Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

Analysis Of Opinion Sentiment Towards Electric Vehicle Tax On Social Media X Using The Support Vector Machine Method

Dara Taqa Assajidah Jusli[1)*], Rakhmat Kurniawan[2)]
[1)2)3)]Computer Science Study Program, State Islamic University of North Sumatra, Indonesia
[1)]darataqa71@gmail.com, [2)]rakhmat.kr@uinsu.ac.id

ABSTRACT

Electric vehicle tax is increasingly becoming an important issue related to environmental and fiscal policies. Electric vehicles are considered an environmentally friendly solution to reduce greenhouse gas emissions and dependence on fossil fuels. However, public perception of electric vehicle tax is still mixed. This study aims to analyze public sentiment about electric vehicle tax based on data from social media platform X, using the Support Vector Machine (SVM) method. The data used was taken through a crawling technique with a total of 1,014 valid data. The data was then classified into positive and negative classes with a transformer. In this analysis, the data was divided with a ratio of 8:2 between training data and test data. 811 were used as training data and 203 as test data. The research stages involved data preprocessing, sentiment labeling, data separation into training and test data, and weighting using TF-IDF. After that, SVM was applied to classify tweets into positive and negative sentiments. The test results showed that the SVM algorithm had an accuracy of 79%, precision of 85%, recall of 89%, and F1-score of 87%. Based on the results of this research, some people feel unsure about the government's policy regarding electric vehicle tax, because it is considered unfair to the lower middle class. Electric vehicles are considered more expensive than fuel-powered vehicles, so this policy is considered unprofitable.

Keywords: Confussion Matrix; Electric Vehicle Tax; Sentiment Analysis; Support Vector Machine (SVM); X

1. INTRODUCTION

Electric vehicle tax is one of the issues that is increasingly emerging in the context of fiscal policy and environmental protection. Along with technological developments, electric vehicles are an attractive alternative to reduce greenhouse gas emissions and dependence on fossil fuels due to the impact of engine vehicles on the environment (Ramadhina and Najicha, 2022). Over time, vehicles have developed into one of the key elements in opening up professional opportunities, both conventional and modern.

With the ever-evolving advancement of technology, there is a shift from the use of fuel oil to electric power that can help stabilize CO2 levels in the atmosphere (Prianjani & Sutopo, 2018). Although electric vehicles are still not many, the increase in sales of these vehicles is projected to continue to increase as time goes by. Data from the Indonesian Automotive Industry Association (GAIKINDO) in 2023 shows that throughout 2023 until the middle of the year, more than 5,000 units of electric cars have been sold. Thus, the number of electric vehicles in Indonesia is expected to continue to grow over time. The advantages of electric vehicle technology lie in their environmentally friendly characteristics, having a quiet engine, and more economical operating costs thanks to the use of batteries. The transformation from fuel oil to electric power provides additional advantages to this technology, with longer durability and higher cost efficiency in the electric vehicle industry in Indonesia (Veza et al., 2022).

The government has issued Government Regulation (PP) Number 55 of 2019 concerning Electric Vehicles, which aims to accelerate the battery-based electric vehicle program for road transportation. The issuance of this PP triggered various discussions and responses from the public, especially on social media such as Twitter or X, who expressed various opinions regarding the presence of electric vehicles. The data generated from X has important value and can be explored to generate new knowledge (Azizi Hakim et al., 2022).

Through a sentiment analysis approach using the Support Vector Machine (SVM) method, a machine learning technique that is able to classify texts based on positive and negative sentiment, this research can provide a deeper understanding of people's perceptions and responses to electric vehicle taxes on social media X. By digging into

* Corresponding author

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

widespread opinions on social media platforms, we can identify trends, patterns, and relevant preferences, which can be the basis for better decision-making in designing tax policies that are more effective and responsive to the needs and expectations of the community.

According to a previous study by Pranatato analyze the sentiment contained in X about prospective presidential candidates, the results of the study showed that positive sentiment reached 57.8%, which is equivalent to 1022 reviews, while neutral sentiment was 16.6% or 295 reviews, and the negative sentiment was 25.4% or 450 reviews. Testing using the SVM method with a comparison of 80% of the training data and 20% of the test data resulted in an accuracy of 60%. Testing with a comparison of 70% of the training data and 30% of the test data resulted in an accuracy of 59%. Meanwhile, with a comparison of 60% of the training data and 40% of the test data, the resulting accuracy is 58% (Pranata et al., 2023).

