Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANALYSIS OF INDONESIAN PEOPLE'S SENTIMENT TOWARDS ELECTRIC CARS ON SOCIAL MEDIA

MUHAMMAD DAFFA ATTARIQ[1], RIYANTO JAYADI[2]
[1,2] Department of Information Systems Management,
Binus Graduate Program – Master of Information Systems Management,Bina Nusantara University
E-mail: [1]muhammad.attariq@binus.ac.id, [2]riyanto.jayadi@binus.edu

ABSTRACT

One of the steps that can be taken to reduce climate change is to reduce greenhouse gas emissions, which are often caused by the fulfillment of fossil energy. Because fossil energy is energy that will run out in the future, it is necessary to reduce its use. Electric cars are one form of transportation that can reduce the use of fossil energy. However, the presence of electric cars has caused pros and cons that are widely discussed, one of which is on social media Twitter. Based on the many responses, sentiment analysis can be carried out to find out the public's views regarding the presence of electric cars based on data taken from Twitter totaling 22783 tweets from January 2019 to December 2023. Sentiment analysis is carried out to analyze the text of opinions so as to produce information that is positive, neutral, or negative. Therefore, this research aims to analyze public sentiment using LSTM and lexicon based. Based on the results of the study, the highest accuracy was obtained by the LSTM algorithm with an accuracy of 96% with a precision in the negative class of 95%, neutral class 95%, positive class 98%. Recall for negative class is 93%, neutral class is 96%, positive class is 98%. And the f1-score of the negative class is 94%, neutral class 96%, positive class 98%. Meanwhile, the lexicon-based algorithm obtained an accuracy of 37% with precision in negative classes, namely 29%, neutral classes 46%, positive classes 43%. Recall negative class is 75%, neutral class 7%, positive class 54%. And the f1-score of the negative class is 41%, neutral class 13%, positive class 48%. So that the tendency of public sentiment towards electric cars on Twitter social media produces a positive trend.

Keywords: *Sentiment Analysis, Machine Learning, Electric Car, LSTM, Lexicon Based*

1. INTRODUCTION

Climate change is an issue that many countries are discussing seriously nowadays. Damage to various ecosystems, extreme weather changes, natural disasters, are the biggest consequences of climate change. One of the best measures to curb climate change is to reduce greenhouse gas emissions [1]. The increasing greenhouse effect is often caused by the fulfillment of fossil-based primary energy sources that are used as final energy (electricity and transportation) [2]. Fossil energy is non-renewable and will run out in the next few years. The availability of fossil energy is reduced, so it is necessary to convert fossil fuels into electricity [3].

Electric cars are a means of transportation that can reduce the use of fossil energy and remain environmentally friendly, this is because electric cars have no pollution or exhaust emissions [4]. In Indonesia, the presence of electric cars has been supported by the government, this is evidenced by PP Number 55 of 2019 related to electric vehicles issued by the Government. With the issuance of the PP, there are various responses conveyed by the public through social media Twitter. Twitter is one of the most popular social media that acts as a forum for communication in society [5]. The presence of electric cars raises many pros and cons that are discussed by many people, as quoted from the news portal, one of the visitors to IIMS 2023 named Bachtiar argues that "electric cars have an allure because of the low cost of use, can be an alternative to overcome depleted petroleum reserves and motorists no longer need to think about fuel prices because they have switched to electric cars". Meanwhile, another visitor named Tama argued that "the transition to electric cars is still less effective in improving air quality because the coal power plants that supply energy for electric cars also produce pollution". Then from social media Twitter, electric cars are still a hot issue being discussed until now, some think that electric cars are the interests of the government and have not been able to overcome congestion and pollution in Indonesia, especially in DKI Jakarta, such as Twitter user @sekarlangit585 who tweeted "poverty alleviation and health services are far from good and successful, the government is busy thinking about electric cars that are even subsidized", and there are many other responses from Twitter social media users about this electric

car. From the public's responses conveyed through Twitter social media, sentiment analysis can be carried out to find out the responses, or views of the public about the presence of electric vehicles. Sentiment analysis is the process of analyzing text from opinions that contain popularity which will produce data in the form of positive and negative information. There are several methods used in sentiment analysis, namely Long Short Therm Memory (LSTM) and Lexicon Based.

The application of Long Short Therm Memory (LSTM) and Lexicon Based in sentiment analysis has been carried out by several previous studies, including research conducted by Aryal & Bhattarai [6] which discusses sentiment analysis of Covid-19 vaccination tweets using Naïve Bayes and LSTM. The study used data derived from tweets on Twitter in March-April 2021. Based on the results of this study, it shows that the accuracy obtained is LSTM with an accuracy of 84.13% and Naïve Bayes with an accuracy of 77.25%. These results show that the accuracy of LSTM is higher and more efficient. Further related research conducted by Himawan, Putri, Kaswidjanti [7] which discusses the lexicon-based method and SVM in analyzing sentiment on social media as a recommendation for favorite souvenirs. Opinions contained in social media are analyzed by conducting sentiment analysis that can assess the sentiment of the opinion. The data used comes from Twitter and Instagram as much as 1000 data for training data and 50 for test data. The test results show that the greatest accuracy is obtained by using lexicon based which is 88% while using SVM produces an accuracy of 86%.

