DOI: https://doi.org/10.34123/jurnalasks.v16i2.790

Available online at https://jurnal.stis.ac.id/

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 16 No. 2, December 2024
e- ISSN: 2615-1367, p-ISSN:2086-4132

Aspect-Based Sentiment Analysis of Transportation Electrification Opinions on YouTube Comment Data

Rahmi Elfa Adilla[1*], Muhammad Huda[2], Muhammad Aziz[3], Lya Hulliyyatus Suadaa[4]

[1]BPS-Statistics Indonesia, Jakarta, Indonesia, [2]PT PLN (Persero), Jakarta, Indonesia, [3]The University of Tokyo, Tokyo, Japan, [4]Politeknik Statistika STIS, Jakarta, Indonesia
**Corresponding Author: E-mail address:* [1]rahmielfa1010@gmail.com

ARTICLE INFO

Article history:
Received 27 August, 2024
Revised 29 October, 2024
Accepted 09 November, 2024
Published 31 December, 2024

Keywords:
Transportation Electrification; Electric Vehicles; Aspect-Based Sentiment Analysis; Machine Learning; Transfer Learning

Abstract

Introduction/Main Objectives: This research aims to conduct an aspect-based sentiment analysis of transportation electrification opinions on YouTube comment data. **Background Problems:** It is difficult to summarize the sentiment of many YouTube user comments related to electric vehicles (EVs) based on their aspects; therefore, aspect-based sentiment analysis is needed to conduct further analysis. **Novelty:** This study identifies five aspects of EV and their sentiments at the same time. The aspects are usefulness, ease of use, comfort, cost, and incentive policies. One of this study's methods is the transfer learning model. This model can be a solution to overcome the shortcomings of deep learning in classifying aspect-based sentiment classification on small datasets. **Research Methods:** The sentiment classification model used is a machine learning model, namely support vector machine (SVM) and transfer learning models from pre-trained IndoBERT and mBERT. **Finding/Results:** Based on the experimental results, transfer learning from the IndoBERT model achieved the best performance with accuracy and F1-Score of 89.17% and 52.66%, respectively. Furthermore, the best IndoBERT model was developed with input in the form of a combination of aspects and comment sentences. Experimental results show that there is an improvement in performance with accuracy and F1-Score of 90% and 60.70%, respectively.

1. Introduction

Motor vehicles have a direct effect on global environmental challenges and the depletion of natural resources. The transportation sector plays a crucial role in contributing to air pollution and climate change, particularly in urban regions, due to emissions of greenhouse gases (GHGs). This sector is responsible for approximately 25% of total global GHG emissions, with projections indicating that this figure could rise from 23% to 50% by 2030 [1]. As a result, transportation is identified as a significant obstacle to achieving a sustainable economy [2], [3].

Based on historical data from IQAir, Indonesia in 2022 ranked 26th out of 131 most polluted countries in the world. Meanwhile, according to IQAir real-time data, accessed on August 30, 2023 at 15:40 WIB, Jakarta is ranked second as the city with the poorest air quality globally, with 152 AQI US air indicators in red, which means that the air quality in Jakarta is unhealthy [4]. The order of cities with the poorest air quality globally is illustrated in Figure. 1.

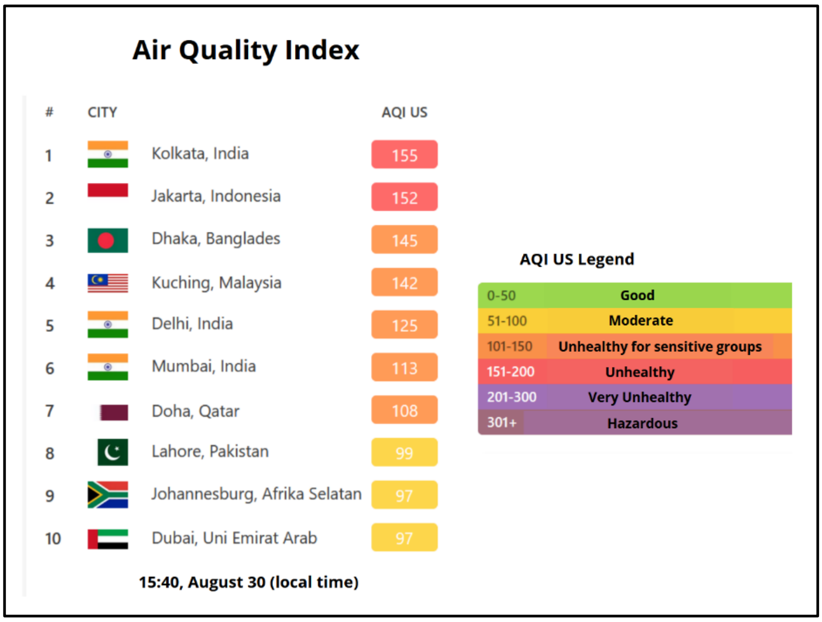

Figure. 1. Ranking of cities with the poorest air quality globally according to IQAir real-time data on August 30, 2023 at 15.40 WIB

Indonesia's poor air quality index cannot be ignored, so it needs special attention from the government, the community, and various other parties. Reducing greenhouse gas emissions from vehicles is one of the solutions [5].This encourages the electrification of road transportation, namely by replacing conventional vehicles with alternative energy options, such as electric vehicles (EVs). EVs are described as eco-friendly vehicles powered by electric motors that utilize energy from batteries and can be recharged from outside sources.[6]. EVs produce lower emissions, so they are environmentally friendly vehicles that can reduce air pollution. The Indonesian government established an EV development program through Presidential Regulation Number 55 of 2019 was issued on August 12, 2019, focusing on the acceleration of the Battery-Based Electric Motor Vehicle Program for road transportation. This regulation is the initial rule as a legal law for EVs in Indonesia.

The electrification of road transportation continues to be socialized by the government, both central and regional, as one way to reduce worsening air pollution. However, there are many parties who disagree with the opinion of transportation electrification. Many argue that switching from fossil fuel vehicles to EVs will not solve the problem and that electrification is a “false solution” because the energy source used for EVs still comes from coal-fired power plants, which are also a contributor to air pollution. In addition, one of the main reasons many parties disagree with the opinion of transportation electrification is the very expensive purchase cost, usage co14st, and maintenance cost of EVs [7].

Based on this background, the author will analyze public opinion regarding EVs. The data analyzed are public comments on transportation electrification opinions obtained from the YouTube platform. The YouTube platform was selected as the data source for this study due to its status as the second most visited website globally. [8], [9]. While there are several video-sharing platforms like Vimeo and Dailymotion, YouTube stands out as the site with the highest volume of videos shared among billions of users around the world [10]. It allows individuals to express their thoughts, ideas, and emotions through video content, for example, related to EVs, and respond by submitting comments. Comments on YouTube videos tend to be more relevant and appropriate to the context of the video shown, so that predetermined aspects of EVs can be extracted more easily and the analysis process can be carried out [11]. The analysis process to find out the public's response to the electrification of transportation is referred to as sentiment analysis. Sentiment analysis is typically used to determine whether a particular opinion conveys a positive, negative, or neutral sentiment. This shows that sentiment analysis in general has not been able to identify sentiment for certain aspects of EVs. If the opinion given by a person on an aspect of an EV has a positive sentiment, it does not necessarily mean that the person also gives positive sentiments to all aspects of EVs, and vice versa. For example, in the sentence “EVs produce low emissions, but the price is very high,” the emission aspect of electric vehicles has a positive sentiment, but gives a negative sentiment towards the price. Therefore, a more complete analysis is needed to determine the sentiment towards EVs based on their aspects. This analysis is called aspect-based sentiment analysis. Constructing a model to understand the sentiment of transportation electrification is essential, as various stakeholders can benefit from this analysis. For instance, the government can utilize the insights to support the development of incentive policies for electric vehicles. The state electricity company (PLN) can gain valuable information regarding the needs for transportation electrification, including the establishment of charging stations and the necessary voltage or wattage for home setups. Additionally, EV suppliers can benefit from insights related to user interests and complaints, which can inform product improvements and customer service strategies. Overall, this model provides crucial information that can guide decision-making for multiple parties involved in the electrification of transportation.

