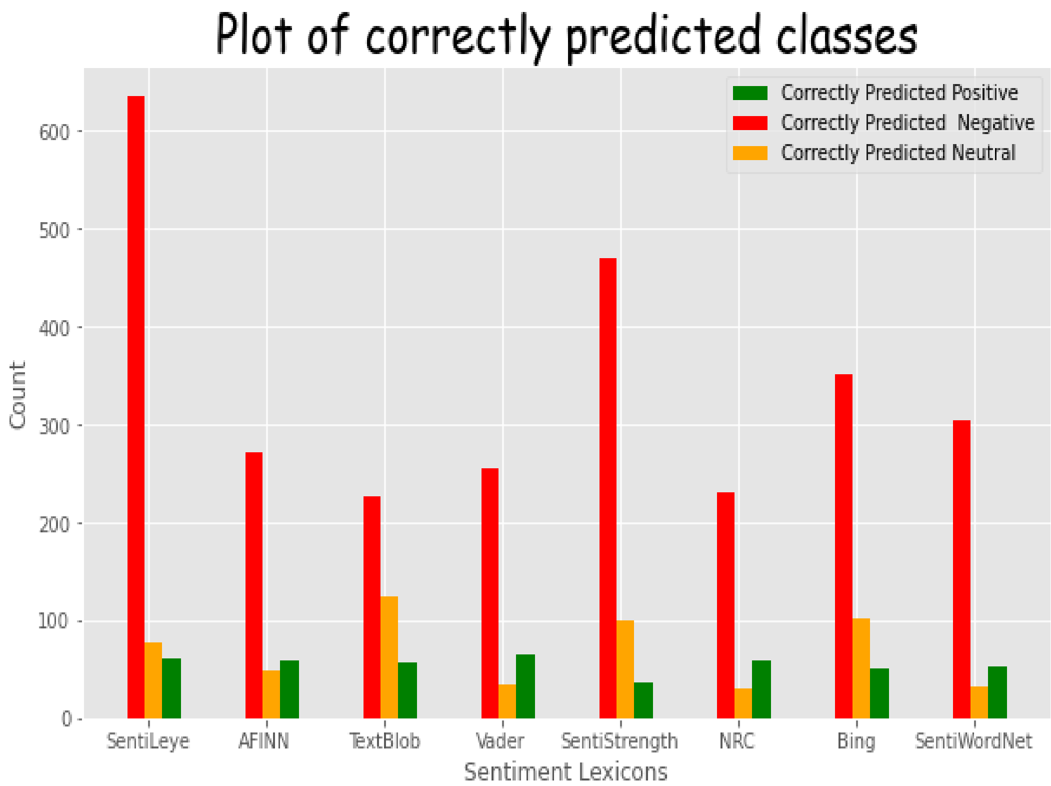

Fig 5. A visual representation of the correctly predicted sentiment classes

From Figure 5 above, we showed SentiLeye captured the negative tweets more precisely. However, struggled with the neutral tweets and performed averagely with the positive tweets. This is because there were tweets where the banks' Twitter handler was trying to apologise for service experienced and praising the customers for their patience whilst trying to resolve the issue. For example, tweets like, *"sorry for inconvenience caused would like assist kindly send account number dm help handle help address quickly possible"* or *"thank you for contacting us, to enable us assist you today kindly fill-in a complaint form and not inquiry form"*. This has been labelled neutral because it only signifies that bank X is now responding to the case. However, it is worth noting that the sentence does not demonstrate that experience is now good therefore we have labelled this neutral in the evaluation dataset (manually annotated test set). However, word like "*inconvenience"* was labelled negative, "*thanks"* is positive in the lexicon and thus misleads the algorithm. As illustration, we observed that our lexicon misclassifies neutral tweets as positive in two different cases. Case one is a situation when objective statement was tweeted politely by the customer. For example, "*kindly send my atm card to this address thank you dear bank X".* The second scenario is when banks try to market their product via their Twitter channels. This is because the tweets come with a lot of positive terms as they intend to sell their products. Thus, justifies why the algorithm in this case predicts as positive which in practice is a marketing tweet. Among the lexicons, TextBlob achieved the highest accuracy in correctly classifying neutral tweets, closely followed by Bing and SentiStrength. SentiLeye, on the other hand, demonstrated average performance in classifying neutral tweets. When it comes to positive tweets, SentiStrength faced challenges in accurately predicting the positive sentiment compared to other lexicons. Surprisingly, VADER emerged as the most effective in correctly classifying positive tweets, with Afinn in the second position. In terms of negative tweets, SentiLeye and SentiStrength achieved the highest accuracy, while TextBlob and NRC struggled significantly. Given the significance of accurately classifying negative tweets in this context, SentiStrength demonstrated commendable performance. However, SentiLeye exhibited the best overall performance among the lexicons.

