Electric Vehicle Sentiment Analysis Using Large Language Models

SHARMA, Hemlata <http://orcid.org/0000-0002-7566-4413>, UD DIN, Faiz and OGUNLEYE, Bayode

Available from Sheffield Hallam University Research Archive (SHURA) at:

https://shura.shu.ac.uk/34572/

Citation:

Electric Vehicle Sentiment Analysis Using Large Language Models

Hemlata Sharma [1] , **Faiz Ud Din** [1] **and Bayode Ogunleye** [2,*]

1   Department of Computing, Sheffield Hallam University, Sheffield S1 2NU, UK; h.sharma@shu.ac.uk (H.S.); f.uddin@barnsley.ac.uk (F.U.D.)
2   Department of Computing & Mathematics, University of Brighton, Brighton BN2 4GJ, UK
*   Correspondence: b.ogunleye@brighton.ac.uk

Abstract: Sentiment analysis is a technique used to understand the public's opinion towards an event, product, or organization. For example, sentiment analysis can be used to understand positive or negative opinions or attitudes towards electric vehicle (EV) brands. This provides companies with valuable insight into the public's opinion of their products and brands. In the field of natural language processing (NLP), transformer models have shown great performance compared to traditional machine learning algorithms. However, these models have not been explored extensively in the EV domain. EV companies are becoming significant competitors in the automotive industry and are projected to cover up to 30% of the United States light vehicle market by 2030 In this study, we present a comparative study of large language models (LLMs) including bidirectional encoder representations from transformers (BERT), robustly optimised BERT (RoBERTa), and a generalised autoregressive pre-training method (XLNet) using Lucid Motors and Tesla Motors YouTube datasets. Results evidenced that LLMs like BERT and her variants are off-the-shelf algorithms for sentiment analysis, specifically when fine-tuned. Furthermore, our findings present the need for domain adaptation whilst utilizing LLMs. Finally, the experimental results showed that RoBERTa achieved consistent performance across the EV datasets with an F1 score of at least 92%.

Keywords: BERT; electric vehicles; large language models; machine learning; natural language processing; RoBERTa; sentiment analysis; XLNet

1. Introduction

Electric vehicles (EVs) are pivotal to attaining the zero-emission target set for 2050 to meet environmental challenges [1]. Since the start of the first mass marketing of an EV, the Nissan Leaf, in 2010, the EV market has grown exponentially [2]. With the great success of companies like Tesla, new companies have emerged in the EV market and shown great potential to provide great products. EV sales have grown from almost three million cars sold worldwide in 2020 to ten million in 2022 [3]. Specifically, 370,000 EVs were sold in the UK in 2022 which evidences the growth of the EV market. Several countries have implemented policies to encourage the adoption of EVs. For example, the UK government constructed several electric charging points across the country. Similarly, China implemented financial subsidies for EV purchases [4]. Despite the potential benefits of adopting EVs, society at large seems doubtful about fully adopting the cars. Previous studies have identified the factors influencing consumers' adoption of EVs as limited awareness [5], battery life [6], and national policies [7]. The authors of [8] found that attitudes to innovation and functional performance are key elements of EV adoption. Furthermore, the findings of Carley et al. [9] showed that people with a high level of education showed favourable intentions towards purchasing plug-in EVs, but an individual's point of view and their understanding of the advantages and disadvantages of EVs are significant factors. However, it is worth noting that the majority of existing studies have utilised survey questionnaires to investigate

people's attitudes towards EVs. Unfortunately, questionnaires are limited to pre-defined variables [10]. Another channel to mine people's perspectives is sentiment analysis performed on user-generated content (UGC) such as user Twitter data (Tweets). This approach has been used extensively in the literature [10,11].

Sentiment analysis (SA) aims to analyse public opinion towards events, issues, and products. It is considered a branch of machine learning, data mining, and natural language processing (NLP). Although NLP began in the 1950s, SA gained more attention in 2005 due to social media's popularity and the availability of big (text) data. SA provides companies the opportunity to understand their business better by extracting insights from customers' data. The authors of [12] stated that consumers consider other people's reviews for purchasing purposes. This evidences the impact of reviews, comments, and information shared on a product publicly. In the automotive industry, newly emerged companies like Lucid Motors have challenged the big EV players like Tesla. Thus, it is worth understanding the perception of consumers. Since the development of BERT [13] and her variants, LLMs (advanced artificial intelligence models designed to understand, generate, and manipulate human language) have taken the world by storm with their performance. Previous studies have proposed the use of LLMs for SA [14,15]. However, there is a dearth of studies on SA in the EV context, specifically with regard to the application of sophisticated approaches like those of LLMs. In addition, understanding public sentiment can significantly impact EV sales, which in turn may contribute to mitigating challenges such as global warming, by reducing air pollution [5]. Thus, this study aims to address this gap. To this end, this paper aims to perform SA to assess consumers' perceptions of Lucid and Tesla EVs and thus evaluate the effectiveness of LLMs. To conclude, the main contributions of this study can be summarised as follows:

- This paper demonstrates the need to fine-tune LLMs for domain adaptation and thus proposes the use of a fine-tuned RoBERTa algorithm for EV sentiment prediction;
- Our paper demonstrates an SA approach that takes advantage of the language understanding of the transformer models to complement a lexicon-based approach when labelled datasets are unavailable;
- We conduct an experimental comparison of LLMs in the EV context and thus present state-of-the-art (SOTA) results.

The rest of the paper is structured into five main parts. Section 2 will centre on a literature review, Section 3 highlights the methodological and evaluation process, the next section presents the results, while the last section presents our conclusions and recommendations.

2. Related Work

SA provides an analytical technique to understand customers' perspectives. SA can be performed at different levels, namely, document, sentence, and aspect levels. The document level involves classifying the sentiment of an entire document into positive, negative, or neutral. The sentence level involves classifying the sentiment of a sentence or Tweet whilst the aspect level involves classifying an entity by recognising the sentiment polarity of its aspect. In general, there are two main approaches to SA, namely, a lexicon based-approach and a machine learning (ML) approach [11]. The former approach relies on bags of words and a set of rules to classify documents, whilst the later approach can be further divided into three main categories, namely, a supervised ML approach, an unsupervised ML approach, and a semi-supervised ML approach. The supervised ML approach is the most popular sentiment classification approach. The approach uses a subset of the labelled dataset (target variable is known) for training purposes and the remainder for testing purposes. For example, the study of [16] used 80% of 2847 Indonesia-labelled Tweets for training and 20% for testing. They showed that a Support Vector Machine (SVM) outperformed a Convolutional Neural Network (CNN), logistic regression, random forest, gradient boosting, and a Recurrent Neural Network (RNN) with an accuracy of 75.08% and F1 score of 78%. In the case of a semi-supervised ML approach, this technique is suitable when only a limited labelled dataset is available for training purposes. An example of

this approach was demonstrated in the study of [17] that used a variational autoencoder for aspect-based SA. The unsupervised ML approach does not require a labelled dataset. The approach focuses on uncovering hidden patterns and making predictions from an unlabelled dataset [18]. In the context of SA, the authors of [19] successfully utilised k-means clustering to perform SA. However, the unsupervised ML approach is unpopular due to word ambiguity which often affects techniques like clustering.