In this study, aspect-based sentiment analysis can be utilized to analyze public responses regarding EVs in various aspects. According to the technology acceptance model (TAM) of EVs and extensions of the model [5], [12], [13], the aspects used in this study are usefulness, ease of use, comfort, cost, and incentive policies. The classification model used to solve the aspect-based sentiment analysis task in this study is a machine learning model, namely the support vector machine (SVM), which provides the best performance in the research of Mustakim and Priyanta [14]. Mustakim and Priyanta [14] conducted aspect-based sentiment analysis to extract aspects of KAI Access user reviews in the review column on the Google Play Store. The methods used are a naïve bayes classifier (NBC) and an SVM. The aspects used are learning ability, memory, efficiency, and errors. The results showed that the majority of user sentiments were negative in each aspect, with the most discussed aspect of errors indicating high system errors. The test results provide the best model in SVM with hyperparameter tuning with an average accuracy score of 91.63%, F1-Score of 75.55%, and recall of 74.47%.

Jeong has carried out an aspect-based sentiment analysis of EVs [15]. The study utilized user satisfaction data from automotive forums (Edmunds.com, Cars.com, Cargurus.com, Carfax.com, and Carbuyers.co.uk) and YouTube reviews. Sixteen main aspect categories were used, namely eight main components of EVs and eight main characteristics of human factors. As a result of the study, it was found that users had positive sentiments on the aspects of acceleration, space, interior, power, safety, ergonomics, price, and power. In addition, users had negative sentiments about seats, battery, charging, noise, winter, and ice.

Related research has also been conducted by Anwar et al. [16] regarding aspect-based sentiment analysis on car reviews. The data used is car review data sourced from the Edmunds website (www.edmunds.com). The results showed that most of the positive sentiments were in the aspects of comfortable driving, excellent fuel efficiency, dependability, comfort, great value, a useful rear camera, a smooth and quiet ride, impressive acceleration, an appealing design, a quality sound system, and a robust build. Some negative sentiment elements share similarities with those in the positive category, although they appear less frequently. Jena [17] also conducted research related to consumer sentiment towards EVs with a big data approach. The data in this study used data for two years (2016 to 2018) collected from various social media platforms by extracting the data into aspects of price, maintenance, and safety. The classification analysis shows that price and maintenance have mostly negative sentiments, while security aspects have mostly neutral sentiments.

In addition, researchers also use transfer learning models from pre-trained models used in the research conducted by Tao and Fang [18]. Tao and Fang [18], who proposed a transfer learning-based approach. First, the proposed approach extends the aspect-based sentiment analysis method with multi-label classification capability. Second, it proposes an advanced sentiment analysis method for categorizing text into sentiment categories while considering the aspects of entities. Third, it extends two transfer learning models, namely Bidirectional Encoder Representations from Transformers (BERT) and XLNet, as analysis tools for aspect-based and multi-label sentiment classification. The data used are three datasets from different domains, namely reviews related to Yelp restaurants, wine, and movies. The findings indicate that the suggested transfer learning model consistently surpasses the baseline SVM and CNN models across all three datasets. This evidence supports the notion that the proposed method is better suited for managing multi-label classification tasks. Liu and Zhao proposed the utilization of the transfer learning approach in aspect-based sentiment analysis using Amazon product review data [19]. The classification models used are CNN and BERT, which use a combination of corpus at the aspect level and corpus at the sentence level to create sequential sentence pairs as input. The results show that the BERT model, which combines aspects and sentences, performs better than the CNN model. The transfer learning model is a solution to overcome the shortcomings of deep learning in performing aspect-based sentiment classification on small datasets [20]. Both types of models are evaluated using accuracy, precision, recall, and F1-Score values. Furthermore, a model comparison is carried out to determine which model performs better at performing aspect-based sentiment classification on public responses to transportation electrification opinions on YouTube comment data.

Based on the described background, this research has several objectives. First, it aims to construct a dataset for sentiment classification focused on aspects of electric vehicles (EVs) using YouTube comment data. Second, the study intends to perform a descriptive analysis of this data to gain insights into public opinion regarding transportation electrification. Third, it seeks to develop an aspect-based sentiment analysis model by applying machine learning and transfer learning techniques to predict public sentiment towards opinions on transportation electrification as reflected in YouTube comments. Finally, the research will evaluate the performance of these aspect-based sentiment analysis models.

The limitations of this research conducted by the researcher are as follows: First, the data utilized in this research was collected through scraping YouTube comments using the YouTube Data API. Second, the data was collected from comments on the top 11 recommended videos based on a YouTube search conducted on September 9, 2023, using the keyword "electric vehicles." Third, the aspects of electric vehicles examined in this research include usefulness, ease of use, comfort, cost, and incentive policies. Fourth, the machine learning algorithm employed for aspect-based sentiment analysis is the Support Vector Machine (SVM). Finally, the transfer learning algorithms used for aspect-based sentiment analysis are IndoBERT and mBERT.

2. Material and Methods

The analytical approach employed in this study is aspect-based sentiment analysis, utilizing machine learning and transfer learning classification models. The research process follows the stages outlined in the Cross Industry Standard Process for developing Machine Learning applications with a Quality assurance methodology (CRISP-ML(Q)). This framework includes steps such as understanding business and data, preparing data, engineering models, evaluating models, and managing model operations [21]. The stages of this research are depicted in Figure. 2.

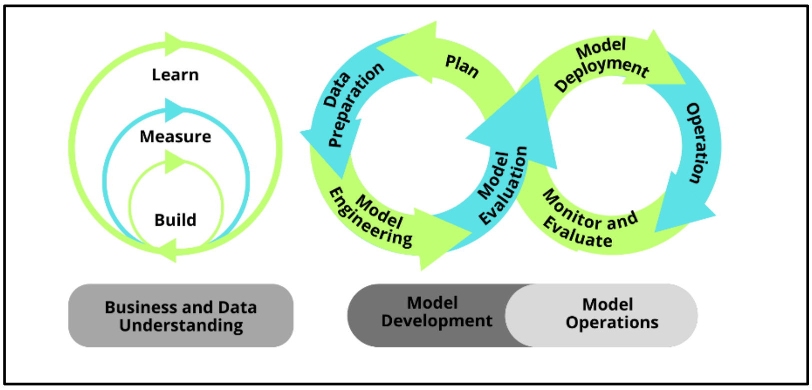

Figure. 2. Life cycle process of research stage

2.1. Business and data understanding

This research starts from the importance of supporting information about the level of public acceptance of transportation electrification, which is a government effort to decarbonize and reduce dependence on fossil fuels in Indonesia. YouTube user comments on EVs can be used to conduct aspect-based sentiment analysis, which can be used as input to overcome shortcomings and add advantages to the use of EVs in the future. It is difficult to summarize the sentiment of many YouTube user comments related to EVs based on their aspects; therefore, aspect-based sentiment analysis is needed to conduct further analysis.

a b

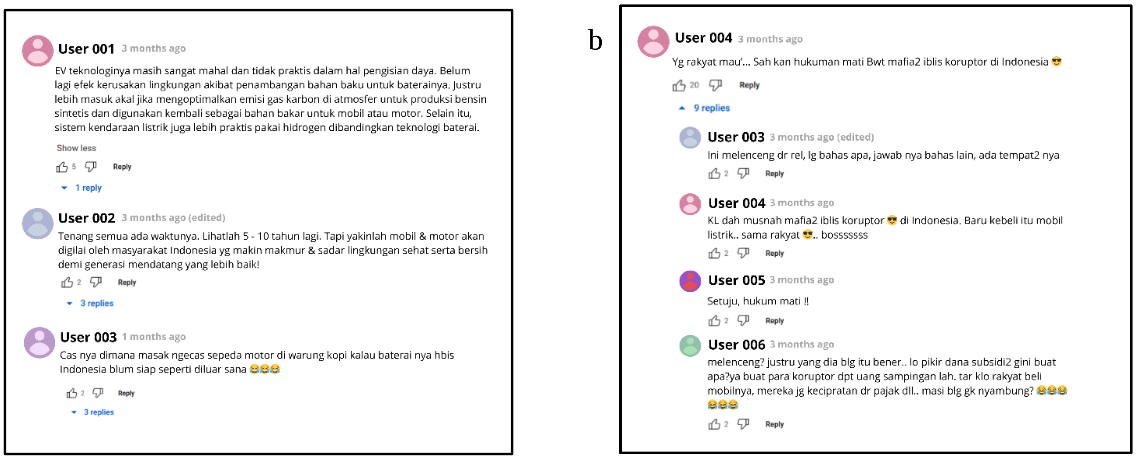

Figure. 3. (a) example of YouTube user comments on EVs; (b) examples of irrelevant YouTube user "reply" comments