A major limitation of lexicon-based approach is the fixed numbers of words in the lexicon and their label. This means the lexicon classifier cannot recognise words not included in the lexicon thus does limit the classification of such words. For example, tweets like, *"that's a sure bank"* is a positive tweet in Pidgin English context. However, the tweet was classified by Afinn, Bing and SentiStrength as neutral, while NRC reported this as zero sentiment and emotion category trust (1). The tweet was misclassified because the word *"sure"* does not exist in most of the lexicon. NRC recognises the word but was labelled trust in their lexicon and not as positive or negative sentiment. Another challenge encountered by the lexicon is the context factor. For example, tweets like *"dear beloved @bank X, I can no longer stand the long queue, unreliable atm"* The tweet was classified by Afinn as positive, SentiStrength as positive, NRC as positive and Bing as zero (0). The tweet was misclassified because these lexicons recognise the word "*dear beloved*" as positive, "*no*" as negative and "*unreliable*" does not exist in lexicons like Afinn. The word "*long*" is a negative word in the banking context; however, the lexicons do not recognise the word. Words like "*long*" are context based and are used differently. An example to show the contextual difference is the word "*long*" as regards long battery life in the mobile industry and long queue in the banking industry which means positive and negative sentiment respectively. Furthermore, some of the lexicons were unable to recognise the intensified adjective as an accompanying word. For example, tweets like *"Nigerian banks are the greatest greatest fraud, deduct acct maintenance fee"* was classified by SentiStrength, Afinn as positive and Bing as neutral. In this case, sentiment lexicons up toned the word *"greatest"* with the previous word *"greatest"* and does not affect the word "fraud" which strengthens the positive sentiment scores rather than strengthen the negative word "fraud". As a result, the lexicons classified the tweet as a positive sentiment. It will be helpful to adjust the algorithm to understand the word *"fraud"* in this context so that the lexicon algorithm can understand that the *"greatest"* values should affect *"fraud"*. Lastly, an important problem encountered was Pidgin English terms. The presence of Pidgin English observed in the tweet has affected the sentiment classifiers. An example is *"I taya oooo .....we pay for deposits, sms, cheques, ATM cards, like everything*" This was classified by as positive by Afinn, SentiStrength, NRC and Bing. In this example above, the bank customer expressed his/her frustration towards payment for everything in a sarcastic manner using the Pidgin English term *"taya"* to express the negativity. Unfortunately, the lexicon algorithms could not recognise the Pidgin English and had therefore misclassified as positive sentiment.

During the experimentation phase, SentiStrength, Afinn, Bing, and NRC wordlists were updated to incorporate Pidgin English terms. Unfortunately, these lexicons showed minimal improvement of approximately 2% accuracy. This can be attributed to the fact that lexicon-based approaches heavily rely on their wordlists, and if a word is not included in the lexicon, its semantic orientation cannot be captured. Some of the lexicon algorithms could have achieved even better results if they had considered opinionated-objective words. For instance, the term *"debited"* was categorized as a neutral word in the Afinn corpus, whereas SentiStrength classified it as negative, which explains why SentiStrength produced better results. In summary, our study aligns with the recommendation of Lin (2022), emphasizing the importance of developing lexical resources in under-resourced domains to enhance the performance of sentiment analysis. It is crucial to consider language, context, and domain-specific words to improve the effectiveness of lexicon-based approaches in sentiment analysis. General-purpose lexicons like Bing, SentiWordNet, VADER, TextBlob, and Afinn are less accurate when applied to specific contexts compared to SentiLeye. For example, a tweet such as *"the thing swallow atm card no gimme money received debit alert like min ago"* would be misclassified by these general-purpose lexicons, as they were developed using broad lexical knowledge. This finding is consistent with that of Shang et al. (2023) & Palmer et al. (2020) that showed domain specific dictionaries perform better than the general-purpose dictionaries for SA, and specifically in financial context. Wu et al. (2023) added that results from domain specific algorithms evidence that general models cannot replace them. Thus, we summarise our findings that there is need for sentiment lexicons to consider updating their wordlist (with domain-dependant terms, opinionated-objective terms, Pidgin English words, and negation) regularly to remain competitive, as lexicon-based approach remains a simple, fast, and efficient approach to perform sentiment analysis where large training dataset is unavailable.