Most studies found use a supervised ML approach. For example, the authors of [20] showed that CNN outperformed SVM, the Doc2Vec (paragraph vector) algorithm, and RNN in a consumer sentiment classification task in the Indian EV market. In China, ref. [4] used a Weibo dataset to mine consumers' sentiments towards EVs. They showed that Bidirectional Long Short-Term Memory (Bi-LSTM) with an attention layer outperformed SVM, CNN, Bi-LSTM, and a combination of CNN and LSTM with an F1 score of 86%. Their findings showed that growth rates towards EVs varied across regions and in terms of gender: men pay more attention to EVs. The problem of utilising a supervised approach is the fact that ML algorithms require large amounts of data for training purposes. Unfortunately, the large, labelled sets are not readily available in most cases. However, there are various ways to label data for sentiment classification tasks. Firstly, there is human labelling, where experts are recruited for the purpose of providing sentiment labels for data. This process yields high quality labelled datasets which can also be regarded as the ground truth labels. For example, the authors of [21] employed expert-annotated data and compared BERT, XLNet, LSTM, and CNN, and thus showed that BERT achieved the best results with an F1 score of 83%. However, the human (expert) annotation of data labelling is expensive and time-consuming [22]. Alternatively, crowdsourcing is another means of data labelling. This involves recruiting online annotators to provide labels for the data. There are popular platforms for this purpose. For example, Amazon Mechanical Turk and Rent-A-Coder. However, it is worth stating that data obtained from this process are prone to error, bias, and are of low-quality labels. This approach is unreliable because the availability of labellers is not certain. Based on these challenges, past studies used lexicons for data labelling. For example, Ref. [23] compared AFINN, TextBlob, and VADER (Valence Aware Dictionary and sEntiment Reasoner) for assigning sentiment labels. In their experiment, they showed the AFINN sentiment labels are close to human labels. Similarly, VADER [24,25], TextBlob [25,26], and AFINN [25,27] have been used for assigning sentiment labels to datasets. The authors of [28] constructed two datasets from Reddit and Twitter between January 2011 and December 2020 to understand the public's perception of EVs. Thus, they used the VADER, LIWC (Linguistic Inquiry and Word Count), and AFINN lexicons to generate insights from the data.

To conclude, the application of SA in the EV context is relatively new and as such research associated with this current topic is limited. Our literature review findings showed that LLMs have not been fully deployed in this context. More specifically, the existing literature has not utilised transformer models for targeted YouTube SA regarding EV corporations (such as Lucid Motors and Tesla Motors). It remains unclear which architecture would perform best in this niche domain. A domain-specific benchmarking of transformer capabilities on automotive brand-related YouTube comments would provide both industry and academic value, especially for research benchmarking. Moreover, if companies gain insight into public sentiment, it can influence EV sales, ultimately helping to address issues like global warming. This study aims to address this gap and also generate actionable insights on consumer perceptions regarding new automotive brands.

3. Methodology

This section presents the methods applied in this study. For our experiment, we collected 19,991 YouTube comments randomly which spanned from 2014 to 2024. Due to limited traction on social media before 2014 on YouTube, we utilised YouTube API [29] to scrape these comments from YouTube (YouTube Data API | Google for Developers).

SA involves preprocessing text to remove noise from the data, followed by assessing subjectivity. Polarity is then determined using either ML or lexical methods, categorising content as positive, negative, or neutral. Context-dependent knowledge is crucial, as words can have multiple meanings. Proper contextual application improves sentiment classification accuracy [30].

We performed several data preprocessing techniques as follows:

- Duplicate data: ensuring that duplicate data is removed using unique comment IDs;
- Removing unnecessary items: eliminating irrelevant elements from the text, including blank spaces, stop words (e.g., "a", "the", "is", "are"), hashtags, emojis, URLs, numbers, and special characters;
- Lowercasing: converting all text to lowercase for smoother processing;
- Whitespace removal: eliminating unnecessary or excessive white spaces in the text.

3.1. Data Labelling

There are many Python libraries to perform data labelling for an SA task. The most popular libraries are AFINN, TextBlob, and VADER. These lexicon approaches have been praised for their performance across several domains which is due to their general lexical knowledge [11]. Several studies have used these lexicons for data labelling [23–27]. More recently, the study of [31] compared the sentiment labels of TextBlob, VADER, and Azure fitted into ML algorithms. Their study showed that SVM fitted with TextBlob labels achieved the best performance. In this study, we consider TextBlob for data labelling. The dictionary consists of 2918 words. The polarity ranges from -1 to 1, where -1 represents a very negative sentiment, 1 represents a very positive sentiment, and 0 represents a neutral sentiment. We calculated the sentiment polarity of the comments. For the evaluation of the data labelling, we randomly selected 500 samples and compared the labels against ground truth labels (conducted by the authors). TextBlob labels achieve 67% in accuracy and 64% in F1 score.

3.2. Transformer-Based ML Models

This paper employed three main transformer models, namely, BERT, XLNet, and RoBERTa based on their performance in an NLP task. In subsequent sections, we discuss these three LLMs.

3.2.1. Bidirectional Encoder Representations from Transformers (BERT)

BERT (proposed by [13]) is a self-supervised autoencoder (AE) language model for training NLP systems. BERT is pre-trained on a large-scale Wikipedia corpus using a masked language model (MLM) and next-sentence prediction tasks. The base version consists of 12 layers of transformer blocks, 768 hidden layer sizes, and 12 self-attention heads, while the large version consists of 24 layers of transformer blocks, 1024 hidden layer sizes, and 16 self-attention heads. The objective is to predict the actual vocabulary ID of a masked word only based on its context after randomly masking some of the tokens from the input. For example, in the sentence, "*I went to UoBrighton to meet a [MASK] last week*", BERT aims to predict the masked word (token) by outputting embeddings. The MLM's intent permits the representation to combine the left and the right context, in contrast to the left-to-right language model pre-training, which allows us to pre-train a deep bidirectional transformer. To fine-tune the pre-trained BERT model, the model is first instantiated with default parameters (used when pre-trained), and then the parameters are fine-tuned using labelled data from downstream tasks (text classification in our case). The main components of the iteration can thus be summarised as follows:

- Input embedding: In this stage, the process of tokenisation occurs. This is the breaking down of text into smaller tokens for numerical encoding. Afterwards, the tokens are transformed into continuous vector representations (token embedding).
- Positional encoding: the position encoding of the tokens is calculated using sine or cosine functions (as an example) and thus added to the token embeddings.

- Self-attention: The aim is to detect how similar each token is to others. The process involves generating the query and key matrices. Afterwards, the value is calculated (vectors) using the dot product.
- Normalisation layer: the SoftMax function helps normalise the vectors.
- Classification head: Converting sequential outputs into classification results. The SoftMax function helps normalise class scores into probability values.
- Training loss: measuring the difference between predicted probabilities and true labels, often using loss functions like cross-entropy.
- Optimisation: updating model parameters to minimise loss using the Adam algorithm (backpropagation).

