

Article

Not peer-reviewed version

Electric Vehicle Sentiment Analysis Using Large Language Models

[Hemlata Sharma](#) , Faiz Ud Din , [Bayode Ogunleye](#) *

Posted Date: 9 August 2024

doi: 10.20944/preprints202408.0723.v1

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



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

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, United Kingdom; h.sharma@shu.ac.uk; faizfakhar@outlook.com

² Department of Computing & Mathematics, University of Brighton, Brighton BN2 4GJ, United Kingdom

* Correspondence: author email: 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, positive or negative opinion or attitude towards electric vehicle (EV) brands. This provides companies with valuable insight about the public's opinion of their products and brands. In the field of natural language processing (NLP), transformer models have shown great performances over the 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 [1]. In this study, we present a comparative study of large language models (LLMs) including bidirectional encoder representations from transformers (BERT), robustly optimized BERT approach and generalized autoregressive pretraining for language understanding using Lucid motors and Tesla motors YouTube datasets. Results evidenced LLMs like BERT and her variants are off-the-shelf algorithms for sentiment analysis, specifically, when fine-tuned. Furthermore, our findings presents the need for domain adaptation whilst utilizing LLMs. Finally, the experimental results showed that RoBERTa achieved consistent performance across the EV datasets with a F1 score of at least 92%.

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

1. Introduction

The electric vehicles (EV) are pivotal to attaining the zero-emission target set for 2050 to improve the environmental challenges. Since the start of the first mass-market of 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 showed 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 EV were sold in the UK in 2022 which evidences the growth in the EV market. Several countries have implemented policies to encourage the adoption of EV. For example, the UK government constructed several electric charging points across the country. Similarly, China implemented financial subsidiary for EV purchase [4]. Despite the potential benefits of adopting the EV cars, the society at large seems doubtful about fully adopting the cars. Previous studies identified the factors influencing the consumers' adoption of EV as limited awareness [5], battery life [6], and national policy [7]. Authors in [8] found out that attitude to innovation and functional performance are key elements of EV adoption. Furthermore, the findings of Carley et al. [9] showed that people with high level of education showed favourable intention to purchase plug-in EV but individual point of view or understanding of the advantages and disadvantages of EV is a significant factor. However, it is worth noting that majority of the studies utilised survey questionnaire to investigate peoples' attitude towards EV. Unfortunately, questionnaires are limited to pre-defined variables [10]. Another channel to mine peoples' perspective is by performing sentiment analysis 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 available of big (text) data. SA provides companies the opportunity to understand their business better by extracting insights from customers' data. Authors in [12] stated consumers consider other peoples' review for purchasing purpose. This evidences the impact of reviews, comments and information shared on a product publicly. In the automotive industry, newly emerged company 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, the LLMs have taken the world by storm with their performances. Previous studies have proposed the use of LLMs for SA [14,15]. However, there is dearth of study on SA in the EV context, specifically with the application of sophisticated approaches like the LLMs. Thus, this study aims to address this gap. To this end, this paper aims to perform SA to assess consumers' perception of Lucid and Tesla EV and thus, evaluate the effectiveness of the LLMs. To conclude, this study summarises the main contributions as follows.

- This paper demonstrates the need to fine-tune LLMs for domain adaptation and thus propose the use of a fine-tuned RoBERTa algorithm for EV sentiment prediction.
- Our paper demonstrate a SA approach that took advantage of language understanding of the transformer models to complement the 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 results.

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

2. Related Work

SA provides an analytical technique to understand customers' perspective. 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, lexicon based approach and machine learning (ML) approach [11]. The former approach relies on bag of words and set of rules to classify document. Whilst, the latter approach can be further divided into three main categories namely, supervised ML approach, unsupervised ML approach and 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 purpose and the remainder for testing purpose. For example, the study of [16] used 80% of 2847 Indonesia labelled tweets for training and 20% for testing. They showed support vector machine (SVM) outperformed convolutional neural network (CNN), logistic regression, random forest, gradient boosting, and recurrent neural network (RNN) with an accuracy of 75.08% and F1 score of 78%. In the case of semi-supervised ML approach, this technique is suitable when only a limited labelled dataset is available for training purpose. An example of the approach was demonstrated in the study of [17] that used variational autoencoder for aspect-based SA. Whilst, the unsupervised ML approach does not require labelled dataset. The approach focuses on uncovering hidden patterns and make prediction from unlabeled dataset [18]. In the context of SA, authors in [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 used supervised ML approach. For example, authors in [20] showed CNN outperformed SVM, Doc2Vec algorithm, and RNN in a consumers' sentiment classification task in the indian EV market. In China, [4] used weibo dataset to mine consumers' sentiment towards EV. They showed bidirectional long short-term memory (Bi-LSTM) with attention layer outperformed SVM, CNN, Bi-LSTM, and a combination of CNN and LSTM with a F1-score of 86%. Their findings