5 Conclusion

Our paper proposes SentiLeye, a novel sentiment lexicon algorithm specifically developed to capture Nigerian Pidgin English and English terms within the banking domain. We showed that SentiLeye outperformed off-the-shelf sentiment lexicons with an impressive F1 score of 76%. Our study outlines the limitations of general-purpose lexicons in context-specific domains such as banking, as they are derived from general lexical knowledge and lack accurate labelling of words in their lexicon. Furthermore, our research emphasizes the importance of incorporating opinionated-objective terms into existing lexicons, particularly when working with service industry datasets like banking. This finding aligns with the work of Ogunleye (2021) and Liu (2015), both of which emphasized the need for the creation of opinionated-objective words. Additionally, our comparison of lexicons highlights the significance of considering negation within lexical algorithms. This finding is consistent with previous studies (Bos & Frasincar, 2022; Gupta & Joshi, 2021; Garg & Subrahmanyam, 2021; Mukherjee et al., 2021; Singh & Paul, 2021) that highlighted the improvements in sentiment classification performance achieved through effective negation handling. Lastly, our research highlights the importance of creating wordlists for non-English text, where necessary, as this significantly enhances the performance of sentiment analysis (SA) lexicons. This finding aligns with the work of Kaity & Balakrishnan (2019), who emphasized that the development of non-English lexicons greatly enhances SA performance in domains where non-English terms are prevalent. In conclusion, our study provides evidence supporting the inclusion of domain-specific terms, opinionated factual words, negation handling, and consideration of language in the design of future lexicon algorithms for sentiment analysis.

5.1 Theoretical and Practical Implications

Our paper demonstrates the concepts of using simple implementation to achieve human interpretable intelligent sentiment analysis system. We evidenced that adjectives like good or bad are not sufficient to detect the sentiment polarity in text especially in the service industry where customers express their experience of service operation. We established that objective terms that interprets to opinions of service experience improves the sentiment classification of customer discussions. Furthermore, we established that domain specific systems cannot be replaced by general purpose systems especially in a narrow context like the banking due to the diversity of bank sentiment words and domain specific terms. Our lexical resources can be used as external knowledge for machine learning models as explored by Shang et al. (2023). In practice, our implementation is useful for the government to inform their decision-making processes. Currently, Nigeria is unsettled due to the redesign of currency notes (CNN, 2023). Our implementation provides an intelligent system to the government to understand the citizens feelings and their perception about the new currency policy. This can then inform their decision prior and post implementation of the redesign policy. On the other hand, our algorithm is useful for banks to monitor and generate insight on their products and service and most importantly, understand customers' attitude as the banks rely on them for stability and profitability.

5.2 Limitation and Future work

For evaluation of our lexicon algorithm, unfortunately, we were unable to obtain another domain specific dataset to suit the context of this study. This could have helped us extend the evaluation process of our algorithm. It is important to note that the implementation of our algorithm is currently limited to the banking domain, as it has not been validated in other domains. The SentiLeye lexicon is specifically designed for English and Pidgin English words and has not yet been validated for words in Nigerian local languages such as Yoruba, Igbo, and Hausa. Unfortunately, the local language words were found in the tweets used in this study. For future research, the implementation can be enhanced by employing knowledge transfer techniques to adapt the lexicon for use in other domains. Secondly, topic models (Ogunleye et al. 2023) can be deployed to improve the lexicon performance as this will detect the aspect customers are talking about. In addition, we encourage the use of pre-trained large language models (LLMs) such as BERT on pidgin English. Lastly, our algorithm can be extended to detect customers' attitude on multimodal dataset.