Furthermore, Haranto uses sentiment objects for Telkom and Biznet services. The evaluation results show that Telkom has a positive percentage rate of 41.2% and a negative percentage of 58.8%, while Biznet has a positive percentage rate of 35.2% and a negative percentage of 64.8%. To assess the performance of the Support vector machine model used, the K-10 Folde Cross validation and Confusion Matrix methods were applied. The results show that the average accuracy for Telkom reached 79.6%, with a precision of 76.5%, a recall of 72.8%, and an f1-score of 74.6%. As for Biznet, the accuracy level reached 83.2%, with 78.8% precision, 71.6% recall, and 75% F1-score (Haranto and Sari, 2019).

Then Tineges took the object of service sentiment towards Indihome. An accuracy level of 87% was obtained, with a prediction accuracy level of 86% and a system success rate in predicting data (recall) of 95%. The error rate for all predicted data (error rate) is 13%, while the f1-score value, which is the average comparison between precision and recall, is 90%. The results of sentiment analysis of Indihome services based on the latest data show that the percentage of positive values reached 18.4%, while negative values reached 81.6%. Based on these findings, it can be concluded that the level of satisfaction of Indihome service users is relatively low (Tineges et al, 2020).

It can be concluded by using the SVM method, this study aims to produce a deep understanding of people's attitudes towards electric vehicle taxes and the factors that affect changes in sentiment.

2. LITERATURE REVIEW

The regulation of electric vehicle taxes in Indonesia is explained in Government Regulation Number 74 of 2021, which sets the sales tax rate on luxury goods (PPnBM) for electric cars at 15% of the Tax Imposition Basis which is 0%. This action shows the government's commitment to providing fiscal incentives to support the transition to the use of electric vehicles. In addition, electric vehicles are also considered a sustainable investment in the long term (Ananta et al., 2024).

Sentiment analysis is a computational process that aims to identify and group opinions contained in a text, especially to assess whether the author has a positive, negative, or neutral view of a topic (Furqan et al., 2022). The main process in sentiment analysis involves grouping the text in a sentence or document, and then assessing whether the opinion conveyed is positive, negative, or neutral (Hasugian et al., 2023).

(SVM) is a machine learning technique created by Vapnik in 1992, which is used for classification as well as regression. SVM is known as an effective method in solving non-linear classification problems and is famous for its ability to efficiently handle large data dimensions. The basic principle of SVM is to find an optimal hyperplane that can separate two data classes within a feature space. This hyperplane was chosen to have a maximum margin, which is the largest distance between data from two different classes (Kurniawan et al., 2023). Data that is at the margin limit is referred to as support vectors. After mapping, the SVM will look for the hyperplane with the largest margin to separate the classes (Handayani & Zufria, 2023).

* Corresponding author

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

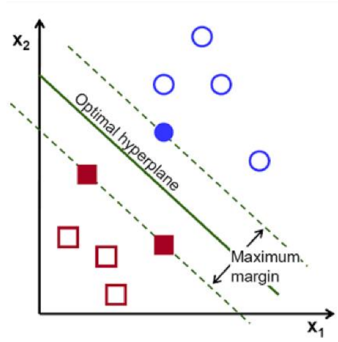

Fig 1. Hyperplane on SVM
Source: (Ginabila & Fauzi, 2023)

For binary classification problems, SVMs look for hyperplanes that maximize margins. Suppose we have a dataset with n features, the hyperlane can be expressed as:

$$w \cdot x + b = 0 \tag{1}$$

Where:
w : weight vector.
x : feature vector.
b : bias or offset.

X is one of the social networking platforms that can connect humanity around the world. The social networking and microblogging platform that first appeared in March 2006 under the name Twitter has undergone a major transformation in July 2023, when the platform was renamed X. The choice of the name "X" was made because it is considered to represent the uniqueness of each individual and has a strong futuristic and provocative connotation in popular culture (Kartino et al., 2021).