showed growth rate toward EV varies across regions and in terms of gender, men pay more attention to EV. The problem of utilising supervised approach is the fact that ML algorithms require large amount of data for training purpose. Unfortunately, the large, labelled set are not readily available in most cases. However, there are various ways to data labelling for sentiment classification task. Firstly, is the human labelling, where experts are recruited for the purpose of providing sentiment labels to data. This process yields high quality labelled dataset and also regarded as the ground truth labels. For example, authors in [21] employed expert annotated data and compared BERT, XLNet, LSTM and CNN and thus, showed BERT achieved the best result with a 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 to 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 is prone to error, biased and are of low quality labels. The approach is unreliable because availability of labellers is not certain. Based on these challenges, past studies used lexicons for data labelling. For example, [23] compared Afinn, Textblob and Vader for assigning sentiment labels. In their experiment, they showed Afinn sentiment labels are close to human labels. Similarly, Vader [24,25], TextBlob [25,26] and Afinn [25,27], and have been used for assigning sentiment labels to dataset. Authors in [28] constructed two datasets from Reddit and Twitter between January 2011 and December 2020 to understand public's perception on EV. Thus, they used the VADER (Valence Aware Dictionary and sEntiment Reasoner), LIWC (Linguistic Inquiry and Word Count) and Afinn lexicons to generate insight 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 the LLMs have not been fully deployed in this context. More specifically, existing literature have not utilised transformer models for targeted YouTube SA regarding EV corporations (such as Lucid Motors & Tesla Motors). It remains unclear which architecture would perform best in this niche domain. A domain-specific benchmarking of transformer capabilities on automotive brand YouTube comments would provide both industry and academic value, most especially for research benchmarking. This study aims to address this gap and also generate actionable insights on consumer perceptions regarding the new automotive brand.

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 utilized 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, categorizing 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 pre-processing 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 SA task. The most popular libraries are Afinn, TextBlob and Vader. These lexicon approaches have been praised for their performances across several domain 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 SVM fitted with Textblob labels achieved the best performance. In this study, we consider TextBlob for data labelling. The dictionary consists of 2,918 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 evaluation of the data labelling, we randomly select 500 samples and compare the labels against ground truth labels (done by 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 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 the Masked Language Model (MLM) and next-sentence prediction tasks. The base version consists of 12 layers of transformer blocks, 768 hidden layer size, and 12 self-attention heads. While the large version consists of 24 layers of transformer blocks, 1024 hidden layer size, 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 tokenization 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 are 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, calculates value (vectors) using the dot product.
- Normalization Layer - the SoftMax function helps normalize the vectors.
- Classification Head – Converting sequential output into classification results and the SoftMax function helps normalize class scores into probability values.
- Training loss – Measuring the difference between predicted probabilities and true labels, often using loss functions like cross-entropy.
- Optimization – updating model parameters to minimize loss using the Adam algorithm (backpropagation).

3.2.2. Robustly Optimized BERT Approach (RoBERTa)

Facebook AI Research (FAIR) proposed Robustly optimized BERT approach (RoBERTa) in 2019. Authors in [32] criticized BERT as undertrained, and thus, modified the training method by (i) using a dynamic masking pattern instead of static, (ii) training with more data with large batches, (iii) removing next sentence prediction, and (iv) training on longer sentences and proposed 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 task [33,34].

3.2.3. XLNet

BERT was criticized that it ignores the dependency between the masked positions and leads to pretrain-finetune discrepancy because it relies on masking the input to corrupt it. Thus, authors in [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. XLNet training objective calculates 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, the 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 is shown as follows.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{F1-score} = \frac{2TP}{2TP+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

Where TP = True positive, TN = True Negative, FP = False Positive and FN = False Negative.

3.3. 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 training objective. This method captures bidirectional context without the need for masked tokens, enhancing the model's ability to utilize information in the text comprehensively. XLNet integrates autoregressive (AR) and autoencoding (AE) methods, solving 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 Pretraining 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 modeling, XLNet improves performance in understanding contextual information.
 - RoBERTa: With dynamic masking and extensive training datasets, RoBERTa optimizes BERT's approach, resulting in higher performance in NLP applications.
3. State-of-the-Art Achievements:

These models have achieved state-of-the-art 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. Result

The results from BERT, XLNet and RoBERTa using both datasets of Lucid and Tesla motors Youtube comments before fine-tuning are presented in Figures 1, 2, 3, 4, 5, and 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 purpose respectively. From Figure 1, the BERT model applied to the Lucid dataset shows that a higher proportion of the comments are positive, whilst, a significant proportion are neutral. There are no comments classified as very positive, negative or very negative.

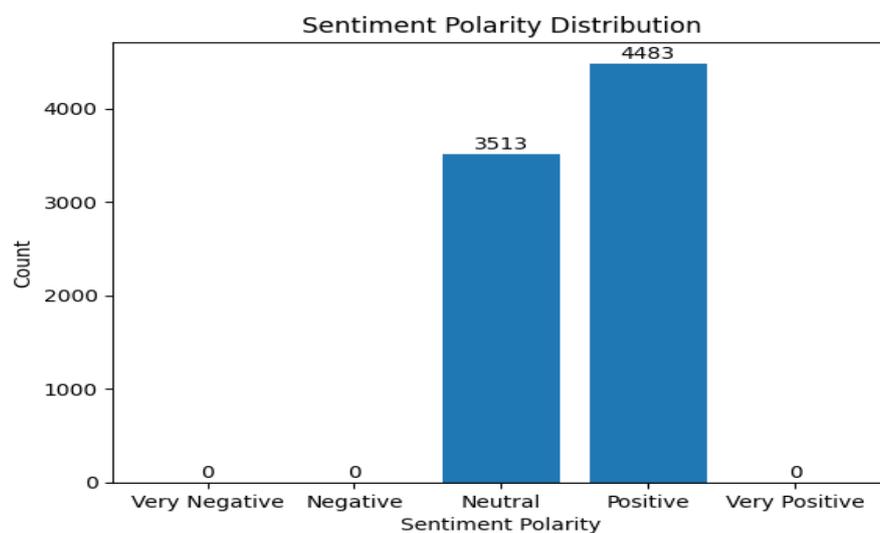


Figure 1. SA result of BERT without fine-tuning (Lucid motors).