Resources

1). https://www.kaggle.com/datasets/batoog/dirtybanktweets-for-research
2). https://www.kaggle.com/datasets/batoog/labelled-pidgin-english-wordlist-for-research
3). https://www.kaggle.com/datasets/batoog/labelled-pidgin-bank-tweet-1000

References

1. Afolabi, I. T., Sowunmi, O. Y., & Adigun, T. (2019). Semantic Text Mining using Domain Ontology. In proceedings of the World Congress on Engineering and Computer Science (pp. 1-6).
2. 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.
3. Ahmad, M., Aftab, S., Muhammad, S. S., & Waheed, U. (2017). Tools and techniques for lexicon driven sentiment analysis: a review. International Journal of Multidisciplinary Sciences and Engineering, volume 8(1), 17-23.
4. Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., & Vilares, J. (2021). Sentiment analysis for fake news detection. Electronics, 10(11), 1348.
5. 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).
6. Al-Twairesh, N., Al-Khalifa, H., & Al-Salman, A. (2016). AraSenTi: large-scale twitter-specific Arabic sentiment lexicons. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 697-705).
7. Amusa, L., Yahya, W. B., & Balogun, A. O. (2016). Data mining of Nigerians' sentiments on the administration of federal republic of Nigeria.
8. Anastasiou, D., & Katsafados, A. (2023). Bank deposits and textual sentiment: When an European Central Bank president's speech is not just a speech. The Manchester School.
9. Atoum, J. O. (2023). Detecting Cyberbullying from Tweets Through Machine Learning Techniques with Sentiment Analysis. In Advances in Information and Communication: Proceedings of the 2023 Future of Information and Communication Conference (FICC), Volume 2 (pp. 25-38). Cham: Springer Nature Switzerland.
10. Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN Computer Science, 3, 1-20.
11. Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec (Vol. 10, No. 2010, pp. 2200-2204).
12. Bhattacharya, S., Sarkar, D., Kole, D. K., & Jana, P. (2022). Recent trends in recommendation systems and sentiment analysis. Advanced Data Mining Tools and Methods for Social Computing, 163-175.
13. Bos, T., & Frasincar, F. (2022). Automatically building financial sentiment lexicons while accounting for negation. Cognitive Computation, 14(1), 442-460.
14. Boukes, M. (2020) The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms.
15. Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. Electronics, 11(3), 374.
16. Catelli, R., Pelosi, S., Comito, C., Pizzuti, C., & Esposito, M. (2023). Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy. Computers in Biology and Medicine, 106876.
17. Choi, Y., & Cardie, C. (2009). Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In Proceedings of the 2009 conference on empirical methods in natural language processing (pp. 590-598).
18. Choudhary, C., Singh, I., & Kumar, M. (2023). SARWAS: Deep ensemble learning techniques for sentiment-based recommendation system. Expert Systems with Applications, 216, 119420.