3.2.2. Robustly Optimised BERT Approach (RoBERTa)

Facebook AI Research (FAIR) proposed a robustly optimised BERT approach (RoBERTa) in 2019. The authors of [32] criticised BERT for being undertrained, and thus modified the training method by (i) using a dynamic masking pattern instead of a static one; (ii) training with more data with large batches; (iii) removing next sentence prediction; and (iv) training on longer sentences and proposing RoBERTa. As a result, RoBERTa outperforms BERT in terms of the masked language modelling objective and performs better on downstream tasks. RoBERTa has been shown to outperform BERT in several NLP tasks [33,34].

3.2.3. XLNet

BERT was criticised for ignoring the dependency between the masked positions which leads to pre-train–finetune discrepancies because it relies on masking the input to corrupt it. Thus, the authors of [35] proposed XLNet. XLNet is a permutation-based autoregressive transformer that combines the finest aspects of autoencoding and autoregressive language modelling while seeking to get around their drawbacks. In autoregressive models, the past values of a variable are used to predict future values. XLNet training objectives calculate the likelihood of a word based on all possible word permutations in a sentence, rather than only those to the left or right of the target token. Furthermore, this approach intends to capture bidirectional context; as such, each position learns contextual data from all positions.

3.3. Evaluation Metrics

We employ common classification evaluation measures such as accuracy, precision, recall, and F1 score to evaluate the performance of the transformer language models. The formulae are shown as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{F1-score} = \frac{2TP}{2TP + FP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

where *TP* = true positive, *TN* = true negative, *FP* = false positive, and *FN* = false negative.

3.4. Criteria to Choose BERT, XLNet, and RoBERTa

The criteria used to choose BERT, XLNet, and RoBERTa as the models for comparison are based on their innovative architectures, performance enhancements, and state-of-the-art achievements in NLP:

1. Innovative architecture and techniques:
   - BERT: BERT was chosen for its bidirectional training mechanism, which allows it to understand the context within text from both directions. This innovation

significantly improves its performance in various NLP tasks by developing knowledge of the relationship between words in a sentence;

- XLNet: XLNet was selected because it addresses the limitations of BERT by using a permutation-based (predicts each word by considering all possible permutations of the words in a sentence) training objective. This method captures bidirectional context without the need for masked tokens. (These are used as proxies in training language models to hide specific words in a sentence. The model's task is to predict the hidden word using the surrounding context). This enhances the model's ability to utilise information in the text comprehensively. XLNet integrates autoregressive (AR) and autoencoding (AE) methods, addressing the disadvantages of BERT's masked language model [36];

- RoBERTa: RoBERTa was included due to its improvements over BERT, such as dynamic masking, increased training data, and longer training durations. These enhancements lead to superior performance in downstream tasks, making RoBERTa a robust model for comparison.

2. Performance and pre-training enhancements:

- BERT: the model's ability to understand the context and meaning of text through self-attention mechanisms makes it a strong baseline for NLP tasks;

- XLNet: by overcoming BERT's limitations with permutation language modelling, XLNet improves performance in understanding contextual information;

- RoBERTa: with dynamic masking and extensive training datasets, RoBERTa optimises BERT's approach, resulting in higher performance in NLP applications.

3. State-of-the-art (SOTA) achievements:

These models have achieved state-of-the-art (denotes the newest and most innovative technologies that currently outperform all others in terms of effectiveness, accuracy, or efficiency on specific benchmarks or tasks) performance on several NLP tasks [37,38]. Their continuous improvements and refinements make them suitable for comparison in the context of SA on YouTube data related to Lucid Motors and Tesla. These criteria highlight the selection of BERT, XLNet, and RoBERTa based on their advanced architectures, enhanced performance, and leading positions in the field of NLP.

4. Results

The results from BERT, XLNet, and RoBERTa using both datasets of Lucid and Tesla Motors YouTube comments before fine-tuning (precise adjustments to achieve the highest level of performance or effectiveness) are presented in Figures 1–6 below. The models underwent a single training cycle on a 60–40 data split for both Tesla and Lucid Motors comments for training and testing purposes, respectively. In Figure 1, the BERT model applied to the Lucid dataset shows that the highest proportion of the comments are positive, whilst a significant proportion are neutral. There are no comments classified as very positive, negative, or very negative.

Figure 2 presents the XLNet model results (without fine-tuning) using the Lucid dataset. The model reveals comments towards Lucid Motors are mostly neutral while there is a tiny proportion of negative and positive comments. In Figure 3, RoBERTa without fine-tuning classified all comments into the positive class. This result is surprising as the classification result deviates from the results from the other models.

Figures 4–6 show results regarding Tesla Motors. In Figure 4, the BERT model shows that comments towards Tesla Motors are mixed. This is because there is a high proportion of comments that are very negative and a significant proportion that are very positive.

Figure 5 presents the XLNet results (without fine-tuning) of comments towards the Tesla dataset. The XLNet results show that comments towards Tesla are mostly neutral or positive whilst other sentiment classifications are minimal.
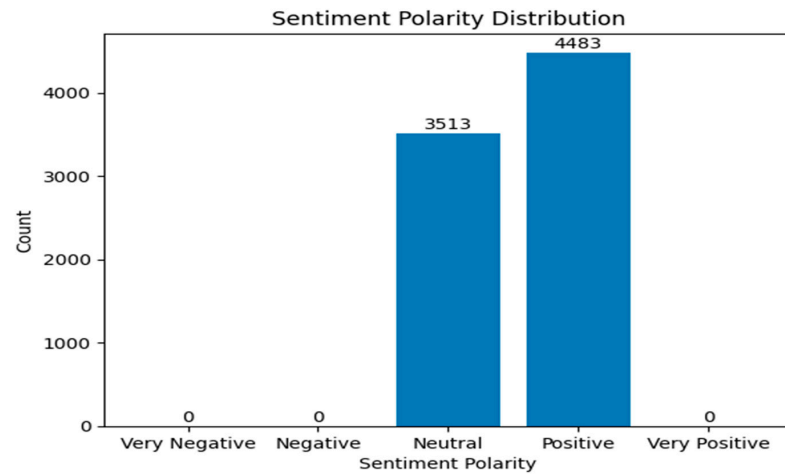
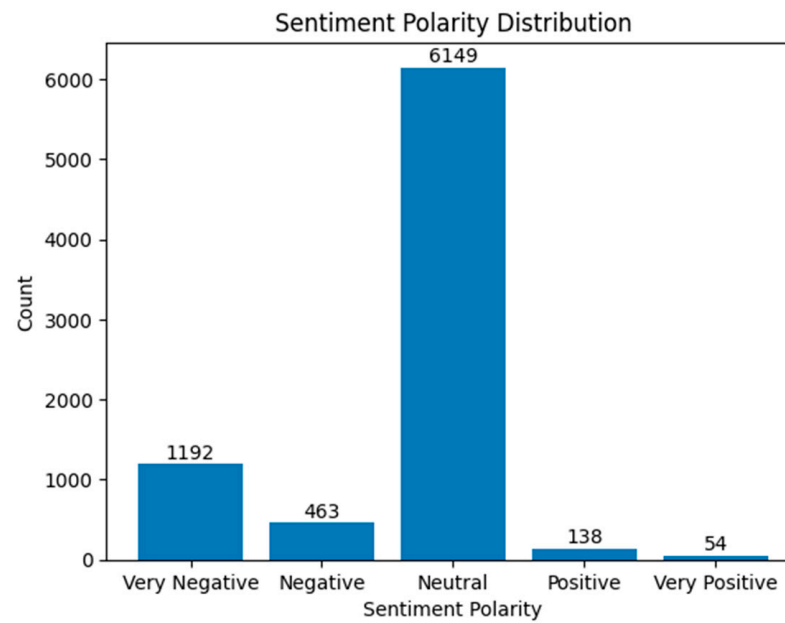**Figure 1.** SA results of BERT without fine-tuning (Lucid Motors).