Based on the problems that occur, this research will analyze sentiments related to public opinion on electric cars using a comparison of the Long Short Term Memory (LSTM) and Lexicon Based methods. The use of LSTM and Lexicon Based algorithms in this study is because both algorithms have better performance and higher accuracy than previous studies. From this research, it is hoped that the general public can find out the results of sentiment towards the presence of electric cars and can find out the best algorithm that can be used for sentiment analysis.

2. LITERATURE REVIEW

Literature review refers to several previous studies related to sentiment analysis. The first research conducted by Alayba and Palade [8] discusses the utilization of Arabic language classification using enhanced CNN-LSTM and effective Arabic text preparation. This research aims to take an approach by combining CNN and LSTM to improve sentiment classification, by excluding the max-pooling layer from CNN. This layer can reduce the length of the feature vectors generated after combining the input data filters. Thus, the LSTM network will receive the captured vectors from the feature map. The results show that the highest result is obtained on the Main-AHS dataset of 0.9483 by using Farasa Lemmatisation as word normalization. Further related research was conducted by Kusumaningrum and Wibowo [9] which discussed sentiment analysis using Word2vec and LSTM for Indonesian hotel reviews. The availability of many reviews owned by online travel agents related to the facilities used by customers, causes problems in knowing the percentage of reviews that have an influence on the services provided. Therefore, the research aims to conduct sentiment analysis using LSTM with the Word2vec model. The parameter combination for Word2vec is Skip-gram as the architecture, Hierarchiccal Softmax as the evaluation method, and 300 as the vector dimension. While the LSTM parameter combination is dropout value of 0.2, pooling type as average polling, and learning rate of 0.001. The results showed that an accuracy of 85.86% was obtained. Further related research was conducted by Mahadevaswamy and P Swathi [10] who discussed sentiment analysis using BiLSTM using Amazon product review data. The analysis was carried out on product reviews that describe customer ratings on mobile electronic products. Reviews are classified into two categories, namely positive and negative. Based on the test results in this study, it shows that the best accuracy is 91.4%. Further related research was conducted by Hernandez, Ojeda-Hernandez, Lopez-Rodriguez, and Mora [11] who discussed lexicon-based sentiment analysis in text using Formal Concept Analysis (FCA) to create a dictionary for classification. The dataset used is a collection of tweets that will be categorized into positive and negative polarity. The results show that the proposed dictionary has a better overall performance in AUC value than the standard dictionary and other standards used in the study. FCA can be a dictionary in detecting tweet polarity, while being efficient in terms of computational time. Further related research was conducted by Bhowmik, Arifuzzaman, and Mondal [12] who discussed sentiment analysis on Bangla text using lexicon dictionaries and deep learning algorithms. The results showed that the proposed LSTM model was very accurate in performing sentiment analysis with the best accuracy of 84.18%. Further related research was conducted by Pradhan, Senapati, and Sahu [13]

which discussed the improvement of sentiment analysis by learning concepts from concepts, lexicon patterns, and negation. The research uses the Latent Dirichlet Allocation (LDA) and Probabilistic Latent Sematic Analysis (PLSA) algorithms. The results show that the accuracy obtained on the SemEval2014 dataset on restaurant data is 84.73% with an f1 value of 81.28%. Similarly, for the SemEval2014 dataset on laptop data, the accuracy is 82.06% and the f1 value is 80.71%.

Based on the literature review, it is found that LSTM and Lexicon Based have good performance and make a reference for the use of these two algorithms in this study, but there are differences between this research and previous research, namely in this study a comparison of performance between LSTM and Lexicon Based which has not been done in previous studies. As well as other differences are on the topic of sentiment analysis regarding the presence of electric cars.

3. METHODOLOGY

In conducting research, researchers create a framework in the form of stages in solving a problem. The framework in conducting sentiment analysis using lexicon based and LSTM in this study is made in the form of a flowchart as in Figure 1.

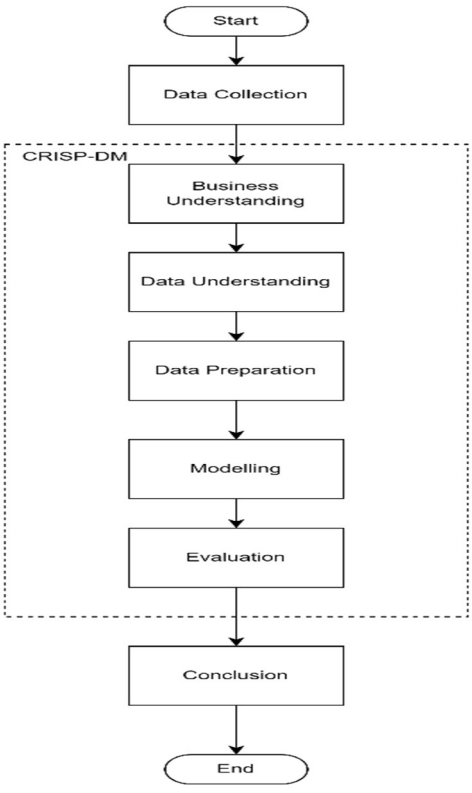

Figure 1 Research Methodology

A. Data Collection

In this study, researchers collected two types of data, namely primary data and secondary data. Primary data in this study is in the form of data collection from Twitter which contains tweets of public opinion regarding electric cars. The data used in this study amounted to 22784 Indonesian tweets taken from January 2019 to December 2023. Meanwhile, secondary data in this study is in the form of literature studies of previous research journals related to this research.