19. CNN (2023), Nigeria delays plans to replace its banknotes after chaotic scenes at ATMs. https://edition.cnn.com/2023/02/08/business/nigeria-new-banknotes-delay-intl-lgs/index.html accessed on 7th March, 2023
20. Consoli, S., Barbaglia, L., & Manzan, S. (2022). Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. Knowledge-Based Systems, 247, 108781.
21. Correa, R., Garud, K., Londono, J. M., & Mislang, N. (2021). Sentiment in central banks' financial stability reports. Review of Finance, 25(1), 85-120.
22. Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2023). Mining and searching app reviews for requirements engineering: Evaluation and replication studies. Information Systems, 114, 102181.
23. Das, S. R. (2014). Text and context: Language analytics in finance. Foundations and Trends in Finance, 8(3), 145-261.
24. Das, S. R., Donini, M., Zafar, M. B., He, J., & Kenthapadi, K. (2022). FinLex: An effective use of word embeddings for financial lexicon generation. The Journal of Finance and Data Science, 8, 1-11.
25. Dashtipour, K., Gogate, M., Gelbukh, A., & Hussain, A. (2022). Extending persian sentiment lexicon with idiomatic expressions for sentiment analysis. Social Network Analysis and Mining, 12, 1-13.
26. De Jesus, D. P., & Da Nóbrega Besarria, C. (2023). Machine learning and sentiment analysis: Projecting bank insolvency risk. Research in Economics.
27. De-Melo, T. (2022). SentiLexBR: An Automatic Methodology of Building Sentiment Lexicons for the Portuguese Language. Journal of Information and Data Management, 13(3).
28. Denecke, K., & Reichenpfader, D. (2023). Sentiment analysis of clinical narratives: A scoping review. Journal of Biomedical Informatics, 104336.
29. Du, K., Xing, F., & Cambria, E. (2023). Incorporating Multiple Knowledge Sources for Targeted Aspect-based Financial Sentiment Analysis. ACM Transactions on Management Information Systems.
30. Farah, H. A., & Kakisim, A. G. (2023). Enhancing Lexicon Based Sentiment Analysis Using n-gram Approach. In Smart Applications with Advanced Machine Learning and Human-Centred Problem Design (pp. 213-221). Cham: Springer International Publishing.
31. Fehle, J., Schmidt, T., & Wolff, C. (2021). Lexicon-based sentiment analysis in german: Systematic evaluation of resources and preprocessing techniques.
32. Gandhi, P., Loughran, T., & McDonald, B. (2019), Using Annual Report Sentiment as a Proxy for Financial Distress in US Banks. Journal of Behavioral Finance, 1-13.
33. Garg, S. B., & Subrahmanyam, V. V. (2021). A Survey on Various Negation Handling Techniques in Sentiment Analysis. Smart and Sustainable Intelligent Systems, 259-280.
34. Ghosh, P., Dutta, R., Agarwal, N., Chatterjee, S., & Mitra, S. (2023). Social Media Sentiment Analysis on Third Booster Dosage for COVID-19 Vaccination: A Holistic Machine Learning Approach. Intelligent Systems and Human Machine Collaboration, 179-190.
35. Gouthami, S., & Hegde, N. P. (2023). Automatic Sentiment Analysis Scalability Prediction for Information Extraction Using SentiStrength Algorithm. In Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems (pp. 21-30). Springer, Singapore.
36. Gui, L., & He, Y. (2021). Understanding patient reviews with minimum supervision. Artificial Intelligence in Medicine, 120, 102160.
37. 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.
38. Hatzivassiloglou, V., & McKeown, K. (1997). Predicting the semantic orientation of adjectives. In 35th annual meeting of the association for computational linguistics and 8th conference of the European chapter of the association for computational linguistics (pp. 174-181).
39. 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). ACM.
40. 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
41. Hutto, C., & Gilbert, E. (2014). 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).
42. Islam, M. R., & Zibran, M. F. (2018). SentiStrength-SE: Exploiting domain specificity for improved sentiment analysis in software engineering text. Journal of Systems and Sof tware, 145, 125-146.
43. Iwendi, C., Mohan, S., Ibeke, E., Ahmadian, A., & Ciano, T. (2022). Covid-19 fake news sentiment analysis. Computers and electrical engineering, 101, 107967.
44. Jha, A. K., & Shah, S. (2021). Disconfirmation effect on online review credibility: An experimental analysis. Decision Support Systems, 145, 113519.
45. Kaity, M., & Balakrishnan, V. (2019). An automatic non-English sentiment lexicon builder using unannotated corpus. The Journal of Supercomputing, 1-26.