Figure 2. SA results of XLNet without fine-tuning (Lucid Motors).
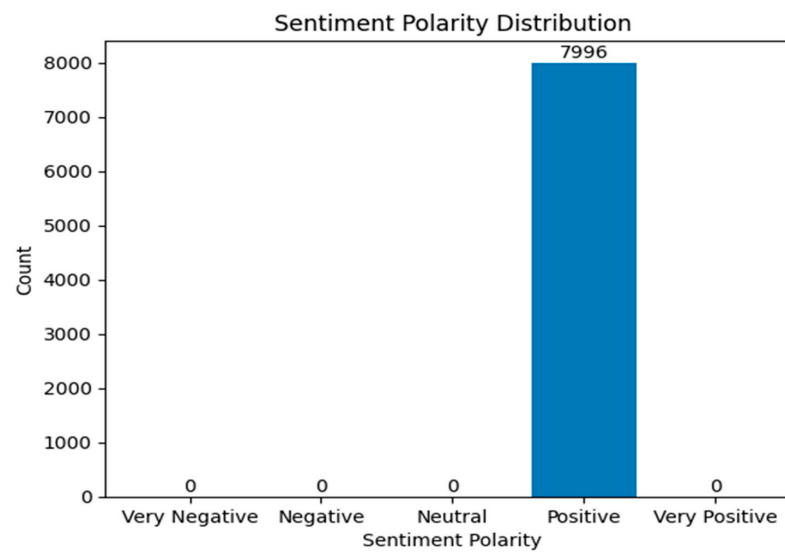

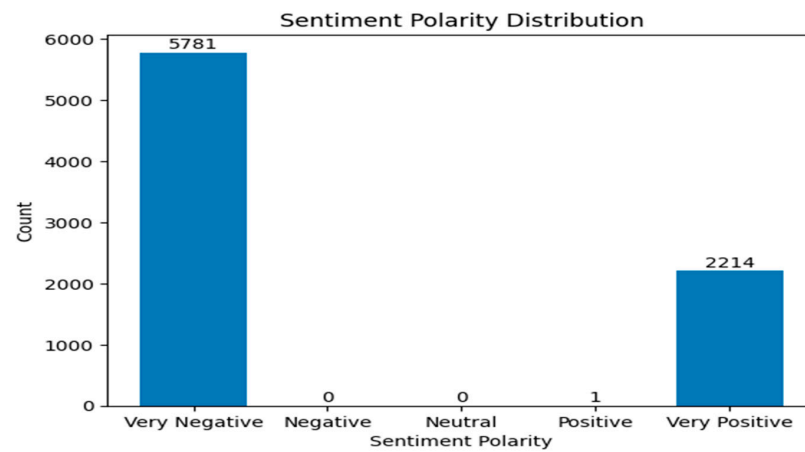**Figure 3.** SA results of RoBERTa without fine-tuning (Lucid Motors).

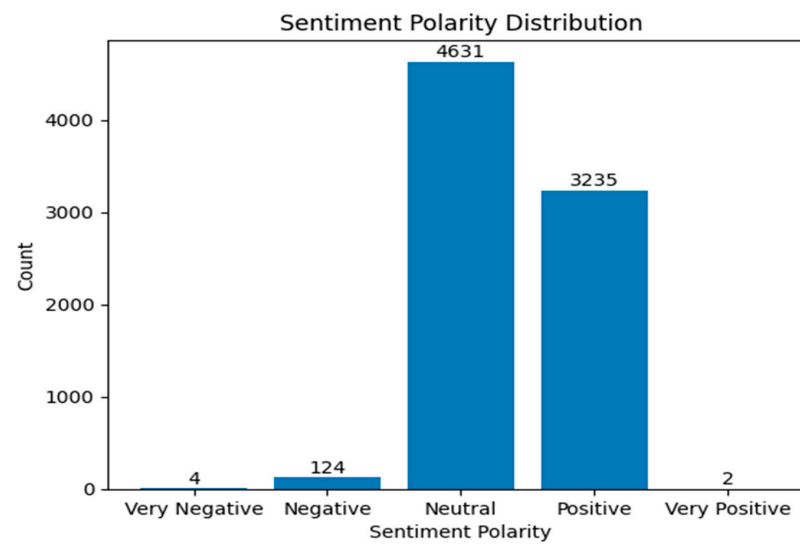**Figure 4.** SA results of BERT without fine-tuning (Tesla Motors).

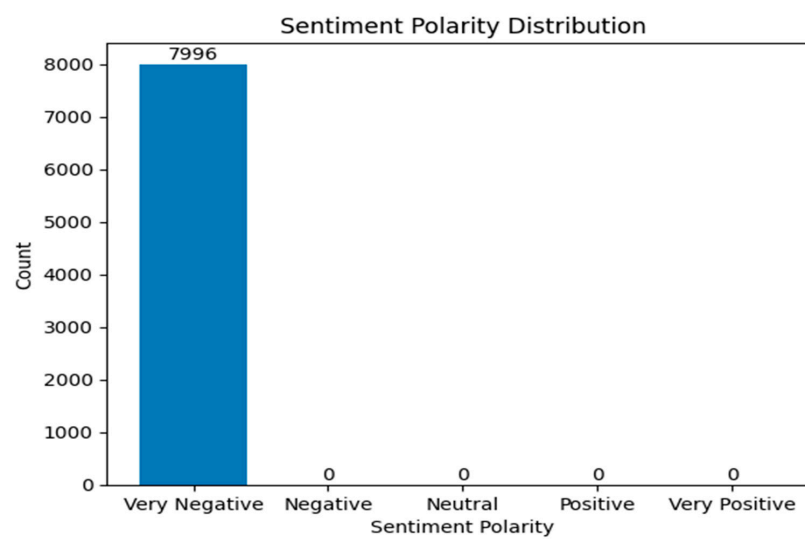**Figure 5.** SA results of XLNet without fine-tuning (Tesla Motors).