B. CRISP-DM Method

In carrying out the implementation, researchers used CRISP-DM. CRISP-DM is a method that uses a data development process model that is widely used by experts to solve problems [14].

1. Business Understanding
   Electric cars are transportation that has no pollution or exhaust emissions. Currently, the presence of electric cars has been supported by the government, this is evidenced by PP Number 55 of 2019 related to electric vehicles issued by the Government. With the issuance of the PP, there are various responses submitted by the public through social media Twitter. From the public responses conveyed through Twitter social media, a sentiment analysis can be carried out on the public's responses. The business purpose of data processing in research is to find out the responses, or views of the public about the presence of electric vehicles so that it can be seen in which direction the tendency of public responses or views is positive, neutral or even negative.

2. Data Understanding
   This stage is the process of collecting initial data in the form of excel documents in .csv format obtained by scraping from twiter social media, where the data obtained amounted to 22784 tweets taken from January 2019 to December 2023, then analyzing the data and evaluating the quality of the data used in the study.

3. Data Preparation
   The next way to prepare data is that the data that has been collected is then labeled and entered into the preprocessing stage. The labeling process is a stage to determine the response of tweets in the dataset which is done by creating a program based on the corpus of words to determine the type of tweet to be a

positive review, neutral review, or negative review. The labeling aims to provide classification of the tweets obtained. Furthermore, preprocessing in this study is case folding, tokenizing, stopwords, and stemming. The purpose of data pre-processing is to overcome various problems in the data such as noisy data, data redundancy, missing data values, and others. Pre-processing is a stage where the data obtained is collected into one document for further analysis [15]. The steps in doing preprocessing are as follows:

1. Case Folding
   Case folding is the process of converting all existing letters into lowercase letters [16]. Examples of case folding can be seen in the table 1.

Table 1 Result of Case Folding

Before	After
RT @barikade_98: Dukung Penuh Penggunaan Kendaraan Dinas Listrik	rt @barikade_98 dukung penuh penggunaan kendaraan dinas listrik

2. Tokenizing
   Tokenizing is the process of breaking the word into several parts. The results of tokenizing are also used to remove punctuation marks that will not be used in preprocessing [17]. Examples of tokenizing can be seen in the table 2.

Table 2 Result of Tokenizing

Before	After
dukung penuh penggunaan kendaraan dinas listrik	‘dukung’, ‘penuh’, ‘penggunaan’, ‘kendaraan’, ‘dinas’, ‘listrik’

3. Stopwords
   Stopwords namely the removal of words that are not important or not needed in the form of adverbs and conjunctions. Examples of stopwords can be seen in the table 3.

Table 3 Result of Stopwords

Before	After
dukung penuh penggunaan kendaraan dinas listrik	‘dukung’, ‘penuh’, ‘kendaraan’, ‘listrik’

4. Stemming
   stemming is the process of converting a word into its original form or base word [18]. Examples of stemming can be seen in the table 4.

Table 4 Result of Stemming

Before	After
dukung penuh penggunaan kendaraan dinas listrik	‘dukung’, ‘penuh’, ‘guna’, ‘kendara’, ‘dinas’, ‘listrik’

Then the data division stage is carried out, where the data is divided into two, namely training data and testing data. The testing data will be grouped into 3 classes to be categorized according to the type of positive sentiment, negative sentiment, or neutral sentiment. The use of a split data ratio of 80:20 is because the larger the training data, the more it can represent the overall data set with different characteristics [19].

4. Modelling
   The research model used in this research is a combination of descriptive and analytical research models. Descriptive research is used because the data collected in the form of words or opinions of the Indonesian people regarding electric cars obtained from Twitter and then this research is studied analytically. Meanwhile, the analytical method explains the analysis method used to solve research problems after data is obtained. The analysis method in this study uses the Long Short-Term Memory (LSTM) sentiment analysis method and the approach method using Lexicon Based. The analysis method or classification method in the LSTM method is used after obtaining training data and testing data. While the classification in the Lexicon Based method does not require dataset training to find sentiment polarity [20]. The classification process with Lexicon Based is done by

validating the words contained in the dataset with the words in the lexicon dictionary that has been prepared previously. After the classification process is carried out, tweets will appear along with the results of the sentiment category, be it positive sentiment, neutral sentiment, or negative sentiment. The stages of algorithm implementation in this research can be seen in Figure 2.