Text Preprocessing is the process of processing text data to make it more structured, which involves a series of steps such as cleaning, case folding, tokenizing, stop removal, and stemming (Destitus et al., 2020).

TF-IDF (Term Frequency-Inverse Document Frequency) is a weighting method used in text mining and information retrieval to evaluate how important a word in a document is relative to a collection or corpus of documents (Septiani & Isabela, 2022). The following is the equation in calculating TF-IDF.

$$tf - idf = tf \ x \ log \frac{N}{df} \tag{2}$$

Where:
tf : Term Frequency
Idf : Inverse Document Frequency
N : Total Number of Documents
Df : Number of Documents Containing a term of a word

Python is a programming language that can be run on various platforms, with a focus on code readability. The use of Python is mainly aimed at data analysis, data visualization, and the development and creation of artificial intelligence (Ikhsan et al., 2022).

In this study, the programming language used is Python. The study used Google Colaboratory as the main application, which is a web-based Python editor that can be accessed through a browser. Python was chosen because of its ability to provide a variety of libraries that support the data pre-processing process.

3. METHOD

This study applies a quantitative research approach that focuses on testing theories by exploring the correlation between certain variables. The purpose of this study is to develop an analysis of opinion sentiment towards electric vehicle tax on social media platform X using the Support Vector Machine method. The stages of the research can be seen in the following figure.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

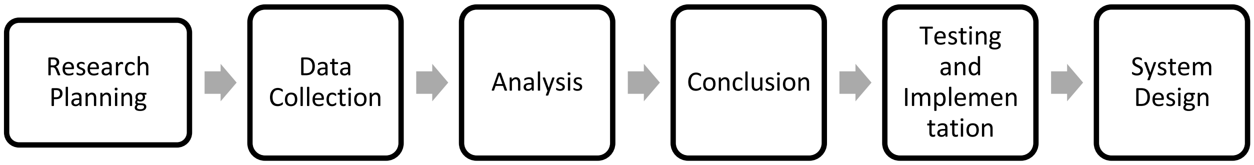

Fig 2. Flow in the framework of the research

The planning in the study uses the Support Vector Machine method for sentiment analysis on X, which focuses on the topic of electric vehicle tax, with the aim of generating sentiment data with positive or negative values.

Data collection is through crawling techniques using the Python programming language. After that, the dataset will be saved in csv format where it will be further processed. Here are some examples of datasets that are crawled using the keyword "electric vehicle tax".

Data analysis using the Support Vector Machine method, this research requires software. The software used includes Google Chrome, Operation Windows 11, Google Colab and the Python programming language. Word weighting will be done using TF-IDF, and sentiment analysis will be applied using the Support Vector Machine.

System design is a stage that researchers carry out after obtaining a dataset. In this system, sentiment analysis will be used to classify electric vehicle taxes using the Support Vector Machine.