Figure 6. SA results of RoBERTa without fine-tuning (Tesla Motors).

Figure 6, given below, shows the results of the RoBERTa model using the Tesla dataset. It shows that all the sentiments are classified as very negative. This result is surprising. However, it will be worth comparing these results with those of the fine-tuned model.

In subsequent plots, this paper presents the results of the fine-tuned BERT, XLNet, and RoBERTa models applied to the Lucid dataset in Figures 7–9, respectively. As shown in Figure 7, the BERT model shows that the majority of the sentiments are classified as neutral while there is a noticeable proportion in other sentiment classes.
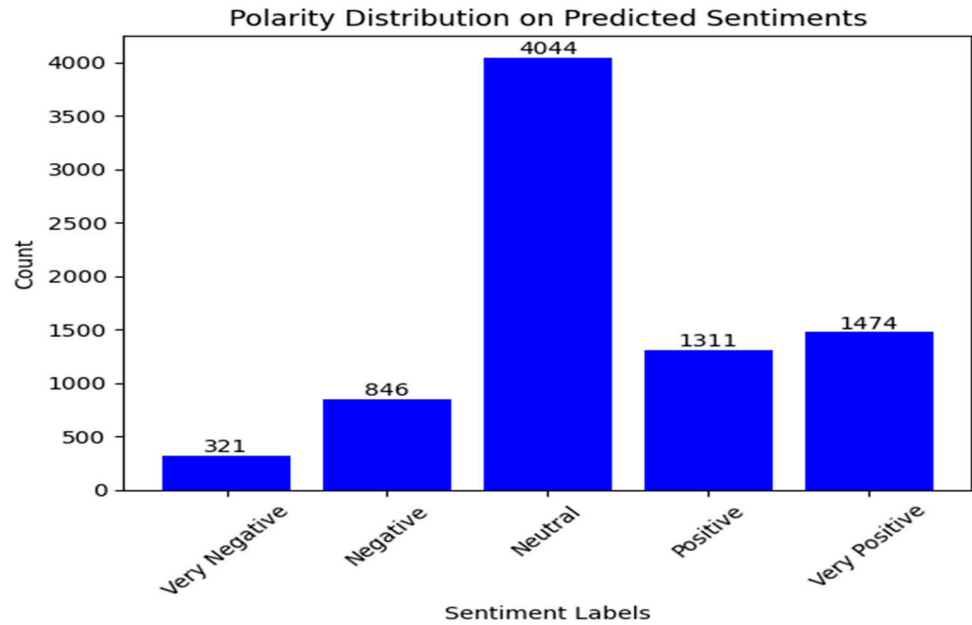

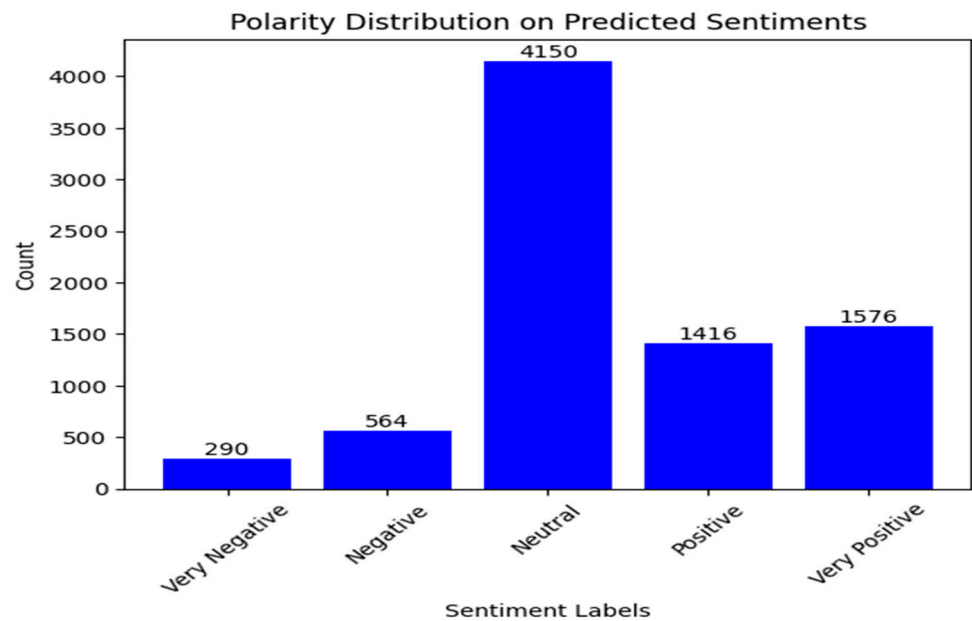**Figure 7.** SA results of BERT with fine-tuning (Lucid Motors).

Figure 8. SA results of XLNet with fine-tuning (Lucid Motors).

Similarly, XLNet results in Figure 8 show that the majority of the sentiments are classified as neutral while there is a noticeable proportion in other sentiment classes.

RoBERTa results in Figure 9 show that the majority of the sentiments are classified as neutral while there is a noticeable proportion in other sentiment classes.

In general, the models show similar results for the Lucid comments in terms of their sentiment classification. It is worth stating that most of the comments are neutral which indicates that users were objective rather than subjective. This might be a case where

comments might have been made requesting information or giving a review without being subjective. Furthermore, our analysis produced model sentiment classification results for Tesla comments (Figures 10–12). As shown in Figure 10, the fine-tuned BERT model shows that the majority of the comments are classified as neutral while there is a noticeable proportion in other sentiment classes.
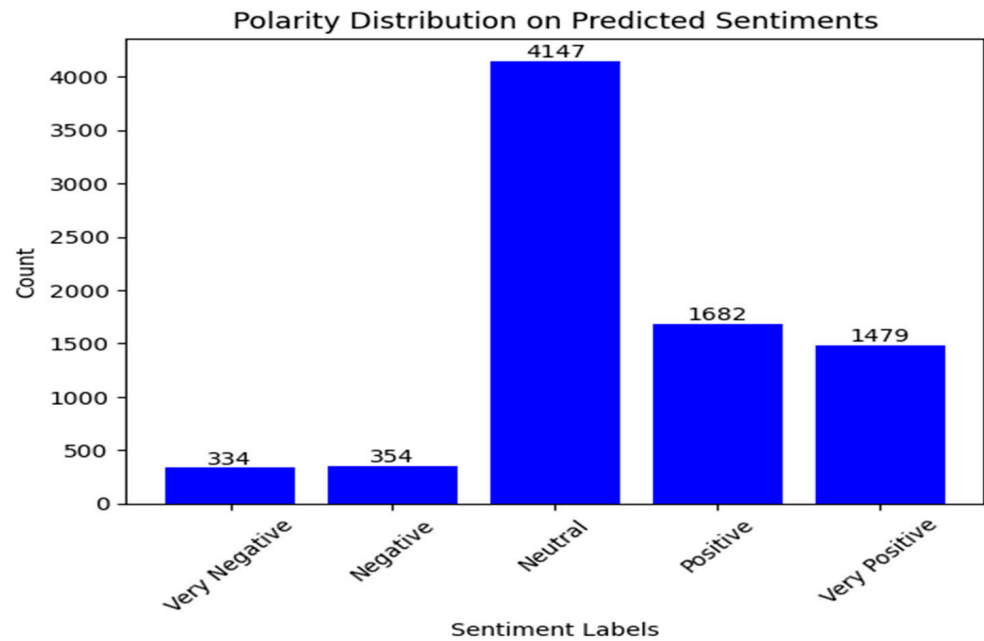