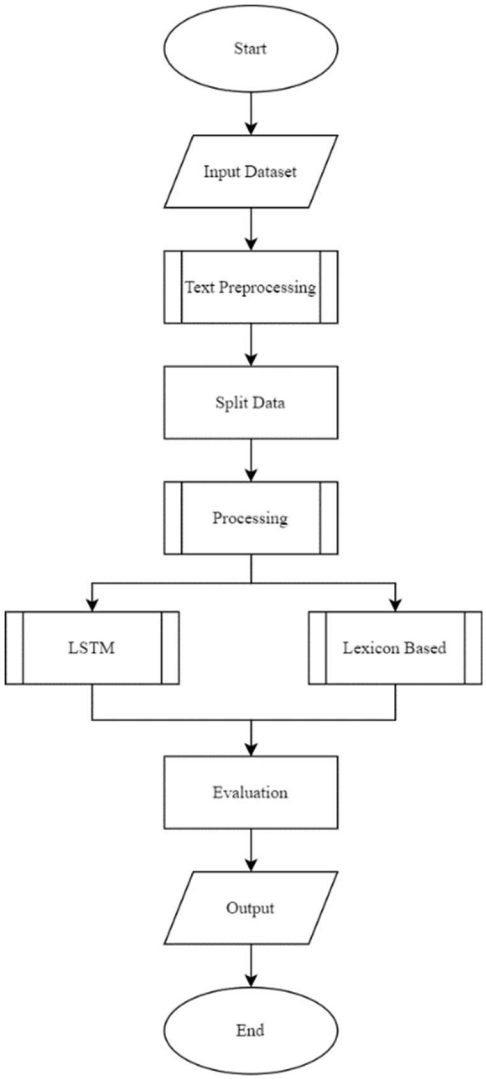

Figure 2 Stages of Algorithm Implementation

5. Evaluation

At this stage, an evaluation is carried out to test the performance of the model using accuracy and error rate calculations. Calculation of accuracy and error rate using confusion matrix method [21]. Confusion matrix is a table that states the classification of the number of correct test data and the number of incorrect test data [22].. From the results of the Confusion Matrix evaluation, it can be concluded regarding the performance of the use of the Lexicon Based and LSTM algorithms from the Sentiment Analysis of Indonesian People Towards Electric Cars on Social Media.

C. Conclusion

After doing all the research as a whole, the last stage is drawing conclusions, where this conclusion is the answer to the problem that has been formulated.

4. RESULTS AND DISCUSSION

A. Results

At this stage, the results of the LSTM and Lexicon Based methods are implemented into the Python programming language. The first thing to do is to import the Python library that will be used and prepare the dataset that will be used. The dataset used can be seen in Figure 3.

	No	Username	date	tweet
0	1.0	Fadhilah Rusmaputeri	1/22/2023	Dari pada subsidi mobil listrik mending subsid...
1	2.0	Ikrar Agung	1/22/2023	Kemarin minta mobil listrik, sekarang minta ba...
2	3.0	Murni Mumuñ	1/22/2023	Kayaknya akar masalah ini udah kelihatan dari ...
3	4.0	Latifah Munawaroh	1/22/2023	Tukang mobil listrik bukannya udah turun tahta...
4	5.0	Mai	1/22/2023	Yepp... plis dah Indo aja sumber listriknya ba...
...	...	...	...	...
22778	NaN	Bams Soesilo	25-Aug	Katanya udh 5G trus mobil listrik bahkan mobil...
22779	NaN	Pero	25-Aug	Lagi gak aktif kemarin wkwk disuruh nya pake m...
22780	NaN	kohryan.eth	25-Aug	kalau itu mobil listrik mungkin masih oke, emi...
22781	NaN	bakanosan1	25-Aug	Cara biar laku kendaraan listrik, padahal dulu...
22782	NaN	Achmad Royandy	25-Aug	Society 5.0 juga membawa implikasi bagi indivi...

Figure 3 Dataset

Furthermore, the following steps are carried out:

1. Labelling

At this stage, data labeling is carried out by determining the type of tweet including positive, neutral, or negative based on the word corpus.

2. Preprocessing

At this stage, preprocessing is carried out to prepare the data so that it becomes a structured and easy to understand format. Preprocessing is done by going through the stages of case folding, tokenizing, stopwords, and stemming. The library used in preprocessing is nltk which is a stopwords dictionary for Indonesian and English [23]. Library nltk.tokenize imported word_tokenize which is used in the tokenizing process. Library Sastrawi imported

Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

StemmerFactory which is used in the stemming process [24]. As well as the re library used in the case folding process to replace the matching characters in the search pattern with the specified string. Preprocessing is carried out with the results of preprocessing which can be seen in Figure 4.