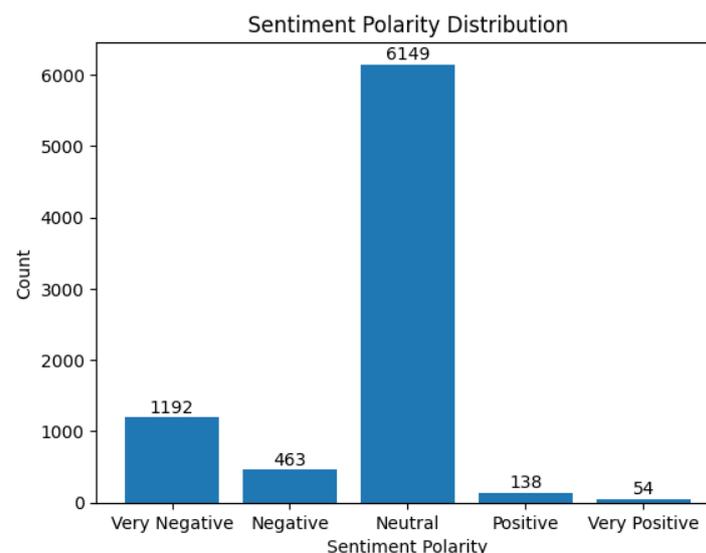


Figure 2. SA result of XLNet without fine-tuning (Lucid motors).

Figure 2 presents the XLNet model without fine-tuning using 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 to results from the other models.

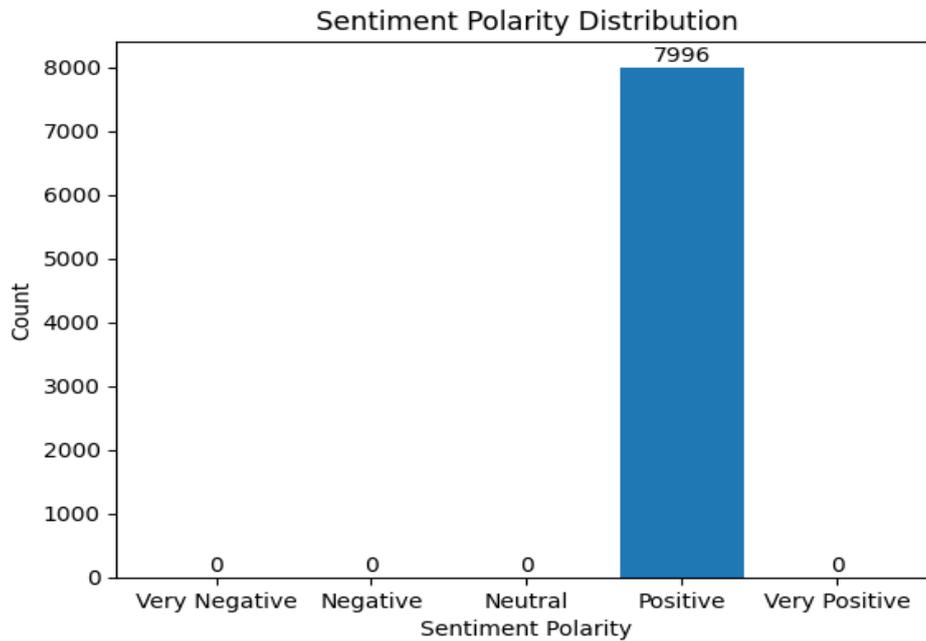


Figure 3. SA result of RoBERTa without fine-tuning (Lucid motors).

Figure 4, 5 and 6 show results with Tesla motors. In figure 4, the BERT model show that comments towards the Tesla motors are mixed. This is because there is a high proportion of comments that are very negative and a significant proportion is very positive.

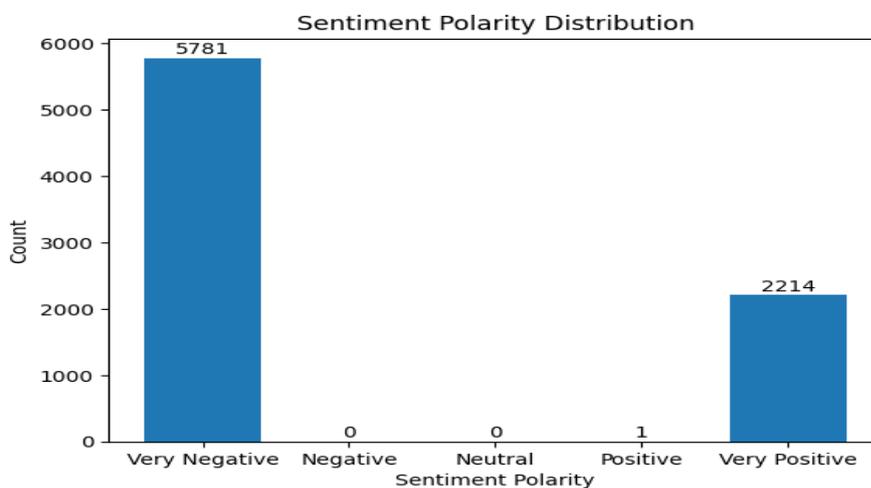


Figure 4. SA result of BERT without fine-tuning (Tesla motors).