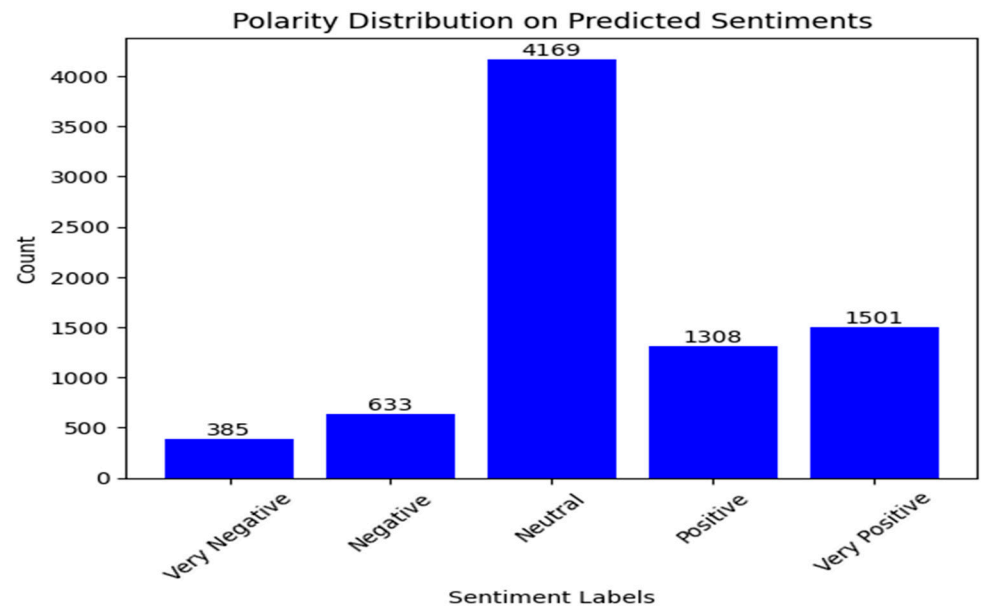**Figure 9.** SA results of RoBERTa with fine-tuning (Lucid Motors).

Figure 10. SA results of BERT with fine-tuning (Tesla Motors).

Similarly, plots in Figures 11 and 12 show that the fine-tuned XLNet and RoBERTa models show that the majority of the comments are classified as neutral while there is a noticeable proportion in other sentiment classes, especially the positive class. The sentiment classification results for the Tesla comments are similar to those of the Lucid comments. Thus, one of the main findings of our paper is that there is a high proportion of neutral comments made about EVs, while there is a higher proportion of positive comments compared to negative comments.
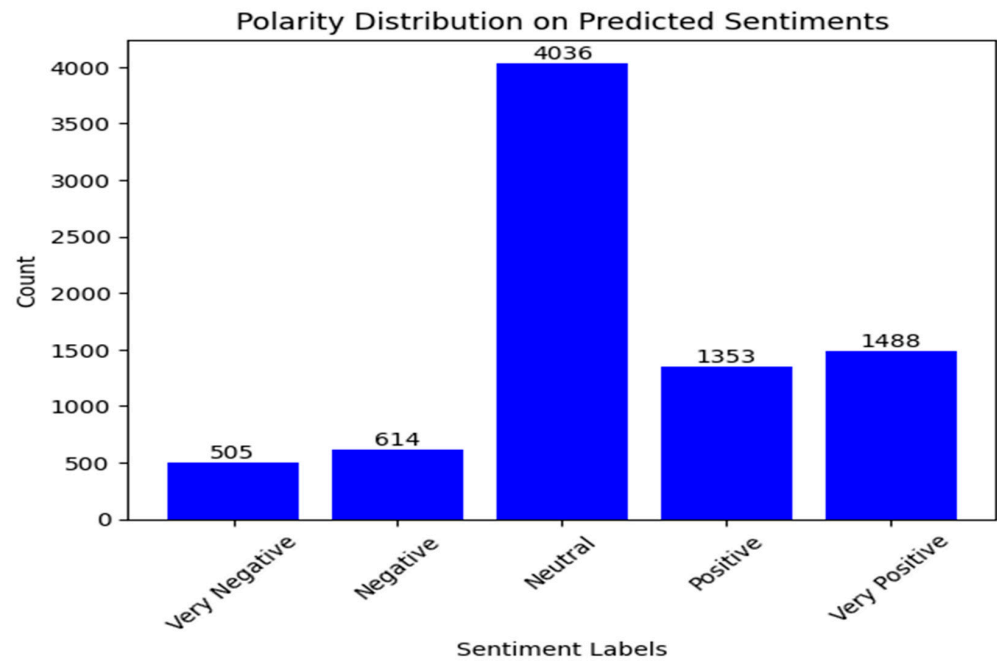
Polarity Distribution on Predicted Sentiments

Figure 11. SA results of XLNet with fine-tuning (Tesla Motors).

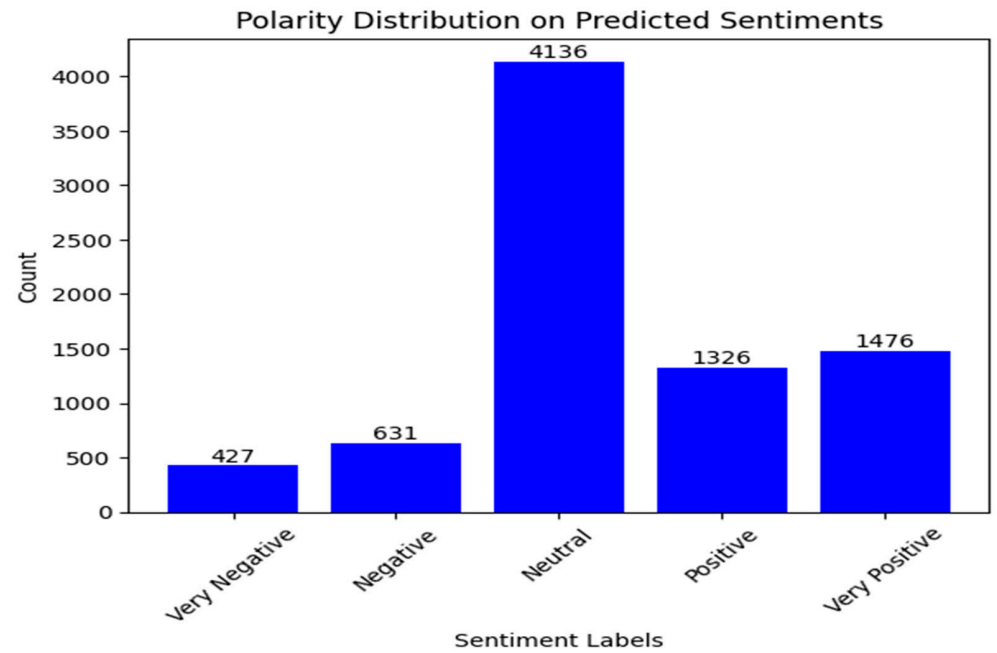**Polarity Distribution on Predicted Sentiments**

Figure 12. SA results of RoBERTa with fine-tuning (Tesla Motors).