The dataset utilized in this research is data collected through scraping YouTube comments using YouTube Data API, which is relevant to the topic of EVs. The dataset used in this research is data obtained from scraping YouTube comments using YouTube Data API, which is relevant to the topic of EVs. The data is collected from the comments of the top 11 recommended videos from YouTube search results. Limiting the analysis to comments from the top 11 recommended videos from the YouTube search results because it will compare machine learning and transfer learning model, with machine learning as the base model. Machine learning requires a lot of training data to build a good model, while transfer learning can overcome the availability of small datasets with better model result [20]. So to validate that transfer learning can overcome this, only the top 11 recommended videos from the YouTube search results were used. The keyword used in video searches is "kendaraan listrik". The keyword was chosen so that the video search results obtained cover the topic of EVs in general so that they can be classified into their aspects. An example of comments on a YouTube video is illustrated in Figure 3 (a). In the data collection process, comments in the form of "replies" were not collected in this study. This is because "replies" tend to deviate from the observed keywords, which would make the data collected irrelevant to EVs. An example of an irrelevant reply is illustrated in Figure. 3 (b).

2.2. Data preparation

2.2.1. Data labelling

After collecting data, the next step is data labeling. Data labeling is performed for each comment based on specific aspects. Based on the technology acceptance model (TAM) of EVs and extensions of the model [5], [12], [13], the aspects used in this study are usefulness, ease of use, comfort, cost, and incentive policies. Usefulness is defined as the extent to which consumers believe that electric vehicles can improve their lives, particularly in environmental performance. Uses of electric vehicles include reducing carbon emissions, controlling engine fuel consumption, improved health quality due to protection from air pollution/smog, and so on. Ease of use is how an item can be understood, used, or operated. The ease of use of electric vehicles includes easy operation, automatic transmission, easy access to charging infrastructure, speed of charging time, battery life, and so on. Electric vehicle comfort reflects how users overall perception of the exterior, interior, and technical electric vehicles such as seat comfort, sound noise level, space, electric vehicle model, and safety. Meanwhile, the cost of electric vehicles includes purchase costs, usage costs, and maintenance costs. To overcome the problem of the high price effect, the government provides incentive policy support in the form of purchase subsidies, tax deductions in the use of EVs, the use of special number plates, and so on.

To reduce subjectivity, data labeling is performed by no fewer than three annotators, and the final outcomes is determined based on majority voting. The sentiment labels used for each aspect consist of "positive," "negative," or "neutral" labels. For comments that do not contain predefined aspects, the sentiment label given is "none." An illustration of data labeling can be seen in Figure. 4.

Comment				
EV teknologinya masih **sangat mahal** dan **tidak praktis dalam hal pengisian daya**. Belum lagi efek **kerusakan lingkungan akibat penambangan bahan baku untuk baterainya**. Justru lebih masuk akal jika mengoptimalkan emisi gas karbon di atmosfer untuk produksi bensin sintestis dan digunakan kembali sebagai bahan bakar untuk mobil atau motor. Selain itu, sistem kendaraan listrik juga lebih praktis pakai hidrogen dibandingkan teknologi baterai. *EV technology is still **very expensive** and **cumbersome in terms of charging**. Not to mention the **environmental damage caused by mining the raw materials for the batteries**. Instead, it makes more sense to optimize carbon gas emissions in the atmosphere for the production of synthetic gasoline and reuse it as fuel for cars or motorcycles. In addition, EV systems are also more practical with hydrogen than battery technology.*				
Aspects				
Usefulness	Ease of Use	Comfort	Cost	Incentive Policies
negative	negative	none	negative	none

Figure. 4. Example of data labeling

To see the level of consistency of several annotators in labeling data, an Inter-Rater Reliability test was conducted using Krippendorff's alpha. Krippendorff's alpha's calculation can be done using the following formula [22]. D_o represents observed disagreement, and D_e represents expected disagreement that occurs by chance. The value of α can range from 0 to 1, with 0 representing an indication of no agreement (no reliability), and 1 representing perfect agreement among raters (perfect reliability).

$$\alpha = 1 - \frac{D_o}{D_e} \tag{1}$$

2.2.2. Data preprocessing

Before Preprocessing
Terbukti subsidi mobil listrik tidak hanya untuk orang yg mampu..pupuk atau yg lain lebih bermanfaat. #pakai akal sehat *Evidently electric car subsidies are not only for the well-off..fertilizer or something else is more useful. #use common sense.*
After Preprocessing
terbukti subsidi mobil listrik tidak hanya untuk orang yg mampu pupuk atau yg lain lebih bermanfaat *it is proven that electric car subsidies are not only for people who can afford fertilizer or other more useful things*

Figure. 5. Example of data preprocessing

Following the data collection and labeling, the next step is preprocessing. Preprocessing is a technique for cleaning data. The preprocessing procedures applied in this study are outlined below.

1) Cleaning, removing irrelevant characters including usernames, hashtags, and URLs.
2) Removing excess punctuation marks, letters, or whitespace.
3) Case folding, the process of converting text into entirely lowercase letters.

In this research, the stopword removal step is not used because the process can remove some important words or phrases that are relevant to keywords in sentiment classification. For example, the word “tidak” can be a determinant of a negative sentiment, so if removed, it can classify the sentiment as positive. An example of the results from preprocessing is depicted in Figure. 5.

2.2.3. Feature extraction

In this study, the model classification relies on the Term Frequency-Inverse Document Frequency (TF-IDF) word weighting method for feature extraction. TF-IDF is a technique used to assess the significance of a word within a collection of documents. Term frequency (TF) indicates that a word's weight increases with its frequency in a document, while inverse document frequency (IDF) suggests that a word's weight decreases as it appears more frequently across the document set. Thus, the TF-IDF score rises when a word occurs more frequently in a specific document and is less common in the overall collection of documents [23]. Feature extraction is applicable solely to machine learning models. In transfer learning, the extraction process does not start from scratch; rather, it leverages token vector weights from models that have previously been trained on a more extensive dataset. The following is the formula for calculating TF-IDF [24].

$$TF - IDF = tf_{ij}idf_i = tf_{ij} \times log\frac{D}{df_i} \quad (2)$$

2.3. Model engineering

In this research, the classification model is developed using machine learning techniques, specifically SVM and transfer learning models derived from pre-trained IndoBERT and mBERT.

1. Machine Learning
   The machine learning model used is SVM. SVM is a technique in machine learning based on vector space that aims to identify the decision boundary between two classes, ensuring it is as far away as possible from any point in the training dataset. In essence, SVM seeks to determine the optimal hyperplane by maximizing the separation between the classes. [25].
2. Transfer Learning
   - IndoBERT
     IndoBERT is a bidirectional Transformer model that has been pre-trained on an extensive Indonesian corpus (Indo4B), which includes a mix of formal and informal language sources like Indonesian Wikipedia, websites, news articles, video subtitles, blogs, and social media. [26].
   - mBERT
     Multilingual BERT, also known as mBERT, is a bidirectional Transformers model trained on 102 (uncased) and 104 (cased) languages in Wikipedia.

2.4. Model evaluation

In this research, the evaluation of the model was conducted by comparing the precision, recall, accuracy, and F1-Score metrics of both the machine learning and transfer learning models [25].

- Precision
  Precision is the proportion of true positive data to the total predicted positive data.
  $$precision = \frac{TP}{TP+FP} \quad (3)$$
- Recall
  Recall is the proportion of true positive data compared to the total amount of actual positive data.
  $$recall = \frac{TP}{TP+FN} \quad (4)$$
- Accuracy
  Accuracy is the proportion of the amount of data that is predicted correctly from all data.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

- F1-Score

  F1-Score is a combination measure that combines both precision and recall measures (weighted harmonic mean).