	username	tweet	casefolding	tokenizing	filtering	stem	label
0	Fadhilah Rusmaputeri	Dari pada subsidi mobil listrik mending subsidi...	dari pada subsidi mobil listrik mending subsidi...	[dari, pada, subsidi, mobil, listrik, mending,...	subsidi mobil listrik mending subsidi seminarh...	subsidi mobil listrik mending subsidi seminarh...	Positive
1	Ikrar Agung	Kemarin minta mobil listrik, sekarang minta ba...	kemarin minta mobil listrik sekarang minta bar...	[kemarin, minta, mobil, listrik, sekarang, min...	kemarin mobil listrik barongsai	kemarin mobil listrik barongsai	Neutral
2	Murni Mumuh	Kayaknya akar masalah ini udah kelihatan dari ...	kayaknya akar masalah ini udah kelihatan dari ...	[kayaknya, akar, masalah, ini, sudah, kelihata...	kayaknya akar wacana pengurangan emisi dipisah...	kayak akar wacana kurang emisi pisah mobil lis...	Negative
3	Latifah Munawaroh	Tukang mobil listrik bukannya udah turun tanta...	tukang mobil listrik bukannya udah turun tanta...	[tukang, mobil, listrik, bukannya, sudah, turu...	tukang mobil listrik turun tanta ketemu pengga...	tukang mobil listrik turun tanta ketemu ganti born	Negative
4	Mai	Yepp... pls dah indo aja sumber listriknya ba...	yepp pls dah indo aja sumber listriknya banya...	[yepp, mohon, sudah, indo, saja, sumber, listr...	yepp mohon indo sumber listriknya pakai batu b...	yepp mohon indo sumber listrik pakai batu bara...	Negative
...	...	...	...	...	...	...	...
18051	tiyas mavarti	Solusi terbaik ganti mobil listrik..	solusi terbaik ganti mobil listrik	[solusi, terbaik, ganti, mobil, listrik]	solusi terbaik ganti mobil listrik	solusi baik ganti mobil listrik	Positive
18052	Edison	Cair cair mobil listrik.inflasia macet tambah p...	cair cair mobil listrikinnajia macet tambah parah	[cair, cair, mobil, listrik, rajia, macet, tam...	cair cair mobil listrik rajia macet parah	cair cair mobil listrik rajia macet parah	Negative
18053	IRFAN 6666	Itu masalah teknis... Beda dg konsep. Apa prog...	itu masalah teknis beda dg konsep apa programm...	[itu, masalah, teknis, beda, konsep, apa, prog...	teknis beda konsep programnya langsung terlaks...	teknis beda konsep program langsung laksana ha...	Neutral
18054	my Name is ~	Tp mobil listrik hidrogen boros energi. Dr sul...	tp mobil listrik hidrogen boros energi dr sul...	[tetapi, mobil, listrik, hidrogen, boros, ener...	mobil listrik hidrogen boros energi suling air...	mobil listrik hidrogen boros energi suling air...	Positive
18055	Dodolmeong	pemanasan global di indonesia yg berkontribusi...	pemanasan global di indonesia yg berkontribusi...	[pemanasan, global, indonesia, yang, berkontri...	pemanasan global indonesia berkontribusi indus...	panas global indonesia kontribusi industri def...	Neutral

Figure 4 Preprocessing Result

2. Split Data

At the split data stage, it is done by dividing the data into train data and test data. Split data is done by separating data based on negative, neutral, and positive labels and then recombining them into train data and test data. The data is divided into 80% train data and 20% test data, with commands as in Figure 5.

```
test_size = 0.2
rs        = 42
```

Figure 5 Split Data

In the Lexicon Based method, the data split process is carried out until the data is combined into train data and test data. Whereas in the LSTM method, the data split process is continued by changing the label to 0,1,2 using the enccoder label. After that, vectorization is done using a tokenizer. Then the conversion is made to a 2D Numpy array using pad¬_sequence and converted into a sequence of integers using texts_to_sequence. Labels will be converted into categorical form using to_categorical. The number of train data is negative 2485, neutral 6589, and positive 5353labels. While the amount of test data is labeled negative 622, neutral 1648, and positive 1339.

3. Implementation of LSTM Method

The LSTM architecture is created with commands as shown in Figure 4. The LSTM model uses 128 layers with tanh activation function. And the dense layer uses the number of layers 32 with the relu activation function. as for the lstm architecture used can be seen in the figure 6.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 36, 64)            640064

 spatial_dropout1d (SpatialD  (None, 36, 64)           0
 ropout1D)

 lstm (LSTM)                 (None, 36, 128)           98816

 dropout (Dropout)           (None, 36, 128)           0

 lstm_1 (LSTM)               (None, 128)               131584

 batch_normalization (BatchN  (None, 128)              512
 ormalization)