Figure 5 presents the XLNet result (without fine-tuning) of comments towards the Tesla dataset. The XLNet result shows that comments towards Tesla are mostly neutral and positive whilst other sentiment classification are minimal.

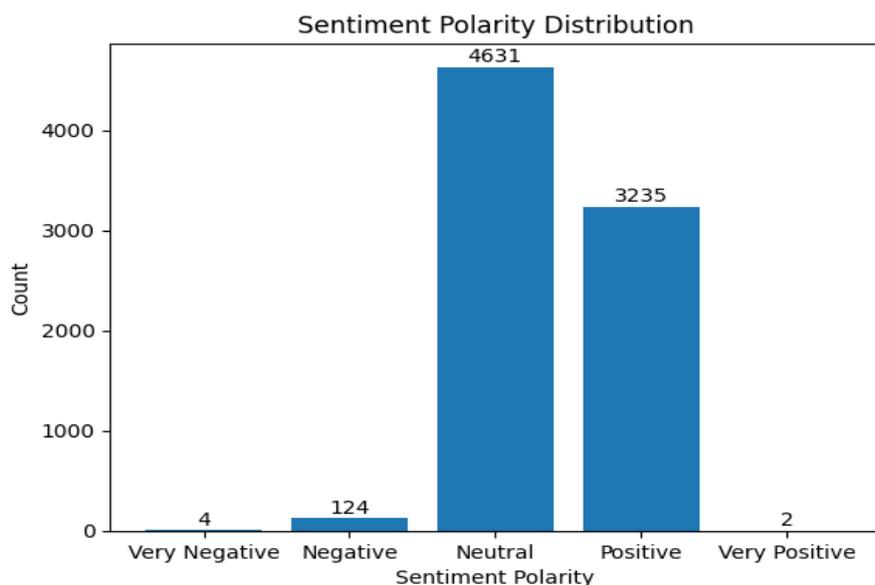


Figure 5. SA result of XLNet without fine-tuning (Tesla motors).

Figure 6 given below shows RoBERTa model using 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.

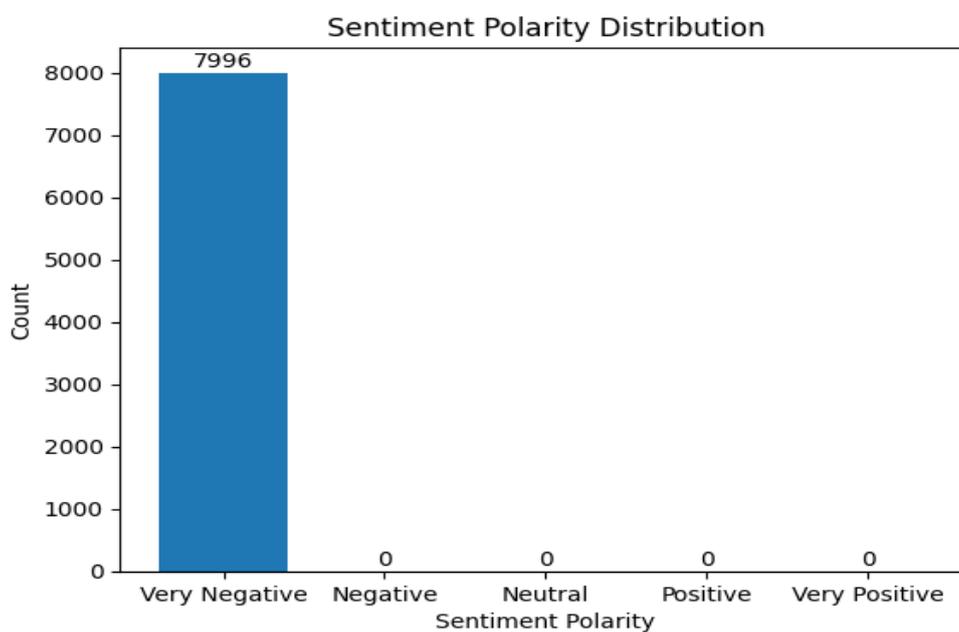


Figure 6. SA result of RoBERTa without fine-tuning (Tesla motors).

In subsequent plots, this paper presents results of the fine tuned BERT, XLNet and RoBERTa models applied to the lucid dataset in Figure 7, 8, and 9 respectively. As shown in Figure 7, the BERT model shows that majority of the sentiments are classified as neutral while there is a noticeable proportion in other sentiment classes.