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (6)$$

Description:

True Positive (TP):	Actual is in the class and predicted is also in the class
False Negative (FN):	Actual is in the class and predicted is not in the class
False Positive (FP):	Actual is not in the class and predicted to be in the class
True Negative (TN):	Actual is not in the class and predicted is not in the class either

The model evaluation stage in this study utilizes the cross-validation method. Cross-validation is a technique employed to identify an suitable model for making predictions. The dataset is split into two segments: one segment is utilized to train the model, while the other is utilized for evaluation the model's predictive performance. The most fundamental form of cross-validation is called k-fold cross-validation [27]. This research uses 5-fold cross-validation.

2.5. Model operation

Based on the best model obtained, the model was implemented to predict aspect-based sentiment labels on a new unlabeled dataset. The new dataset used is obtained from scraping the comments of a YouTube video about the revision of EV regulations. Based on the implementation results, the prediction results are checked by manually calculating the evaluation metric, which compares the aspect-based sentiment label by the annotator along with the results of the predictions of the aspect-based sentiment label by the best model.

3. Result and Discussion

3.1. Business and data understanding

The scraping process was carried out on September 9th, 2023, using the YouTube Data API on each video from the top 11 YouTube video search results with the keyword “electric vehicle” and obtained a dataset of 3,669 comments. The dataset consists of comments that are relevant and irrelevant to EVs. After manual filtering of 3,669 comments, 1,881 comments are relevant and 1,788 comments are not relevant to EVs. In this study, only relevant comments were used in the analysis process.

The relevant comment data consists of Indonesian, English, and other languages. The percentage of languages used by YouTube users in commenting on transportation electrification is shown in Figure. 6. Language identification is done using the langdetect library. The results obtained state that Indonesian is the most widely used language in commenting, with a percentage of 91%. Furthermore, followed by English by 1% and the rest is categorized in other languages with a percentage of 8%. In identifying languages, there are some errors because of the constraints of the model employed by langdetect. For instance, in Indonesian comments that use informal words, the langdetect library identifies the comment as a non-Indonesian comment. Figure. 7 shows a visualization of the most frequently occurring words in the dataset of comments on EVs. The most frequent words are mobil listrik (electric car), kendaraan listrik (EV), subsidi (subsidies), mahal (expensive), infrastruktur (infrastructure), and harga (price).

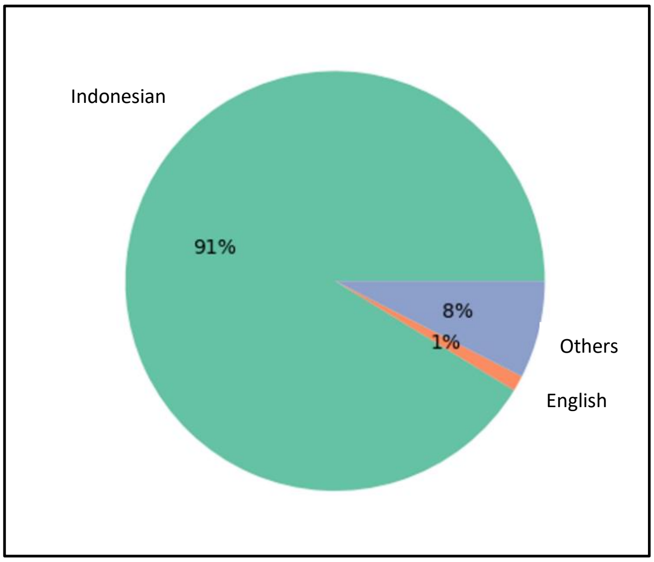

Figure. 6. Percentage of language in YouTube comment based on the langdetect library

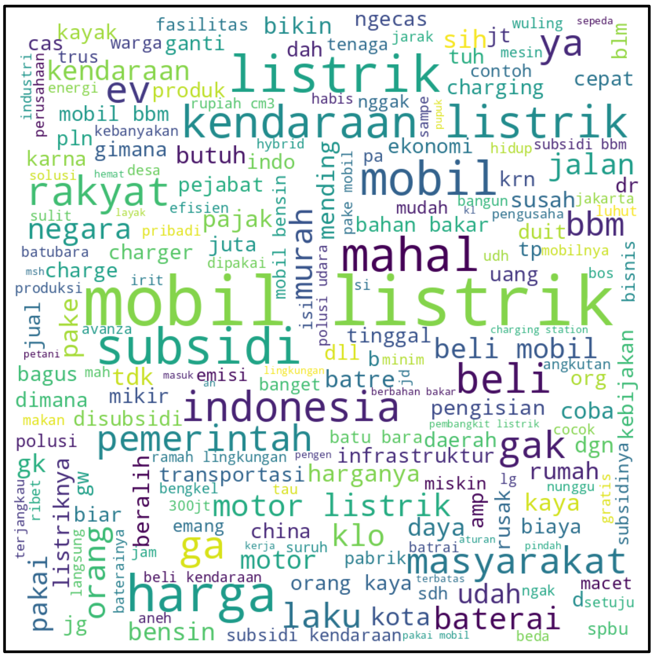

Figure. 7. Word cloud of relevant data set to EV

3.2. Data preparation

3.2.1. Data labeling

Labeling of comment data was done manually by three undergraduate students as annotators, and the final result was determined by majority voting. Labels that received disagreement from at least two annotators were excluded from the analysis process. From 1,881 datasets relevant to EVs, 1839 labeled data were obtained. There is a reduction in the dataset of 42 data due to disagreement by at least two annotators on the labels given. The distribution of labeling results after applying majority voting is shown in Figure. 8.

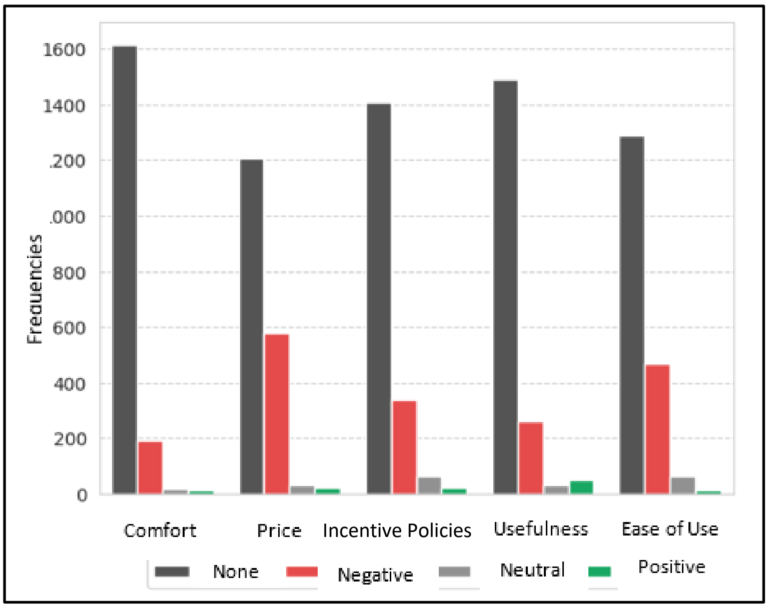

Figure. 8. Distribution of labeling results based on sentiment per aspect

Based on the distribution graph of labeling results for each aspect of EVs in Figure. 8, it is clear that the most commented aspects are the cost aspect and the ease of use aspect. To see the level of consistency of the three annotators in labeling the data, an inter-rater reliability test was conducted using Krippendorff's alpha. The higher the alpha value, the more consistent the annotators are in labeling the data and vice versa. The alpha value for each aspect of EVs is shown in Table 1. According Table 1, each aspect demonstrates a relatively high alpha value, nearing 1. This means that the three annotators have a high level of agreement in labeling, resulting in a reliable dataset.

Table 1. Inter-rater reliability test after majority voting

Aspect	Alpha
Usefulness	0.734
Ease of use	0.823
Comfort	0.776
Cost	0.852
Incentive policies	0.781

3.2.2. Dataset statistics

The characteristics of the obtained dataset are shown in Table 2. According to Table 2, it is clear that on average, there are 23 words per comment, 20 unique words per comment, and two sentences per comment.