 dropout_1 (Dropout)         (None, 128)               0

 dense (Dense)               (None, 32)                4128

 dense_1 (Dense)             (None, 3)                 99

=================================================================
Total params: 875,203
Trainable params: 874,947
Non-trainable params: 256
_________________________________________________________________
```

Figure 6 LSTM Architecture

Next, the LSTM model will be trained with train data and then evaluated using test data. The model is first compiled with categorical crossentropy loss, adam optimizer parameters, and accuracy metrics. Next, the model training is carried out with the batch size (batch_size) used as 32 with an epoch of 100. Epoch stops automatically at epoch 29 because it uses the early stopping function. The following Figure 7 displays a plot between the results of training data accuracy and validation data. The best accuracy is obtained at epoch 19 with the resulting accuracy of 0.9856 on training data and 0.7690 on validation data. The relationship contained in the accuracy value displays a positive relationship, namely the more the number of epochs used, of course, the higher the accuracy value on training data and validation data [25].

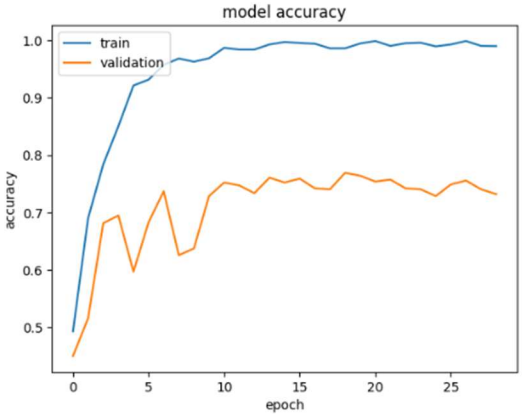

Figure 7 Plot Result Accuracy

The following Figure 8 displays a plot between the loss results of training data and validation data. The best loss value is obtained at epoch 19 with the resulting loss of 0.0474 in training data and 1.6075 in validation data. The relationship between the number of epochs and the loss value is a negative relationship, namely the greater the number of epochs used, the smaller the loss value issued on the training data. So to minimize the expected loss value, it can be done by increasing the number of epochs in the training process [25].

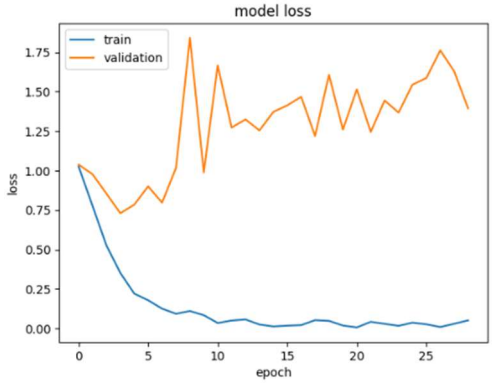

Figure 8 Loss Plot Result

Next, load the best model and predict the test data and take the highest value with np.argmax.

4. Implementation of Lexicon Based Method
   In the lexicon-based model, a lexicon dictionary, InSet (Indonesia Sentiment Lexicon), is used. Polarity determination is done based on the score obtained. If the score is more than 0 then the polarity is positive, the score is less than 0 then the polarity is negative, and the score is equal to 0 then neutral. The lexicon-based model process shows the results of the sentiment polarity obtained, namely negative labels as much as 2019 data, neutral labels as much as 1975 data, and positive labels as much as 308.

5. Evaluation
   1) Evaluation of the LSTM Method
      At this stage, the performance of the LSTM model obtained using testing data is evaluated. Model evaluation is done using a confusion matrix. The following in Figure 8 is a display of the LSTM confusion matrix.

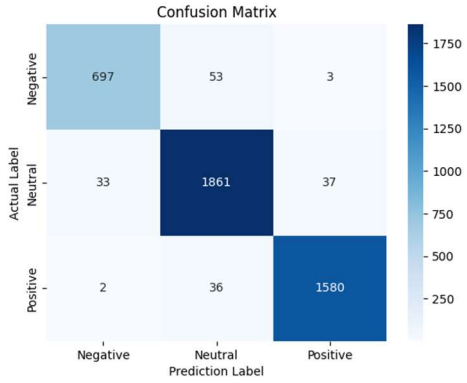

Figure 9 Confusion Matrix View

Based on Figure 9, it can be explained that the LSTM algorithm obtained prediction results on correctly classified negative labels as much as 697 data, while prediction errors with 53 data entered into neutral labels and 3 data entered into positive labels. On the neutral label with a total of 1861 data correctly classified, while the prediction error with 33 data entering the negative label and 37 data entering the positive label. On positive labels with a total of 1580 data classified correctly, while prediction errors with 2 data entering the negative label and 36 data entering the neutral label.

Next, the process is carried out to display the results of the LSTM classification report as in Figure 10.

	precision	recall	f1-score	support
0	0.95	0.93	0.94	753
1	0.95	0.96	0.96	1931
2	0.98	0.98	0.98	1618
accuracy			0.96	4302
macro avg	0.96	0.96	0.96	4302
weighted avg	0.96	0.96	0.96	4302

Figure 10 LSTM Result Classification Report

Based on Figure 10, the results obtained are 96% accuracy with precision in negative classes, namely 95%, neutral classes 95%, positive classes 98%. Recall negative class is 93%, neutral class 96%, positive class 98%. And the f1-score of the negative class is 94%, the neutral class is 96%, the positive class is 98%. Furthermore, the evaluation of negative, neutral, and positive labels is carried out.

	Word	Total
0	mobil	26716
1	listrik	25936
2	pakai	2571
3	beli	2492
4	subsidi	2389
...	...	...
24257	jwbny	1
24258	jwab	1
24259	juwalan	1
24260	jutsu	1
24261	zzzzzzzz	1

Figure 11 Frequency of Word Occurrence

Next is done to display the wordcloud into the image. The following in Figure 12 is a view of the wordcloud obtained.

Figure 12 Wordcloud LSTM

2) Evaluation of the Lexicon Based

At this stage, the performance of the Lexicon Based model obtained using testing data is evaluated. Model evaluation is done using confusion matrix. The following in Figure 13 is a display of the Lexicon Based confusion matrix.

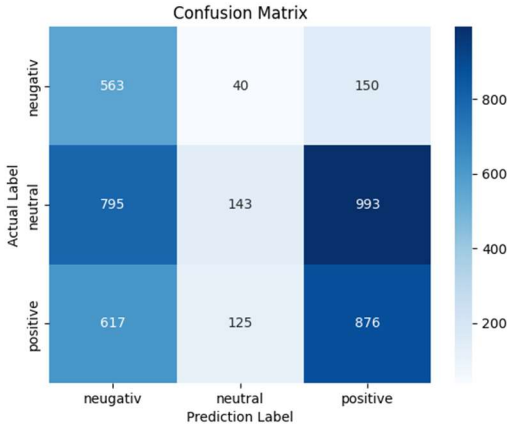

Figure 13 Confusion Matrix View

Based on Figure 13, it can be explained that the Lexicon-based algorithm obtained prediction results on correctly classified negative labels as much as 563 data, while prediction errors with 40 data entered into neutral labels and 150 data entered into positive labels. On the neutral label with a total of 143 data correctly classified, while the prediction error with 963 data entering the positive label and 795 data entering the negative label. On the positive label with a total of 876 data classified correctly, while the prediction error with 125 data entering the neutral label and 617 data entering the negative label.