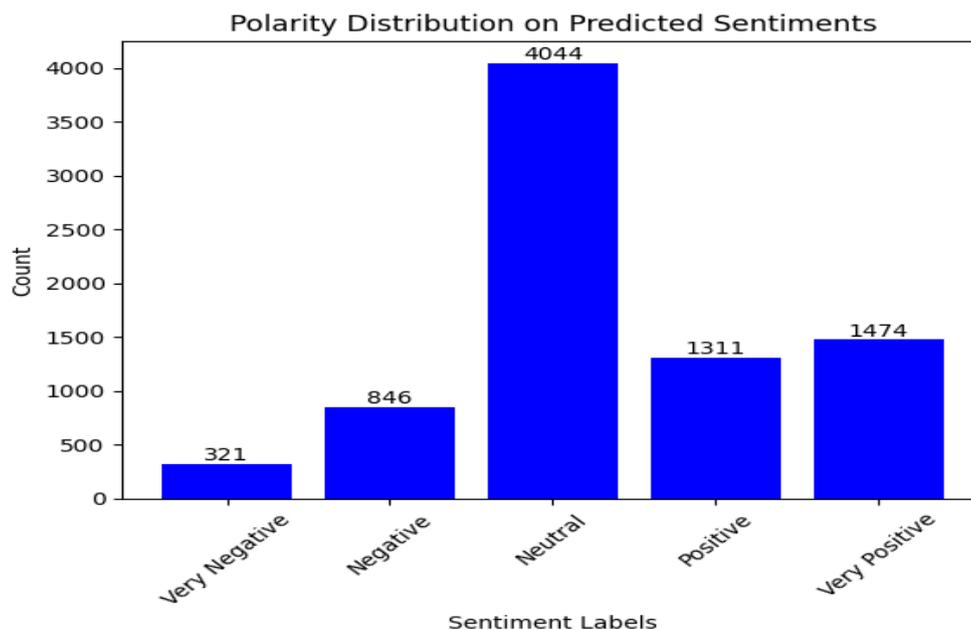


Figure 7. SA result of BERT with fine-tuning (Lucid motors).

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

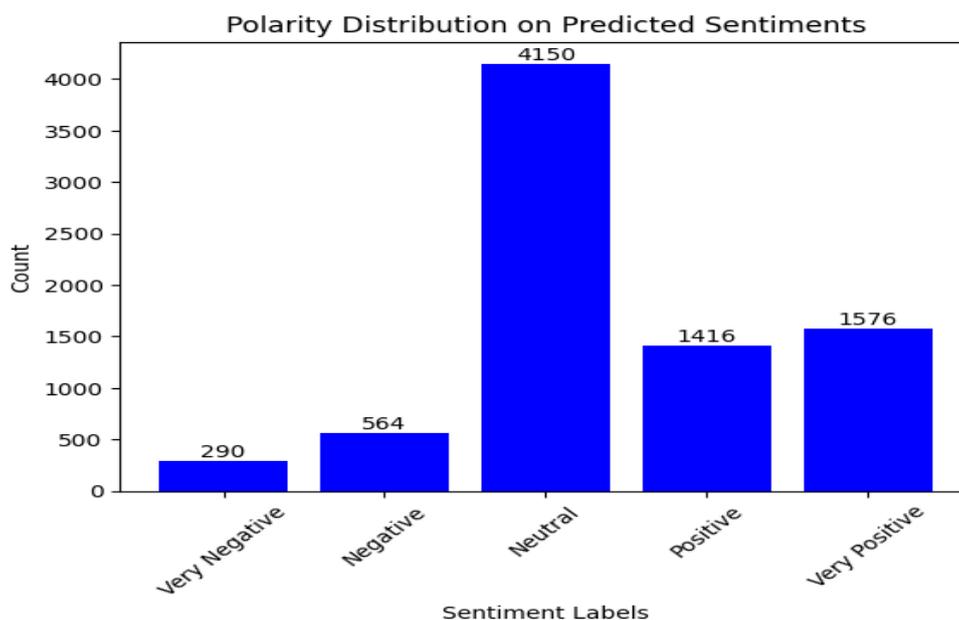


Figure 8. SA result of XLNet with fine-tuning (Lucid motors).

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

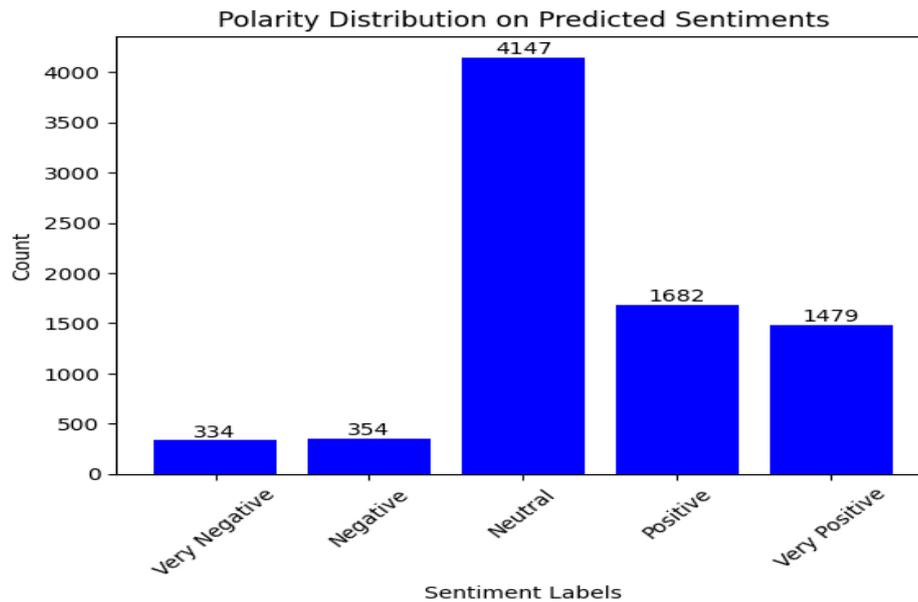


Figure 9. SA result of RoBERTa with fine-tuning (Lucid motors).