To understand model performance, this paper presents in Tables 1 and 2 the evaluation results of the transformer models in terms of accuracy (A), precision (P), recall (R), and weighted F1 score (F). Moreover, it is worth stating that the models were split 60–40 for training and testing purposes, using three epochs, a learning rate of 2×10^{-5}, and a batch size of 3 for fair comparison. Table 1 presents the sentiment transformer model performance evaluation report for the Tesla dataset.

In Table 2 below, we present the sentiment transformer model performance evaluation report in terms of accuracy (A), precision (P), recall (R), and weighted F1 score (F) for the Lucid dataset.

In general, post fine-tuning, BERT and RoBERTa stand out as the best performers. For the Lucid Motors dataset, RoBERTa outperformed other variants of LLMs with an

F1 score of 92% and in the Tesla Motors dataset, both BERT and RoBERTa performed equally well after fine-tuning with an F1 score of at least 92%. The results showed that different model architectures and pre-training procedures showcase unique strengths and weaknesses depending on the dataset and domain. However, in both datasets RoBERTa showed consistent performance with a F1 score of at least 92%.

Table 1. Model evaluation before and after fine-tuning for the Tesla dataset.

		BERT		RoBERTa		XLNet	
		Without Fine-Tuning	Fine-Tuning	Without Fine-Tuning	Fine-Tuning	Without Fine-Tuning	Fine-Tuning
Tesla Motors	A	9.75%	**93.63%**	5.34%	92.12%	42.26%	90.10%
	P	3.89%	**93.77%**	0.29%	92.26%	43.19%	90.47%
	R	9.75%	**93.63%**	5.34%	92.10%	42.26%	90.10%
	F	4.94%	**93.63%**	0.54%	92.15%	37.10%	90.21%

Table 2. Model evaluation before and after fine-tuning for the Lucid dataset.

		BERT		RoBERTa		XLNet	
		Without Fine-Tuning	Fine-Tuning	Without Fine-Tuning	Fine-Tuning	Without Fine-Tuning	Fine-Tuning
Lucid Motors	A	37.06%	90.33%	17.30%	**92.33%**	43.88%	90.90%
	P	33.46%	91.85%	2.99%	**92.90%**	37.78%	91.01%
	R	37.06%	90.33%	17.30%	**92.31%**	43.88%	90.90%
	F	33.00%	90.76%	5.10%	**92.22%**	35.53%	90.92%

5. Conclusions

This study conducted SA on Lucid Motors and Tesla Motors-related YouTube data using BERT, XLNet, and RoBERTa pre-trained transformer models. Our findings showed that fine-tuning significantly improves model performance. Among the LLMs, fine-tuned RoBERTa achieved the highest accuracy of 92.12% with Tesla and 92.33% with Lucid EV datasets. Our findings show that a high proportion of public sentiment towards EVs (using results from Lucid and Tesla) is neutral. This indicates that people were more objective rather than subjective. This suggests that users made comments to seek information or give a review without being subjective. Furthermore, it is worth highlighting that our results suggest that there are considerably more positive comments than negative comments. This indicates that a higher proportion of users with subjective comments feel positive about the EVs.

RoBERTa showed consistent results in several NLP tasks, and our results are consistent with the findings of some other NLP results as evidenced in the studies of [33,34]. Theoretically, this underscores the need for case-specific model evaluation and highlights RoBERTa's effectiveness for SA. In conclusion, fine-tuned RoBERTa excelled in sentiment prediction, offering valuable market insights, particularly the prevalence of neutral and positive sentiment. This research emphasises the importance of fine-tuning LLMs for domain adaptation to achieve accurate sentiment classification and highlights RoBERTa's capabilities for YouTube comment SA. In summary, our approach can be deployed as a benchmark methodology in future studies. Our methodology contributes significantly by deploying a method which utilised a lexicon-based approach for data labelling. This is an alternative approach to the zero-shot text classification approach proposed in the study of Nugroho et al. [39]. Our approach is beneficial in practice as labelled datasets for these powerful machines are not readily available.

This paper presents SOTA results, however, limited by the data labelling technique utilised. In addition, we used small datasets (YouTube comments) due to limited resources. There is a potential bias in terms of social media data collected (YouTube). It is worth applying our methods to other social media datasets like Twitter in future work. In the future, we aim to adopt human annotation by experts to improve on data quality. Furthermore, we aim to compare other transformer models using diverse datasets from various EV companies.

References

1. Wolinetz, M.; Axsen, J. How policy can build the plug-in electric vehicle market: Insights from the REspondent-based Preference And Constraints (REPAC) model. *Technol. Forecast. Soc. Chang.* **2017**, *117*, 238–250. [CrossRef]
2. Mateen, S.; Amir, M.; Haque, A.; Bakhsh, F.I. Ultra-fast charging of electric vehicles: A review of power electronics converter, grid stability and optimal battery consideration in multi-energy systems. *Sustain. Energy Grids Netw.* **2023**, *35*, 101112. [CrossRef]
3. International Energy Agency (IEA). Global EV Outlook. 2023. Available online: https://www.iea.org/reports/global-ev-outlook-2023 (accessed on 21 December 2023).
4. Qin, Q.; Zhou, Z.; Zhou, J.; Huang, Z.; Zeng, X.; Fan, B. Sentiment and attention of the Chinese public toward electric vehicles: A big data analytics approach. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107216. [CrossRef]
5. Su, C.W.; Yuan, X.; Tao, R.; Umar, M. Can new energy vehicles help to achieve carbon neutrality targets? *J. Environ. Manag.* **2021**, *297*, 113348. [CrossRef] [PubMed]
6. Zhao, X.; Ma, Y.; Shao, S.; Ma, T. What determines consumers' acceptance of electric vehicles: A survey in Shanghai, China. *Energy Econ.* **2022**, *108*, 105805. [CrossRef]
7. Hayashida, S.; La Croix, S.; Coffman, M. Understanding changes in electric vehicle policies in the US states, 2010–2018. *Transp. Policy* **2021**, *103*, 211–223. [CrossRef]
8. Morton, C.; Anable, J.; Nelson, J.D. Exploring consumer preferences towards electric vehicles: The influence of consumer innovativeness. *Res. Transp. Bus. Manag.* **2016**, *18*, 18–28. [CrossRef]
9. Carley, S.; Krause, R.M.; Lane, B.W.; Graham, J.D. Intent to purchase a plug-in electric vehicle: A survey of early impressions in large US cites. *Transp. Res. Part D Transp. Environ.* **2013**, *18*, 39–45. [CrossRef]
10. Ogunleye, B.O. Statistical Learning Approaches to Sentiment Analysis in the Nigerian Banking Context. Ph.D. Thesis, Sheffield Hallam University, Sheffield, UK, 2021.
11. Ogunleye, B.; Brunsdon, T.; Maswera, T.; Hirsch, L.; Gaudoin, J. Using Opinionated-Objective Terms to Improve Lexicon-Based Sentiment Analysis. In Proceedings of the 12th International Conference on Soft Computing for Problem-Solving (SocProS 2023), Roorkee, India, 11–13 August 2023; Lecture Notes in Networks and Systems; Springer Nature: Singapore, 2023; Volume 995, pp. 1–23. [CrossRef]
12. Chen, C.C.; Chang, Y.C. What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness. *Telemat. Inform.* **2018**, *35*, 1512–1523. [CrossRef]
13. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
14. Kumawat, S.; Yadav, I.; Pahal, N.; Goel, D. Sentiment analysis using language models: A study. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 984–988.
15. Zhang, B.; Yang, H.; Zhou, T.; Ali Babar, M.; Liu, X.Y. Enhancing financial sentiment analysis via retrieval augmented large language models. In Proceedings of the Fourth ACM International Conference on AI in Finance, Brooklyn, NY, USA, 27–29 November 2023; pp. 349–356.