Table 2. Dataset statistics

Characteristics	Minimum	Maximum	Average
Word count per comment	1	411	23
Number of unique words per comment	1	257	20
Number of sentences per comment	1	63	2

3.2.3. Data preprocessing

The preprocessing stages carried out in this research are cleaning, removing punctuation, letters or excess spaces, and case folding. In this research, the stopword removal step is not used because the process can remove some important words or phrases that are relevant to keywords in sentiment classification.

3.2.4. Feature extraction

Feature extraction is performed using TF-IDF and the TfidfVectorizer library from the Sklearn module to assess the significance of a word within the document. The TF-IDF hyperparameters for each machine learning model are determined from the results of a grid search, namely $min\_df = 1$, $max\_df = 0.25$, and $max\_features = 2000$.

3.3. Model Engineering

In this study, a classification model using a machine learning model, namely SVM, is constructed using a machine learning model, which is already available in the Scikit-learn library, also known as sklearn. Hyperparameters for each machine learning model are derived from the outcomes of the grid search. The SVM model is built with cost 10, gamma 0.1, and radial basis function (rbf) kernel parameters. To build transfer learning models from pre-trained models, the PyTorch and transformer libraries are utilized, employing Adam as the optimization algorithm, with a batch size of 10 and a specified learning rate of $1\ x\ 10^{-5}$ which are the default parameters.

A representation of the architecture of the pre-trained model used is shown in Figure. 9 and Figure**.** 10. Text classification with IndoBERT and mBERT is done by tokenizing or converting sentences into tokens, then converted into vector representations in high-dimensional space through token embeddings, segment embedding, and position embedding. The representation is passed to the transformer layer which allows the model to understand the context of the sentence in depth. The output from the Transformer layer is subsequently sent to the classification layer, which generates probabilities to determine which class the sentence belongs to in terms of aspect and sentiment.

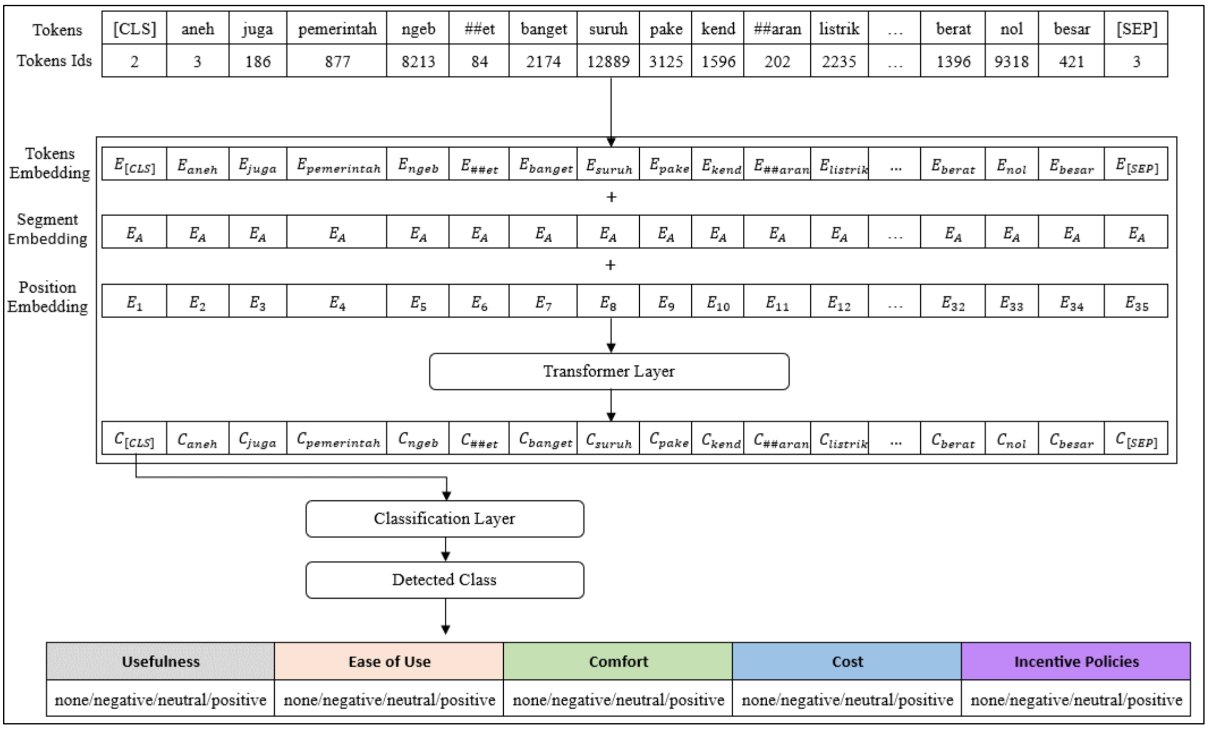

Figure. 9. Illustration of fine-tuned IndoBERT and mBERT for aspect-based sentiment classification

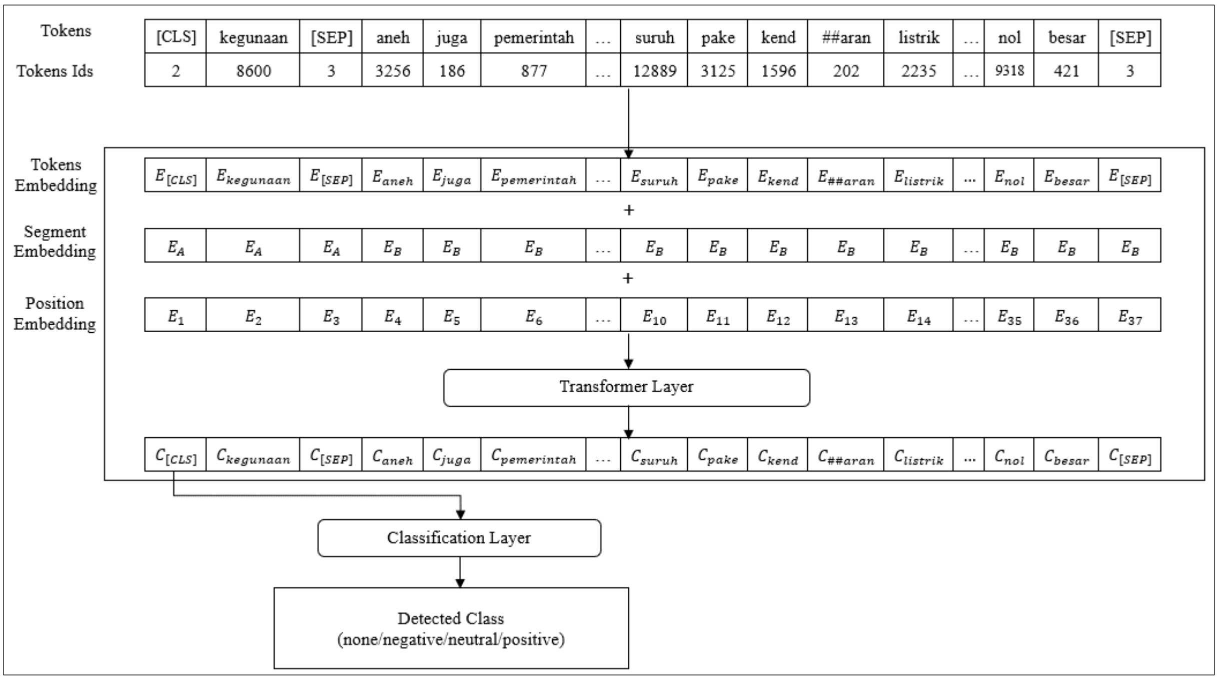

Figure. 10. Illustration of fine-tuned IndoBERT (with aspect and comment combination inputs) for aspect-based sentiment classification

3.4. Model evaluation

Experimental results of the aspect-based sentiment classification model using 5-fold cross-validation is presented in Table 3. Based on Table 3, it is clear that the transfer learning model from pre-trained IndoBERT exhibits the highest accuracy and F1-Score in comparison to the other classification models. This is because the transfer learning model has been trained in advance on a vast language corpus and adjusted its model parameters through a fine-tuning stage with training data for aspect-based sentiment analysis tasks. While the machine learning model learns solely from the training data provided in this study.