Next, the process is carried out to display the results of the Lexicon Based classification report as shown in Figure 14.

	precision	recall	f1-score	support
Negative	0.27	0.73	0.40	622
Neutral	0.47	0.07	0.13	1648
Positive	0.41	0.53	0.46	1339
accuracy			0.35	3609
macro avg	0.39	0.44	0.33	3609
weighted avg	0.42	0.35	0.30	3609

Figure 14 Lexicon Based Classification Report Result

Based on Figure 14, the results obtained are accuracy 35% with precision in the negative class which is 27%, neutral class 47%, positive class 41%. Recall negative class is 73%, neutral class 7%, positive class 53%. And the f1-score of the negative class is 40%, neutral class 13%, positive class 46%.

Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

After that, the word grouping is done. The following in Figure 15 is the frequency of occurrence of the words obtained.

	Word	Total
0	mobil	26716
1	listrik	25936
2	pakai	2571
3	beli	2492
4	subsidi	2389
...	...	...
24257	jwbny	1
24258	jwab	1
24259	juwalan	1
24260	jutsu	1
24261	zzzzzzzz	1

Figure 15 Frequency of Word Occurrence

Based on Figure 15, it can be seen that the highest number of words is a car with a total of 26716 words, followed by the word electricity with a total of 25936. This is in accordance with the research conducted, namely regarding sentiment towards the presence of electric cars. So that the data contained in the study contains a lot of electric car words.

Based on Figure 15, it can be seen that the highest number of words is a car with a total of 26716 words, followed by the word electricity with a total of 25936. This is in accordance with the research conducted, namely regarding sentiment towards the presence of electric cars. So that the data contained in the study contains a lot of electric car words.

Next is done to display the wordcloud into the image. The following in Figure 16 is the wordcloud display obtained.

Figure 16 Wordcloud LSTM

B. Discussion

Based on the results of research that has been carried out in implementing the Long Short-Term Memory (LSTM) and Lexicon Based algorithms in analyzing public sentiment towards electric cars on Twitter social media using the Python programming language with Google Colab tools. The dataset used is data taken from Twitter containing public opinion tweets regarding electric cars totaling 22784 tweets in Indonesian taken from January 2019 to December 2023.

The dataset that has been collected is then labeled on the data and then entered into the preprocessing stage including case folding, tokenizing, stopwords, and stemming. Then the dataset is divided into 80% train data and 20% test data. The data that has been divided is then implemented using the Long Short-Term Memory (LSTM) and Lexicon Based algorithm. The results obtained by evaluating using confusion matrix are the highest accuracy obtained in the LSTM algorithm with an accuracy of 96% with precision in negative classes, namely 95%, neutral classes 95%, positive classes 98%. Recall negative class is 93%, neutral class 96%, positive class 98%. And the f1-score of the negative class is 94%, the neutral class is 96%, the positive class is 98%. While the Lexicon Based algorithm obtained 35% accuracy with precision in negative classes, namely 2%, neutral classes 47%, positive classes 41%. Recall negative class is 73%, neutral class 7%, positive class 53%. And the f1-score of negative classes is 40%, neutral classes are 13%, positive classes are 46%. In addition, the wordcloud obtained in each of the LSTM and Lexicon Based methods produced with the word that has the highest frequency of occurrence is a car with a total of 26716 occurrences of the word, in second position there is electricity with a total of 25936, and in third position there is a use with a total of 2571. The results of the confusion matrix can be seen in Table 1. So that the results of the tendency of public sentiment towards electric cars on Twitter social media produce positive trends in the LSTM algorithm and in the Lexicon Based algorithm.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5 Confusion Matrix Result

	Class	Confusion Matrix			
		Accuracy	***Precision***	***Recall***	***F1-Score***
LSTM	Negative	96%	95%	93%	94 %
	Neutral		95%	96%	96 %
	Positive		98%	98%	98 %
Lexicon Based	Negative	35%	27%	73%	40 %
	Neutral		47%	7%	13 %
	Positive		41%	53%	46 %

5. CONCLUSION

This study aims to use sentiment analysis as a tool to determine public sentiment towards the presence of electric cars in Indonesia. This research uses Twitter as the main data source by scrapping tweets related to public opinion regarding electric cars. The data taken was 22783 tweets taken from January 2019 to December 2023. After going through several stages using the CRISP-DM method, it was found that the model that had the best performance was LSTM with an accuracy of 96% with a precision in the negative class of 95%, neutral class 95%, positive class 98%. Recall for negative class is 93%, neutral class 96%, positive class 98%. And the f1-score of the negative class is 94%, neutral class 96%, positive class 98%. Meanwhile, the lexicon-based algorithm obtained an accuracy of 37% with precision in negative classes, namely 29%, neutral classes 46%, positive classes 43%. Recall negative class is 