16.   Ashari, N.; Al Firdaus, M.Z.M.; Budi, I.; Santoso, A.B.; Putra, P.K. Analyzing Public Opinion on Electrical Vehicles in Indonesia Using Sentiment Analysis and Topic Modeling. In Proceedings of the 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), Jakarta, Indonesia, 16 February 2023; pp. 461–465.

17.   Fu, X.; Wei, Y.; Xu, F.; Wang, T.; Lu, Y.; Li, J.; Huang, J.Z. Semi-supervised Aspect-level Sentiment Classification Model based on Variational Autoencoder. *Knowl. Based Syst.* **2019**, *171*, 81–92. [CrossRef]

18.   John, J.M.; Shobayo, O.; Ogunleye, B. An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics* **2023**, *2*, 809–823. [CrossRef]

19.   Iparraguirre-Villanueva, O.; Guevara-Ponce, V.; Sierra-Liñan, F.; Beltozar-Clemente, S.; Cabanillas-Carbonel, M. Sentiment analysis of tweets using unsupervised learning techniques and the k-means algorithm. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 571–578. [CrossRef]

20.   Jena, R. An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. *Ind. Mark. Manag.* **2020**, *90*, 605–616. [CrossRef]

21.   Ha, S.; Marchetto, D.J.; Dharur, S.; Asensio, O.I. Topic classification of elec-tric vehicle consumer experiences with Transformer-based deep learning. *Patterns* **2021**, *2*, 100195. [CrossRef]

22.   Biswas, S.; Young, K.; Griffith, J. Automatic Sentiment Labelling of Multimodal Data. In Proceedings of the International Conference on Data Management Technologies and Applications, Virtual, 6–8 July 2021; Springer Nature: Cham, Switzerland, 2021; pp. 154–175.

23.   Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassanien, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Appl. Soft Comput.* **2020**, *97*, 106754. [CrossRef]

24.   Saad, E.; Din, S.; Jamil, R.; Rustam, F.; Mehmood, A.; Ashraf, I.; Choi, G.S. Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums. *IEEE Access* **2021**, *9*, 85721–85737. [CrossRef]

25.   Hasan, A.; Moin, S.; Karim, A.; Shamshirband, S. Machine learning-based sentiment analysis for twitter accounts. *Math. Comput. Appl.* **2018**, *23*, 11. [CrossRef]

26.   Hasan, K.A.; Shovon, S.D.; Joy, N.H.; Islam, M.S. Automatic labeling of twitter data for developing COVID-19 sentiment dataset. In Proceedings of the 2021 5th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 17–19 December 2021; pp. 1–6.

27.   Ogunleye, B.; Sharma, H.; Shobayo, O. Sentiment Informed Sentence BERT-Ensemble Algorithm for Depression Detection. *Big Data Cogn. Comput.* **2024**, *8*, 112. [CrossRef]

28.   Ruan, T.; Lv, Q. Public perception of electric vehicles on Reddit and Twitter: A cross-platform analysis. *Transp. Res. Interdiscip. Perspect.* **2023**, *21*, 100872. [CrossRef]

29.   JustAnotherArchivist. *Justanotherarchivist/Snscrape: A Social Networking Service Scraper in Python*; GitHub: San Francisco, CA, USA, 2020. Available online: https://github.com/JustAnotherArchivist/snscrape (accessed on 25 August 2024).

30.   Grzegorzewski, P.; Kochanski, A. Data Preprocessing in Industrial Manufacturing. In *Soft Modeling in Industrial Manufacturing, Studies in Systems, Decision and Control*; Springer Nature: Cham, Switzerland, 2019; Volume 183, pp. 27–41. [CrossRef]

31.   Qorib, M.; Oladunni, T.; Denis, M.; Ososanya, E.; Cotae, P. COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccina-tion Twitter dataset. *Expert Syst. Appl.* **2023**, *212*, 118715. [CrossRef] [PubMed]

32.   Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Stoyanov, V. RoBERTa: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.

33.   Ogunleye, B.; Dharmaraj, B. The Use of a Large Language Model for Cyberbullying Detection. *Analytics* **2023**, *2*, 694–707. [CrossRef]

34.   Bozanta, A.; Angco, S.; Cevik, M.; Basar, A. Sentiment analysis of stocktwits using transformer models. In Proceedings of the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA, 13–16 December 2021; pp. 1253–1258.

35.   Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. XLNet: Generalized autoregressive pretraining for language understanding. *arXiv* **2019**, arXiv:1906.08237.

36.   Dey, L.; Chakraborty, S.; Biswas, A.; Bose, B.; Tiwari, S. Sentiment analysis of review datasets using naïve Bayes' and k-nn classifier. *Int. J. Infor-Mation Eng. Electron. Bus.* **2016**, *8*, 54–62. [CrossRef]

37.   Ye, J.; Zhou, J.; Tian, J.; Wang, R.; Zhou, J.; Gui, T.; Zhang, Q.; Huang, X. Sentiment-aware multimodal pre-training for multimodal sentiment analysis. *Knowl. Based Syst.* **2022**, *258*, 110021. [CrossRef]

38.   Chennafi, M.E.; Bedlaoui, H.; Dahou, A.; Al-qaness, M.A. Arabic aspect-based sentiment classification using Seq2Seq dialect normalization and transformers. *Knowledge* **2022**, *2*, 388–401. [CrossRef]

39.   Nugroho, S.A.; Widianto, S. Exploring Electric Vehicle Adoption in Indonesia Using Zero-Shot Aspect-Based Sentiment Analysis. *Sustain. Oper. Comput.* **2024**, *5*, 191–205. [CrossRef]