Table 3. Experimental results of aspect-based sentiment classification model using 5-fold cross-validation

Model	Accuracy	Precision	Recall	F1-Score
Machine Learning				
SVM	0.8765	**0.7798**	0.4454	0.4818
Transfer Learning				
IndoBERT	**0.8917**	0.6588	**0.4975**	**0.5266**
mBERT cased	0.8722	0.4311	0.4244	0.4188
mBERT uncased	0.8722	0.4843	0.4248	0.4258

Of all the fine-tuning models, the IndoBERT classification model records the highest evaluation scores, achieving accuracy and F1-Score values of 89.17% and 52.66%, respectively. Regarding the language corpus for each model, the mBERT model was trained on 102 languages (uncased) and 104 languages (cased), which includes Indonesian. However, in this study, the IndoBERT model provides better evaluation results. This can happen because multilingual BERT (mBERT) is better suited for multilingual datasets compared to the IndoBERT model. IndoBERT was developed using a substantial Indonesian corpus (Indo4B) that encompasses both formal and informal languages, which makes it more appropriate for the data utilized in this research. In addition, YouTube user comment data generally uses non-formal language and has a percentage of Indonesian language usage of 91%, which is much greater than the utilization of English along with other languages. As a result, the IndoBERT model is more appropriate for this study compared to the uncased or cased mBERT models.

Table 4. Aspect classification results for each label on IndoBERT using 5-fold cross-validation

Aspect	Sentiment	Precision	Recall	F1-Score	Support
Usefulness	None	0.9114	0.9626	0.9362	146
	Negative	0.6843	0.6508	0.6655	29
	Neutral	0.0000	0.0000	0.0000	4
	Positive	0.4000	0.2429	0.2927	6
Ease of Use	None	0.9127	0.9398	0.9260	129
	Negative	0.7678	0.8096	0.7874	46
	Neutral	0.5200	0.1857	0.2480	6
	Positive	0.2000	0.0667	0.1000	2
Comfort	None	0.9409	0.9818	0.9606	162
	Negative	0.7240	0.5517	0.6156	20
	Neutral	0.0000	0.0000	0.0000	1
	Positive	0.0000	0.0000	0.0000	1
Cost	None	0.9280	0.9337	0.9304	123
	Negative	0.8282	0.8565	0.8411	57
	Neutral	0.0000	0.0000	0.0000	2
	Positive	0.4000	0.1400	0.2000	3
Insentive Policies	None	0.9424	0.9619	0.9518	142
	Negative	0.7590	0.8072	0.7729	35
	Neutral	0.4500	0.1836	0.2452	6
	Positive	0.0000	0.0000	0.0000	2

The distribution of sentiment classification scores for every aspect of the IndoBERT model is presented in Table 4. Based on Table 4, it is evident that aspect-based sentiment classification utilizing the IndoBERT model provides varying F1-Score results. There are models that achieve very high scores for each aspect with the sentiment label “none,” each of which has an F1-Score of more than 90%. In addition, there is an F1-Score that exceeds 60% on each aspect with a “negative” sentiment label. There is also the lowest F1-Score reaching 0% on the “neutral” and “positive” sentiment labels in some aspects. This can occur due to the uneven distribution of datasets for each aspect and sentiment label, as indicated in the "support" column of Table 4. In addition, this also occurs due to the large number of “none” sentiment labels for each aspect in the dataset due to the limited types and number of predetermined aspects, which are only limited to the usefulness aspect, ease of use aspect, comfort aspect, cost aspect, and incentive policies aspect.

In the machine learning classification model, SVM provides better F1-Score results than the transfer learning model of pre-trained mBERT. In general, the F1-Score for each aspect and sentiment label in SVM has a similar pattern to the results of the previous IndoBERT model due to dataset imbalance. Following Liu and Zhao [19], the best IndoBERT model was developed for aspect-based sentiment classification with input in the form of a combination of aspects and comment sentences. The results of the experiment are presented in Table 5.

Table 5. Experimental results of aspect-based sentiment classification model on IndoBERT using 5-fold cross-validation

Model	Accuracy	Precision	Recall	F1-Score
IndoBERT	0.8917	0.6588	0.4975	0.5266
IndoBERT (combination of aspect and sentence)	**0.9000**	0.6444	0.5921	**0.6070**

Based on Table 5, it can be seen that the IndoBERT model with a combination of aspects and comment sentences has higher accuracy and F1-Score than the IndoBERT model without a combination of aspect and comment sentences. This is because the presence of aspects in the input can help the model better classify comment sentences into their sentiment labels. Of all the aspect-based sentiment classification models that have been built, the IndoBERT classification model with a combination of aspect and comment sentences has the highest evaluation score, with accuracy and F1-Score values of 90% and 60.70%, respectively. The distribution of sentiment classification scores for each aspect in the IndoBERT model (combination of aspect and sentence) is shown in Table 6. In general, the F1-Score for each aspect and sentiment label in IndoBERT (combination of aspect and sentence) increases and has a similar pattern to the results of the previous IndoBERT without combination input of aspect and sentence.

Table 6. Aspect classification results for each label on IndoBERT (combination of aspect and sentence) using 5-fold cross-validation

Aspect	Sentiment	Precision	Recall	F1-Score	Support
Usefulness	None	0.9220	0.9574	0.9392	146
	Negative	0.7203	0.6690	0.6930	29
	Neutral	0.3333	0.1400	0.1809	4
	Positive	0.4333	0.3262	0.3661	6
Ease of Use	None	0.9263	0.9522	0.9388	129
	Negative	0.8306	0.7909	0.8093	46
	Neutral	0.4655	0.5048	0.4594	6
	Positive	0.2000	0.0667	0.1000	2
Comfort	None	0.9622	0.9660	0.9639	162
	Negative	0.6951	0.6524	0.6652	20
	Neutral	0.1167	0.3000	0.1667	1
	Positive	0.5000	0.4167	0.5470	1
Cost	None	0.9452	0.9491	0.9471	123
	Negative	0.8648	0.8692	0.8662	57
	Neutral	0.0667	0.0667	0.0667	2
	Positive	0.6000	0.2700	0.3633	3
Insentive Policies	None	0.9500	0.9617	0.9555	142
	Negative	0.7975	0.8011	0.7902	35
	Neutral	0.4050	0.3569	0.3520	6
	Positive	0.0000	0.0000	0.0000	2

<table>
<tr><th colspan="5">Comment</th></tr>
<tr><td colspan="5">Kndaraan listrik cm di ousat kl di kmpg giln ngecas di mn itu yg jd maslh trs model motor jg jlk jlk modelnya beat smua

EVs are only in the city center if in the village where to charge, that's the problem, and the motorcycle model is also ugly, all beat models.</td></tr>
<tr><th colspan="5">SVM Prediction</th></tr>
<tr><td>Usefulness</td><td>Ease of Use</td><td>Comfort</td><td>Cost</td><td>Incentive Policies</td></tr>
<tr><td>none</td><td>none</td><td>negative</td><td>none</td><td>none</td></tr>
<tr><th colspan="5">IndoBERT Prediction</th></tr>
<tr><td>Usefulness</td><td>Ease of Use</td><td>Comfort</td><td>Cost</td><td>Incentive Policies</td></tr>
<tr><td>none</td><td>negative</td><td>negative</td><td>none</td><td>none</td></tr>
<tr><th colspan="5">IndoBERT (combination aspect and sentence) Prediction</th></tr>
<tr><td>Usefulness</td><td>Ease of Use</td><td>Comfort</td><td>Cost</td><td>Incentive Policies</td></tr>
<tr><td>none</td><td>negative</td><td>negative</td><td>none</td><td>none</td></tr>
<tr><th colspan="5">Gold Label</th></tr>
<tr><td>Usefulness</td><td>Ease of Use</td><td>Comfort</td><td>Cost</td><td>Incentive Policies</td></tr>
<tr><td>none</td><td>negative</td><td>negative</td><td>none</td><td>none</td></tr>
</table>

Figure. 11. Example of comment 1, gold label and predicted label using SVM and IndoBERT

An example of a comment with the actual label given by the annotator (gold label), as well as the prediction results from the IndoBERT and SVM classification models, are shown in Figure. 11. The prediction example of the comment shows that the IndoBERT model is able to provide the same prediction results as the gold label. Meanwhile, the SVM model did not succeed in identifying sentiment labels on the ease of use aspect. This example leads to the conclusion that the IndoBERT model is better at capturing sentence context than the SVM model.