46. Kaity, M., & Balakrishnan, V. (2020). Sentiment lexicons and non-English languages: a survey. Knowledge and Information Systems, 62(12), 4445-4480.
47. Kanayama, H., & Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In Proceedings of the 2006 conference on empirical methods in natural language processing (pp. 355-363).
48. Kearney, C., & Liu, S. (2014), Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171-185.
49. 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.
50. 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.
51. Krishnamoorthy (2018), Sentiment analysis of financial news articles using performance indicators. Knowledge and Information Systems, 56(2), 373-394.
52. Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A deep learning-based approach to constructing a domain sentiment lexicon: a case study in financial distress prediction. Information Processing & Management, 58(5), 102673.
53. Lin, B., Cassee, N., Serebrenik, A., Bavota, G., Novielli, N., & Lanza, M. (2022). Opinion mining for software development: a systematic literature review. ACM Transactions on Software Engineering and Methodology (TOSEM), 31(3), 1-41.
54. Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge university press.
55. Long, S., Lucey, B., Xie, Y., & Yarovaya, L. (2023). “I just like the stock”: The role of Reddit sentiment in the GameStop share rally. Financial Review, 58(1), 19-37.
56. Loughran, T., & McDonald, B. (2011), When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.
57. Machová, K., Mikula, M., Gao, X., & Mach, M. (2020). Lexicon-based sentiment analysis using the particle swarm optimization. Electronics, 9(8), 1317.
58. Man, X., Luo, T., & Lin, J. (2019). Financial sentiment analysis (FSA): A survey. In 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS) (pp. 617-622).
59. 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.
60. Martinis, M. C., Zucco, C., & Cannataro, M. (2022). An Italian lexicon-based sentiment analysis approach for medical applications. In Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (pp. 1-4).
61. Miller, G. A. (1995), WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.
62. Mittal, D., & Agrawal, S. R. (2022). Determining banking service attributes from online reviews: text mining and sentiment analysis. International Journal of Bank Marketing.
63. Mohammad, S. M., & Turney, P. D. (2013), Crowdsourcing a word–emotion association lexicon. Computational Intelligence, 29(3), 436-465
64. 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.
65. Nielsen, F. Å. (2011), A new ANEW: Evaluation of a word list for sentiment analysis in microblogs
66. O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010), From tweets to polls: Linking text sentiment to public opinion time series. In Fourth International AAAI conference on weblogs and social media.
67. Ogueji, K., & Ahia, O. (2019). Pidginunmt: Unsupervised neural machine translation from west african pidgin to english.
68. Ogunleye, B. O. (2021). Statistical learning approaches to sentiment analysis in the Nigerian banking context (Doctoral dissertation, Sheffield Hallam University).
69. Ogunleye, B., Maswera, T., Hirsch, L., Gaudoin, J., & Brunsdon, T. (2023). Comparison of Topic Modelling Approaches in the Banking Context. Applied Sciences, 13(2), 797.
70. Oyewola, D. O., Oladimeji, L. A., Julius, S. O., Kachalla, L. B., & Dada, E. G. (2023). Optimizing sentiment analysis of Nigerian 2023 presidential election using two-stage residual long short term memory. Heliyon.
71. 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).
72. Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 71(2001), 2001.
73. Qian, C., Mathur, N., Zakaria, N. H., Arora, R., Gupta, V., & Ali, M. (2022). Understanding public opinions on social media for financial sentiment analysis using AI-based techniques. Information Processing & Management, 59(6), 103098.

74. Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016), Sentibench benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Science, 5(1), 23.
75. Ringsquandl, M., & Petkovic, D. (2013), Analyzing political sentiment on Twitter. In 2013 AAAI Spring Symposium Series.
76. Rizk, R., Rizk, D., Rizk, F., & Hsu, S. (2023). 280 characters to the White House: predicting 2020 US presidential elections from twitter data. Computational and Mathematical Organization Theory, 1-28.
77. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5-19.
78. Schmidt, T., Dangel, J., & Wolff, C. (2021). SentText: A tool for lexicon-based sentiment analysis in digital humanities.
79. Shang, L., Xi, H., Hua, J., Tang, H., & Zhou, J. (2023). A Lexicon Enhanced Collaborative Network for targeted financial sentiment analysis. Information Processing & Management, 60(2), 103187.
80. Singh, P. K., & Paul, S. (2021). Deep learning approach for negation handling in sentiment analysis. IEEE Access, 9, 102579-102592.
81. 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.
82. Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (pp. 380-385).
83. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.
84. 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.
85. 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.
86. 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.
87. Veltmeijer, E., & Gerritsen, C. (2023). SentiMap: Domain-Adaptive Geo-Spatial Sentiment Analysis. In 2023 IEEE 17th International Conference on Semantic Computing (ICSC) (pp. 17-24). IEEE.
88. Wang, Q., Su, T., Lau, R. Y. K., & Xie, H. (2023). DeepEmotionNet: Emotion mining for corporate performance analysis and prediction. Information Processing & Management, 60(3), 103151.
89. Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of human language technology conference and conference on empirical methods in natural language processing (pp. 347-354).
90. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564.
91. Yadav, A., Jha, C. K., Sharan, A., & Vaish, V. (2020). Sentiment analysis of financial news using unsupervised approach. Procedia Computer Science, 167, 589-598.
92. Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S. (2021). An analysis of COVID-19 vaccine sentiments and opinions on Twitter. International Journal of Infectious Diseases, 108, 256-262.
93. Zhou, Z., Zhao, W., & Shang, L. (2014). Sentiment analysis with automatically constructed lexicon and three-way decision. In Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9 (pp. 777-788). Springer International Publishing.
94. Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018), The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. ACM Transactions on Management Information Systems (TMIS), 9(2), 5.
95. Alnawas, A. (2022, December). Stacked Bi-directional Long Short-Term Memory model for Multi-Class Arabic Sentiment Analysis on Covid-19. In 2022 3rd Information Technology to Enhance e-learning and Other Application (IT-ELA) (pp. 191-195). IEEE.