In general, the models show similar results with 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 towards requesting for information or passing a review without being subjective. Furthermore, our analysis produces the model sentiment classification results in Figure 10, 11, 12 with Tesla comments. As shown in Figure 10, the fine tuned BERT model shows that majority of the comments are classified as neutral while there is a noticeable proportion in other sentiment classes.

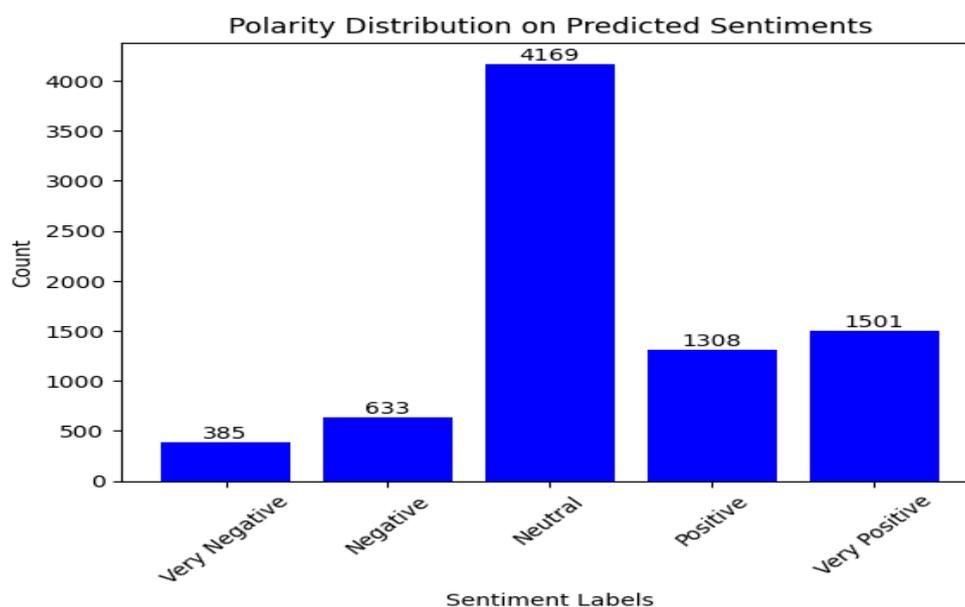


Figure 10. SA result of BERT with fine-tuning (Tesla motors).

Similarly, plots in Figure 11 and 12 shows that the fine tuned XLNet and RoBERTa models show that 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 that 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 EV. Whilst there is a higher proportion of positive comments compared to negative comments.

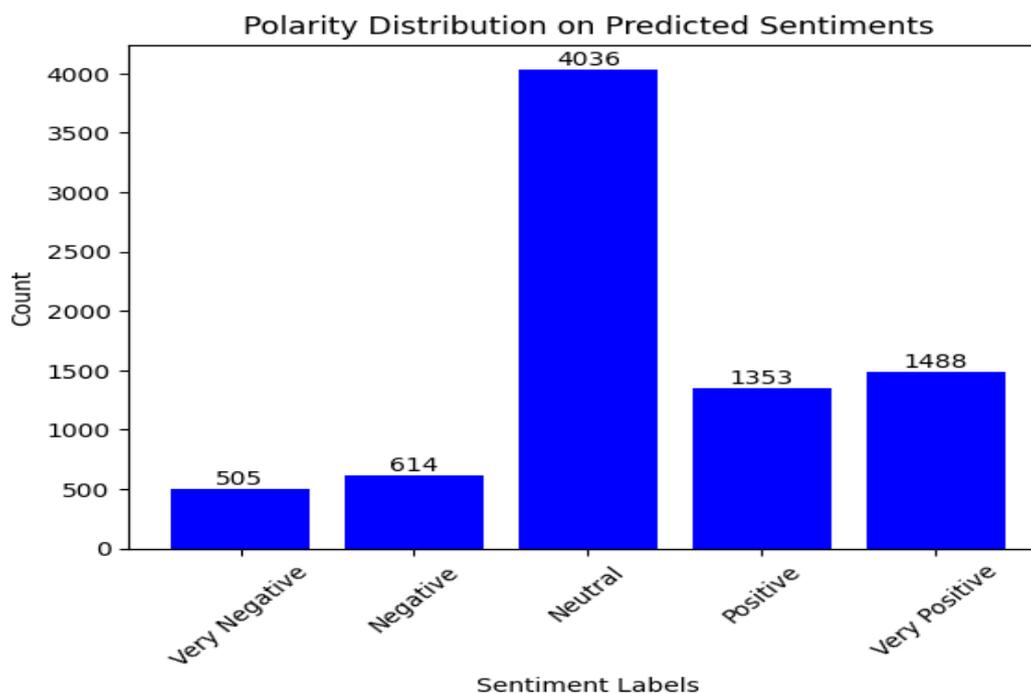


Figure 11. SA result of XLNet with fine-tuning (Tesla motors).