3.5. Model operation

Based on the best model obtained, namely the IndoBERT model with input combinations of aspects and comment sentences, the implementation of the model is carried out to predict aspect-based sentiment labels on a new unlabeled dataset. The new dataset used is 69 comment data obtained from scraping the comments of a YouTube video about the revision of EV regulations.

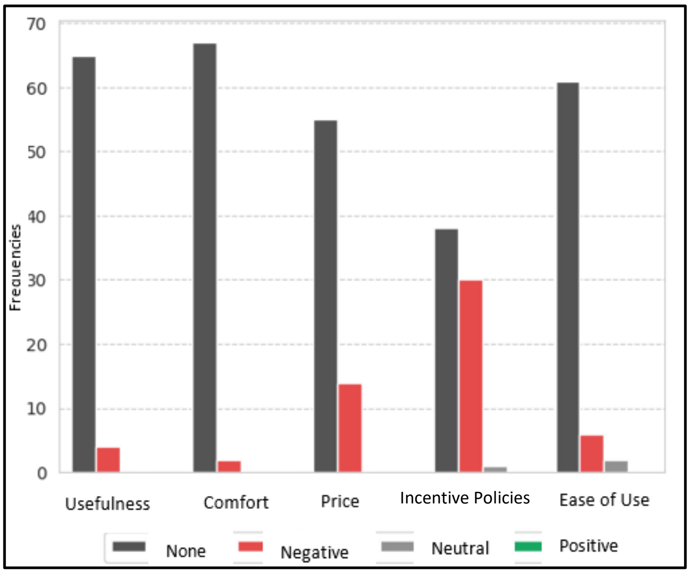

Figure. 12. Distribution of sentiment label prediction results per aspect

The results of the aspect-based sentiment label prediction of the video comments are depicted in Figure. 12. Based on Figure. 12, it can be seen that the comments on the revision of the EV rules are overall negative. The most commented aspects are the incentive policies aspect and the cost aspect.

Based on Figure. 13, it can be seen that the most frequently occurring words in the usefulness aspect are the words “mobil”, “listrik”, “menguntungkan”, “polusi”, and “lingkungan” (“car”, “electricity”, “favourable”, “pollution”, and “environment”). The words that appear most often in the ease of use aspect are the words “listrik”, “kendaraan”, “ekosistem”, “pengisian”, and “insfrastruktur” (“electricity”, “vehicles”, “ecosystem”, “charging”, and “infrastructure”). The most frequently occurring words in the comfort aspect are the words “mobil”, “listrik”, “aneh”, and “bentuk” (“car”, “electricity”, “weird”, and “shape”.). The words that appear most often in the cost aspect are the words “mobil”, “listrik”, and “harga” (“car”, “electricity”, and ’price’). Meanwhile, the words that appear most often in the incentive policies aspect are the words “pajak”, “mobil”, “listrik”, and “subsidi” (“tax”, “car”, “electricity”, and “subsidy”).

Furthermore, the prediction results are checked manually and the evaluation metrics of the aspect-based sentiment label prediction results on IndoBERT are obtained with the input of a combination of aspects and comment sentences from the dataset on the revision of EV rules. The experimental results are presented in Table 7. Based on Table 7, it is clear that the IndoBERT model with the input of a combination of aspects and comment sentences from the dataset on the revision of EV rules has an accuracy and F1-Score that is almost similar to the dataset on EVs. This means that the best model obtained has been able to perform aspect-based sentiment classification well on other EV-related datasets.

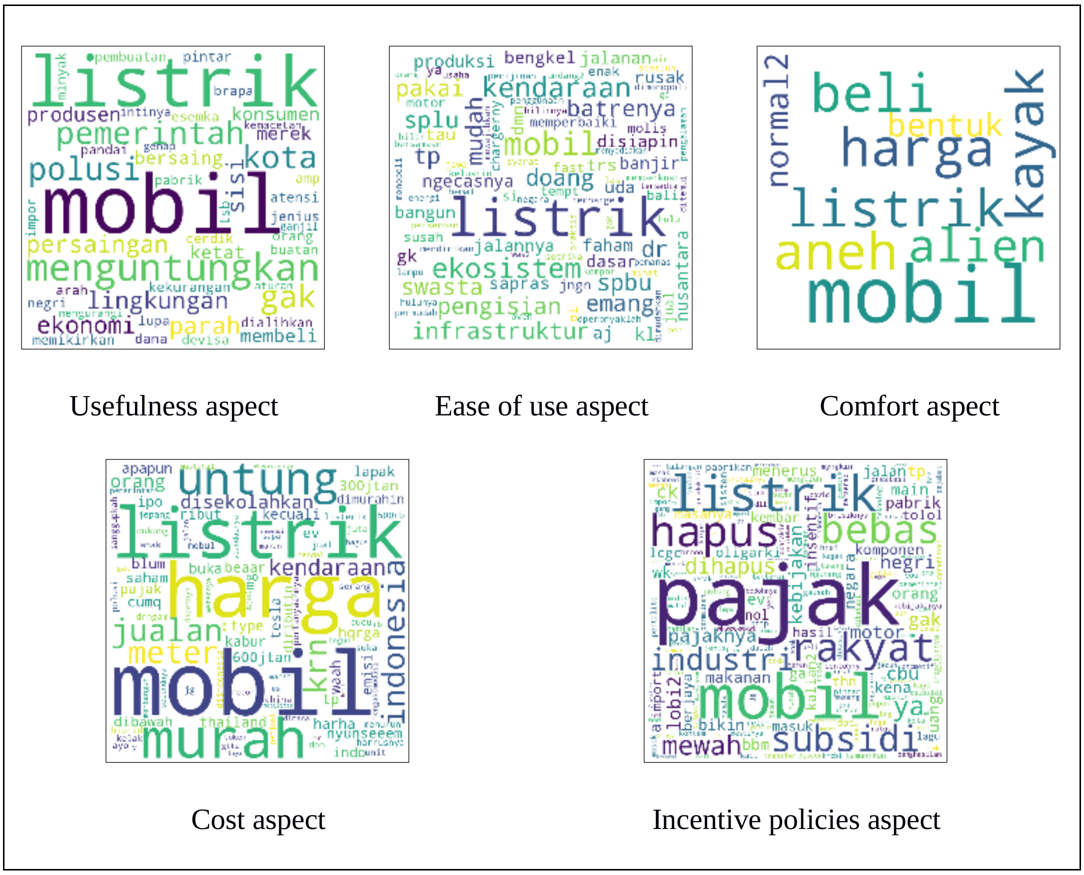

Figure. 13. Word cloud of each aspect of the new dataset

Table 7. Aspect-based sentiment label prediction result evaluation metrics on IndoBERT (combination of aspect and sentence)

Model	Accuracy	Precision	Recall	F1-Score
IndoBERT (combination of aspect and sentence)	**0.9275**	0.5994	0.6166	**0.5878**

4. Conclusion

According to the findings of the research described above, a number of conclusions can be drawn. A dataset of YouTube user comments on transportation electrification has been constructed, although it exhibits imbalanced labels; this study focuses solely on aspect-based sentiment classification. The analysis of the dataset reveals that the sentiment label "none" is the most frequently encountered across all aspects, followed by the "negative" label. The most commented aspects include cost and ease of use, with the predominant language used being Indonesian at 91%, followed by English at 1% and other languages at 8%. For the machine learning models, the hyperparameters identified through grid search include a cost of 10, a gamma of 0.1, and an RBF kernel. Meanwhile, the transfer learning models constructed from pre-trained models utilize the Adam optimization algorithm, with a batch size of 10 and a learning rate of $1 x 10^{-5}$. The top-performing model generated in this research is the fine-tuned IndoBERT model, which combines aspect and comment sentences, achieving an accuracy of 90% and an F1-Score of 60.70%.