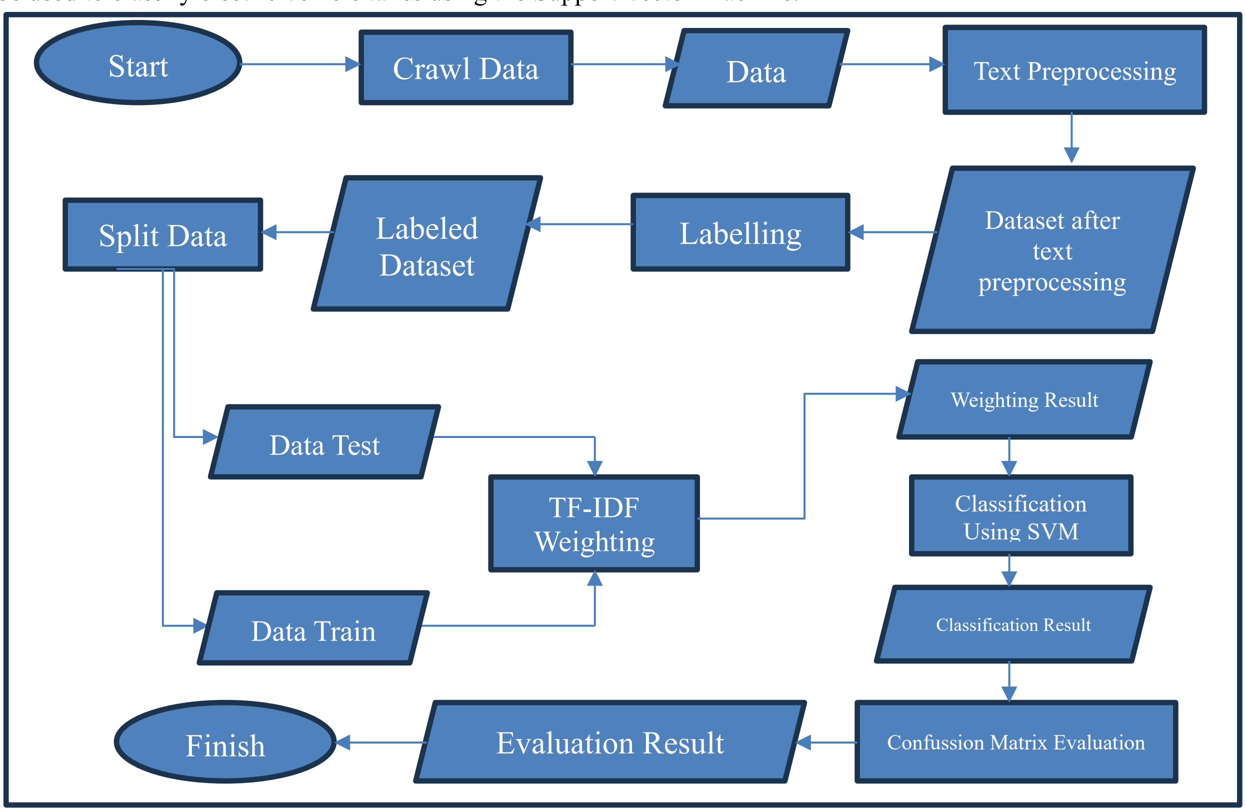

Fig 3. Research Flowchart

The test and implementation included the implementation of a pre-designed program to analyze the sentiment of electric vehicle taxes on social media X using the support vector machine (SVM) method.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

And the conclusion takes the results obtained after the process of analyzing the classification of sentiment related to electric vehicle taxes in social media X, an evaluation matrix such as precision, accuracy, recall and F1-score is obtained to assess how effective the system is in classifying data correctly.

4. RESULT

The data obtained from X is then downloaded in .csv format, with a total of 1038 data that has been successfully collected, then data that has duplicate sentiments will be deleted into data that will be used by 1014 data that has been successfully collected. Before classification, each data will go through a text preprocessing stage which includes data cleaning, case folding, tokenization, normalization, stop words removal, and stemming. After the text preprocessing process is complete.

4.1 Text Pre-processing

In the cleaning stage, the character elements that the user uses in the tweet will be removed from the review that will be included in the analysis stage.

Table 1
Cleaning Stages

Before	After
@gigirenggang If I'm not mistaken, electric vehicles get subsidies, sir, in addition to free annual taxes, the government is indeed boosting electric vehicles (the plates are blue)	If I'm not mistaken, electric vehicles get subsidies, in addition to free annual taxes, indeed the government is boosting electric vehicles with blue plates

At the case folding stage, tweets that contain capital letters will be replaced with lowercase letters on all existing sentiments.

Table 2
Tokenizing Stages

Before	After
Specification of Neta X Electric SUV Vehicle Tax Whitening Jakarta	Specification of Neta X Electric SUV Vehicle Tax Whitening Jakarta

At the normalization stage, where at this stage it will change tweets or sentences that are not standard (slangwords) into standard sentences that are in accordance with KBBI (Great Indonesian Dictionary).