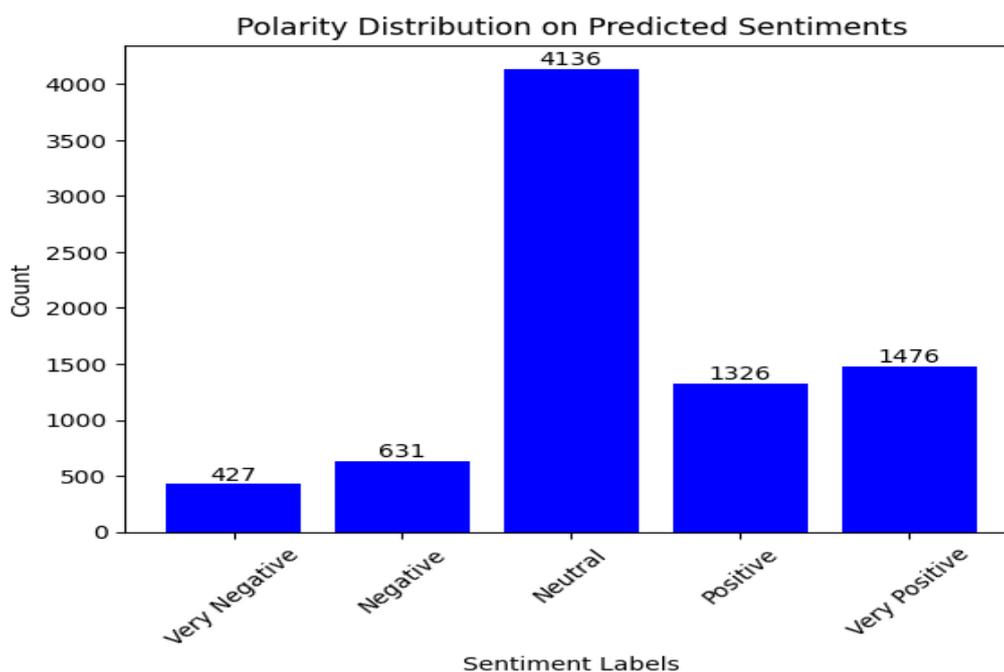


Figure 12. SA result of RoBERTa with fine-tuning (Tesla motors).

To understand the model performance, this paper presents in Table 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 splitted into 60 - 40 for training and testing purpose, using three epochs, a learning rate of $2e-5$, and a batch size of 3 for fair comparison. Table 1 presents the sentiment transformer model performance evaluation report for Tesla Dataset.

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

Model	BERT	RoBERTa	XLNet
-------	------	---------	-------

		Without Fine-tuning	Fine-tuning	Without Fine-tuning	Fine-tuning	Without Fine-tuning	Fine-tuning
	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%

Subsequently, 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 Lucid Dataset.

Table 2. Model evaluation before and after fine-tuning for 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%

In general, post fine-tuning, BERT and RoBERTa stands 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%.

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 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 dataset. Our findings show a high proportion of public sentiment towards EV (using results from Lucid and Tesla) are neutral. This indicates that people were more objective rather than subjective. This suggests that users made comments to seek information or pass a review without being subjective. Furthermore, it is worth highlighting that our results suggest that there are considerable positive comments than negative comments. This indicates that a higher proportion of users with subjective comments feel positive about the EV.

RoBERTa showed consistent results in several NLP tasks, as our result is consistent with the findings of some other NLP results as evidenced in the study 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 emphasizes the importance of fine-tuning LLMs for domain adaptation to achieve accurate sentiment classification and highlights RoBERTa's capabilities for YouTube comment SA.

This paper presents SOTA results, however, limited with the data labelling technique utilized. In addition, we used a 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 dataset 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.

Supporting Materials: Not applicable.

Author Contributions: "Conceptualization, H.S, F.U.D, and B.O.; methodology, H.S, F.U.D, and B.O.; software, F.U.D.; validation, H.S, F.U.D, and B.O.; formal analysis, F.U.D.; investigation, H.S, F.U.D, and B.O.; resources, H.S, F.U.D, and B.O.; data curation, H.S, F.U.D, and B.O.; writing—original draft preparation, H.S, F.U.D, and B.O.; writing—review and editing, H.S, F.U.D, and B.O.; visualization, H.S, F.U.D, and B.O.; supervision, H.S. and B.O.; project administration, H.S. and B.O.; All authors have read and agreed to the published version of the manuscript."