75%, neutral class 7%, positive class 54%. And the f1-score of the negative class is 41%, neutral class 13%, positive class 48%. So that the tendency of public sentiment towards electric cars on Twitter social media produces a positive trend using the LSTM algorithm and the Lexicon Based algorithm. The suggestions that can be given in the development of further research are to use data sources from other social media such as Facebook or Instagram so that they can provide a more accurate picture. In addition, because this research was developed using the LSTM algorithm which is a deep learning algorithm and Lexicon Based, this research can be developed using machine learning algorithms such as Support Vector Machine (SVM) or other methods so that it can be seen which is the best method in analyzing the sentiment of the Indonesian people towards electric cars on Twitter social media.

REFERENCES:

[1] A. F. Riyadi, F. R. Rahman, M. A. Nofa Pratama, M. K. Khafidli dan H. Patria, "Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia," *JURNAL MANAJEMEN SISTEM INFORMASI DAN TEKNOLOGI,* pp. 141-149, 2021.

[2] A. S. Pramudiyanto dan S. W. Agung Suedy, "Energi Bersih dan Ramah Lingkungan dari Biomassa untuk Mengurangi Efek Gas Rumah Kaca dan Perubahan Iklim yang Ekstrim," *JEBT: Jurnal Energi Baru & Terbarukan,* pp. 86-99, 2020.

[3] D. R. Putra, D. Yoesgiantoro dan S. Thamrin, "Kebijakan Ketahanan Energi Berbasis Energi Listrik Pada Bidang Transportasi Guna Mendukung Pertahanan Negara Di Indonesia: Sebuah Kerangka Konseptual," *Jurnal Ilmu Pengetahuan Sosial,* pp. 658-672, 2020.

[4] A. Santoso, A. Nugroho dan A. S. Sunge, "Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization," *Journal of Pratical Computer Science,* pp. 24-31, 2022.

[5] M. I. Fikri, T. S. Sabrila dan Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machinepada Analisis Sentimen Twitter," *SMATIKA Jurnal,* pp. 71-76, 2020.

[6] R. R. Aryal dan A. Bhattarai, "Sentiment Analysis on Covid-19 Vaccination Tweets using Naïve Bayes and LSTM," *Advances in Engineering and Technology: An International Journal,* pp. 57-70, 2021.

[7] H. Himawan, D. G. Putri D D R dan W. Kaswidjanti, "Long Short Term Memory (LSTM) dan Lexicon Based Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit," *Seminar Nasional Informatika 2018,* pp. 235-244, 2018.

[8] A. M. Alayba dan v. Palade, "Leveraging Arabic sentiment classification using an enhanced CNN-LSTM approach and effective Arabic text preparation," *Journal of King Saud University,* pp. 9710-9722, 2021.

[9] P. F. Muhammad, R. Kusumaningrum dan A. Wibowo, "Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews," *Procedia Computer Science,* pp. 728-735, 2021.

[10] U. B. Mahadevaswamy dan S. P, "Sentiment Analysis using Bidirectional LSTM Network," *Procedia Computer Science,* pp. 46-56, 2023.

[11] Hernandez, M. Ojeda-Hernandez, D. Lopez-Rodriguez dan A. Mora, "Lexicon-based sentiment analysis in texts using Formal," *International Journal of Approximate Reasoning,* pp. 104-112, 2023.

[12] N. R. Bhowmik, M. Arifuzzaman dan M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithm," *Array,* pp. 1-14, 2022.

[13] A. Pradhan, M. R. Senapati dan P. K. Sahu, "Improving sentiment analysis with learning concepts from concept, patterns lexicons and negations," *Ain Shams Engineering Journal,* pp. 1-13, 2022.

[14] D. Astuti, A. R. Iskandar dan A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *Journal of Informatics, Information System, Software Engineering and Applications,* pp. 60-72, 2019.

[15] Y. Pratama, D. T. Murdiansyah dan K. M. Lhaksmana, "Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis," *JURNAL MEDIA INFORMATIKA BUDIDARMA,* pp. 529-535, 2023.

[16] Herwinsyah dan A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Sistem Informasi dan Informatika (Simika),* vol. 5, pp. 59-68, 2022.

[17] P. P. O. Muhawardana, G. A. Sasmita dan I. P. A. E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap "Figure Pemimpin" Menggunakan Python," *JITTER- Jurnal Ilmiah Teknologi dan Komputer,* vol. 3, 2022.

[18] S. Lestari dan S. Saepudin, "Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes," *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika),* pp. 163-171, 2021.

[19] R. D. Fitriani, H. Yasin dan Tarno, "Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal)," *JURNAL GAUSSIAN,* vol. 10, no. 1, pp. 11-20, 2021.

[20] Y. N. Prasetya, D. Winarso dan Syahril, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," *JURNAL FASILKOM,* pp. 97-103, 2021.

[21] N. M. A. J. Astari, D. G. H. Divayana dan G. Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *JURNAL SISTEM DAN INFORMATIKA (JSI),* pp. 22-28, 2020.

[22] H. F. Fadli dan A. F. Hidayatullah, "Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM," *Jurnal AUTOMATA,* vol. 2, no. 1, 2021.

[23] D. Arianto dan I. Budi, "Analisis Sentimen Berbasis Aspek dan Pemodelan Topik pada Candi Borobudur dan Candi Prambanan," *JURNAL MULTINETICS ,* pp. 141-150, 2022.

[24] Y. Marwanti dan Badiyanto, "Analisis Sentimen Pencitraan Perguruan Tinggi di Yogyakarta Menggunakan Metode Naïve Bayes Classifier," *Journal of Applied Informatics and Computing (JAIC),* pp. 15-21, 2023.

[25] A. Sholihin, D. I. Mulyana dan M. B. Yel, "Klasifikasi Alat Musik Tradisional Papua menggunakan Metode Transfer LearningDan Data Augmentasi," *Jurnal Sistem Komputer dan Kecerdasan Buatan,* pp. 36-44, 2022.