Looking ahead, this research offers several suggestions for future studies. While this work focused on five aspects related to transportation electrification—usefulness, ease of use, comfort, cost, and incentive policies—subsequent research could explore additional aspects such as security and performance to help reduce the prevalence of the "none" sentiment label. This expansion could provide a more comprehensive understanding of public opinions and preferences. By addressing these additional factors, researchers may be able to capture a broader range of sentiments, thereby helping to reduce the prevalence of the "none" sentiment label. This would enhance the overall analysis and contribute to more informed discussions about the future of transportation electrification. Additionally, applying text augmentation techniques, such as an oversampling approach using back translation techniques to overcome the imbalance of datasets for each sentiment in each aspect. Furthermore, while this study utilized the $IndoBERT_{Base}$ model, future research might consider larger models like the $IndoBERT_{Large}$. Lastly, developing a dashboard or application could enhance user experience by facilitating aspect-based sentiment classification regarding opinions on transportation electrification.

Ethics approval

Not required.

Acknowledgments

We are very grateful for the support and contributions of all parties involved in this research.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

Credit Authorship

Rahmi Elfa Adilla: Conceptualization, Data Collection, Formal Analysis, Writing-Original Draft, Visualization. **Muhammad Huda**: Methodology, Writing-Review & Editing, Supervision. **Muhammad Aziz:** Writing-Review & Editing, Supervision. **Lya Hulliyyatus Suadaa:** Writing - Review & Editing, Supervision.

References

[1] IEA, "World Outlook," *Econ. Outlook*, vol. 11, no. 4, pp. 1–8, 2016, doi: 10.1111/j.1468-0319.1987.tb00425.x.

[2] H. Turton, "Sustainable global automobile transport in the 21st century: An integrated scenario analysis," *Technol. Forecast. Soc. Change*, vol. 73, no. 6, pp. 607–629, 2006, doi: 10.1016/j.techfore.2005.10.001.

[3] A. Ajanovic, "The future of electric vehicles: Prospects and impediments," *Wiley Interdiscip. Rev. Energy Environ.*, vol. 4, no. 6, pp. 521–536, 2015, doi: 10.1002/wene.160.

[4] IQAir, "Live most polluted major city ranking." Accessed: Aug. 30, 2023. [Online]. Available: https://www.iqair.com/world-air-quality-ranking

[5] A. Vafaei-Zadeh, T. K. Wong, H. Hanifah, A. P. Teoh, and K. Nawaser, "Modelling electric vehicle purchase intention among generation Y consumers in Malaysia," *Res. Transp. Bus. Manag.*, vol. 43, no. January, p. 100784, 2022, doi: 10.1016/j.rtbm.2022.100784.

[6] US Department of Energy, "Alternative Fuels Data Center: Electric Vehicle (EV) Definition." Accessed: Sep. 01, 2023. [Online]. Available: https://afdc.energy.gov/laws/12660

[7] BBC, “Kendaraan listrik disebut ‘solusi palsu’ untuk perbaiki kualitas udara di Indonesia.” [Electric vehicles deemed a 'false solutions' for improving air quality in Indonesia] (in Indonesia). Accessed: Aug. 31, 2023. [Online]. Available: https://www.bbc.com/indonesia/articles/c51qrg47241o

[8] J. Eagle, “The Most Popular Websites by Web Traffic (1993 to 2022).” Accessed: Oct. 26, 2023. [Online]. Available: https://www.visualcapitalist.com/cp/most-popular-websites-by-web-traffic/

[9] R. Gomez, “25 YouTube Stats_ Users, Marketing, Demographics [2023] _ Sprout Social.” Accessed: Oct. 26, 2023. [Online]. Available: https://sproutsocial.com/insights/youtube-stats/

[10] A. U. R. Khan, M. Khan, and M. B. Khan, “Naïve Multi-label Classification of YouTube Comments Using Comparative Opinion Mining,” *Procedia Comput. Sci.*, vol. 82, no. March, pp. 57–64, 2016, doi: 10.1016/j.procs.2016.04.009.

[11] K. M. Kavitha, A. Shetty, B. Abreo, A. D’Souza, and A. Kondana, “Analysis and classification of user comments on YouTube videos,” *Procedia Comput. Sci.*, vol. 177, no. 2018, pp. 593–598, 2020, doi: 10.1016/j.procs.2020.10.084.

[12] D. Jaiswal, V. Kaushal, R. Kant, and P. Kumar Singh, “Consumer adoption intention for electric vehicles: Insights and evidence from Indian sustainable transportation,” *Technol. Forecast. Soc. Change*, vol. 173, no. November 2020, p. 121089, 2021, doi: 10.1016/j.techfore.2021.121089.

[13] Z. Yang, Q. Li, Y. Yan, W. L. Shang, and W. Ochieng, “Examining influence factors of Chinese electric vehicle market demand based on online reviews under moderating effect of subsidy policy,” *Appl. Energy*, vol. 326, no. September, p. 120019, 2022, doi: 10.1016/j.apenergy.2022.120019.

[14] H. Mustakim and S. Priyanta, “Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 16, no. 2, p. 113, 2022, doi: 10.22146/ijccs.68903.

[15] S. Jeong, “Aspect-Level Analysis and Predictive Modeling for Electric Vehicle Based on Aspect-Based Sentiment Analysis Using Machine Learning,” 2020.

[16] M. T. Anwar, D. Trisanto, A. Juniar, and F. A. Sase, “Aspect-based Sentiment Analysis on Car Reviews Using SpaCy Dependency Parsing and VADER,” *Adv. Sustain. Sci. Eng. Technol.*, vol. 5, no. 1, p. 0230109, 2023, doi: 10.26877/asset.v5i1.14897.

[17] R. Jena, “An empirical case study on Indian consumers’ sentiment towards electric vehicles: A big data analytics approach,” *Ind. Mark. Manag.*, no. November 2018, pp. 0–1, 2020, doi: 10.1016/j.indmarman.2019.12.012.

[18] J. Tao and X. Fang, “Toward multi-label sentiment analysis: a transfer learning based approach,” *J. Big Data*, vol. 7, no. 1, pp. 1–26, 2020, doi: 10.1186/s40537-019-0278-0.

[19] N. Liu and J. Zhao, “A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/8726621.

[20] G. Xu, Z. Zhang, T. Zhang, S. Yu, Y. Meng, and S. Chen, “Aspect-level sentiment classification based on attention-BiLSTM model and transfer learning,” *Knowledge-Based Syst.*, vol. 245, p. 108586, 2022, doi: 10.1016/j.knosys.2022.108586.

[21] S. Studer *et al.*, “Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology,” *Mach. Learn. Knowl. Extr.*, vol. 3, no. 2, pp. 392–413, 2021, doi: 10.3390/make3020020.

[22] K. Krippendorff and R. Craggs, “The Reliability of Multi-Valued Coding of Data,” *Commun. Methods Meas.*, vol. 10, no. 4, pp. 181–198, 2016, doi: 10.1080/19312458.2016.1228863.

[23] V. N. Gudivada, D. L. Rao, and A. R. Gudivada, “Information Retrieval: Concepts, Models, and Systems,” in *Handbook of Statistics*, 1st ed., vol. 38, Elsevier B.V., 2018, pp. 331–401. doi: 10.1016/bs.host.2018.07.009.

[24] B. Trstenjak, S. Mikac, and D. Donko, “KNN with TF-IDF based framework for text categorization,” *Procedia Eng.*, vol. 69, pp. 1356–1364, 2014, doi: 10.1016/j.proeng.2014.03.129.

[25] A. Shiri, *Introduction to Modern Information Retrieval (2nd edition)*. Cambridge University Press, 2008. doi: 10.1108/00242530410565256.

[26] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” pp. 843–857, 2020, [Online]. Available: http://arxiv.org/abs/2009.05387

[27] P. Reafeilzadeh, L. Tang, and H. Liu, “Cross Validation,” in *Contemporary Interventional Ultrasonography in Urology*, 2009, pp. 1–6. doi: 10.1007/978-1-84800-217-3_1.