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

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. *Technological Forecasting and Social Change*, **2017**, 117, 238-250.
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. *Sustainable Energy, Grids and Networks*, **2023**, 101112.
3. International Energy Agency (IEA). *Global EV Outlook 2023*. <https://www.iea.org/reports/global-ev-outlook-2023> (accessed on 21st 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. *Engineering Applications of Artificial Intelligence*, **2024**, 127, 107216.
5. Su, C. W., Yuan, X., Tao, R., & Umar, M. Can new energy vehicles help to achieve carbon neutrality targets?. *Journal of Environmental Management*, **2021**, 297, 113348.
6. Zhao, X., Ma, Y., Shao, S., & Ma, T. What determines consumers' acceptance of electric vehicles: A survey in Shanghai, China. *Energy Economics*, **2022**, 108, 105805.
7. Hayashida, S., La Croix, S., & Coffman, M. Understanding changes in electric vehicle policies in the US states, 2010–2018. *Transport policy*, **2021**, 103, 211-223.
8. Morton, C., Anable, J., & Nelson, J. D. Exploring consumer preferences towards electric vehicles: The influence of consumer innovativeness. *Research in transportation business & management*, **2016**, 18, 18-28.
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 cities. *Transportation Research Part D: Transport and Environment*, **2013**, 18, 39–45. <https://doi.org/10.1016/j.trd.2012.09.007>.
10. Ogunleye, B. O. *Statistical learning approaches to sentiment analysis in the Nigerian banking context*. Sheffield Hallam University, United Kingdom. (PhD Thesis), **2021**.
11. Ogunleye, B., Brunson, T., Maswera, T., Hirsch, L. and Gaudoin, J. Using Opinionated-Objective Terms to Improve Lexicon-Based Sentiment Analysis. *Proceedings of the 12th International conference on soft computing for problem-solving (SocProS 2023)*. *Lecture Notes in Networks and Systems*, **2023**, vol 995, pp. 1-23. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3292-0_1

12. Chen, C. C., & Chang, Y. C. What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness. *Telematics and Informatics*, **2018**, 35(5), 1512-1523.
13. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding, **2018**. arXiv preprint arXiv:1810.04805.
14. Kumawat, S., Yadav, I., Pahal, N., & Goel, D. Sentiment analysis using language models: A study. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) **2021**, 984-988. IEEE.
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 **2023**, 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 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE) **2023**, pp. 461-465. IEEE.
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. *Knowledge Based Systems*, **2019**, 171, 81-92.
18. John, J. M., Shobayo, O., & Ogunleye, B. An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, **2023**, 2(4), 809-823.
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. **2022**.
20. Jena, R. An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. *Industrial Marketing Management*, **2020**, 90, 605-616.
21. Ha, S., Marchetto, D. J., Dharur, S., & Asensio, O. I. Topic classification of electric vehicle consumer experiences with Transformer-based deep learning. *Patterns*, **2021**, 2(2), 100195. <https://doi.org/10.1016/j.patter.2020.100195>.
22. Biswas, S., Young, K., & Griffith, J. Automatic Sentiment Labelling of Multimodal Data. In International Conference on Data Management Technologies and Applications **2021**, 154-175. Cham: Springer Nature Switzerland.
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. *Applied Soft Computing*, **2020**, 97, 106754.
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.
25. Hasan, A., Moin, S., Karim, A., & Shamshirband, S. Machine learning-based sentiment analysis for twitter accounts. *Mathematical and computational applications*, **2018**, 23(1), 11.
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 2021 5th International Conference on Electrical Information and Communication Technology (EICT) **2021**, 1-6. IEEE.
27. Ogunleye, B.; Sharma, H.; Shobayo, O. Sentiment Informed Sentence BERT-Ensemble Algorithm for Depression Detection. Preprints **2024**, 2024071325. <https://doi.org/10.20944/preprints202407.1325.v1>
28. Ruan, T., & Lv, Q. Public perception of electric vehicles on Reddit and Twitter: a cross-platform analysis. *Transportation research interdisciplinary perspectives*, **2023**, 21, 100872.
29. JustAnotherArchivist. Justanotherarchivist/snsrape: A social networking service scraper in Python. **2020**, GitHub. <https://github.com/JustAnotherArchivist/snsrape>
30. Grzegorzewski, P., & Kochanski, A. "Data Preprocessing in Industrial Manufacturing" *Soft Modeling in Industrial Manufacturing, Studies in Systems, Decision and Control*, **2019**, vol 183, 27-41, Springer Nature Switzerland https://doi.org/10.1007/978-3-030-03201-2_3
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 vaccination Twitter dataset. *Expert Systems with Applications*, **2023**, 212, 118715. <https://doi.org/10.1016/j.eswa.2022.118715>.
32. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., & Stoyanov, V. RoBERTa: A robustly optimized bert pretraining approach, **2019**. arXiv preprint arXiv:1907.11692.
33. Ogunleye, B.; Dharmaraj, B. The Use of a Large Language Model for Cyberbullying Detection. *Analytics* **2023**, 2, 694-707. <https://doi.org/10.3390/analytics2030038>
34. Bozanta, A., Angco, S., Cevik, M., & Basar, A. Sentiment analysis of stocktwits using transformer models. In 2021 20th IEEE international conference on machine learning and applications (ICMLA) **2021**, 1253-1258. IEEE.
35. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. XLNet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, **2019**, 32.

36. Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. Sentiment analysis of review datasets using naïve Bayes' and k-nn classifier. *International Journal of Information Engineering and Electronic Business*, **2016**, 8(4), 54–62. <https://doi.org/10.5815/ijieeb.2016.04.07>.
37. Ye, J., Zhou, J., Tian, J., Wang, R., Zhou, J., Gui, T., Q. Zhang & Huang, X. Sentiment-aware multimodal pre-training for multimodal sentiment analysis. *Knowledge-Based Systems*, **2022**, 258, 110021.
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(3), 388-401.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.