Table 3
Normalization Stages

Before	After
"Ooh" "that" "sorry" "kirain" "car" "electricity" "tax" "jg". " Tp" "necessary", "learned", "lg", "behavior", "society", "especially", "jkt". "Mrk2" "yg" "still" "like" "carry" "vehicle" "personal" "this "emg" "gara2" "vehicle" "generally" "krg" "adequate" "or" "emg" "life style" . " "Feed" "PR" "lg"	"oh" "so" "sorry" "kirain" "car" "electricity" "taxed" "also" "but" "necessary" "learned" "again" "behavior" "the society" "especially" "jakarta". " "they", "who", "still", "like", "carry", "vehicle", "personal", "this", "because", "because", "vehicle", "generally", "lack", "adequate", "or", "emang", "style", "life". " will" "pr" "again"

At the stopword removal stage, words that have no meaning in a sentence, for example such as conjunction and so on.

Table 4
Stages of Stopword Removal

Before	After
"Use" "vehicle" "electric" "in" "IKN" "will" "encouraged" "through" "incentive" "tax"	"user" "vehicle" "electric" "IKN" "incentive" "tax"

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

Finally, through the stemming stage, where this stage is carried out to find the root word of a word by removing words such as affixes.

Table 5.
Stemming Stages

Before	After
Electric vehicles in IKN will get special tax incentives	electric vehicles in IKN will get special tax incentives

4.2 Data Labeling

The next stage after text preprocessing is data labeling. Data labeling in this study will use automatic data labeling with Transformers. Transformer models are used to extract features from each word, and the most popular modern NLP models consist of various models or variants of them. BERT (Bidirectional Encoder Representations from Transformers) is a transformer architecture that specifically uses encoders and has been pre-trained with deep two-way language representations. Here are some samples of data that have been labeled using transformers.

Table 6
Data Labeling Sample

stemming_data	Label
Tesla Big Win India Cuts Import Tax on Electric Vehicles Produced by Car Manufacturers Commit to Realize Million Investment in Domestic Production	POSITIVE
News for electric driving wise issuance order electric vehicle tax incentives regulate VAT incentives DTP incentives PPNBM incentives	POSITIVE
Minimal Structure Costs of Using Expensive Electric Vehicles People Driving Motorcycles Means of Transportation Businesses Consider Wrong	NEGATIVE
Fuel Motorcycle Tax Exemption Fuel Subsidy Eliminate Electric Vehicle Business Yes	NEGATIVE
Issuance of an Official Order to Exempt Import Electric Vehicles	NEGATIVE

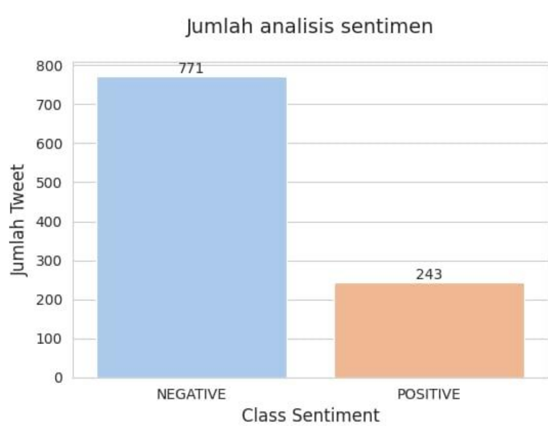

Fig 4. Analysis of Labeling Results

In the image above, negative sentiment results were obtained with a total of 771 comments and positive sentiment as many as 243 comments. It can be concluded that negative sentiment > positive sentiment.

4.3 Split Data

After going through the labeling stage, the next data will be divided or it can also be called splitting data. At this stage, the data will be divided into test data and training data. Where the training data is 80% and the test data is 20%. The following is an example of a sample of training data and test data.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

Table 7
Test Data

Training Data
China Electric Vehicle Tax Free
Thailand Invites Electric Vehicle Tax Discount Incentives

Table 8
Test Data

Test Data
Driving Tax is Expensive If You Use a Tax-Free Electric Car

4.4 TF-IDF word weighting

The next stage after going through data splitting is the TF-IDF weighting stage. At this stage, the technique of calculating the weight of each word (term) in the document is carried out by calculating the frequency of occurrence of each word. Furthermore, each of these words will be multiplied by the IDF (Inverse Document Frequency) value. The following is a sample of TF and IDF value calculations from 2 samples of training data and 1 sample of test data.

Table 9
Sample Training Data

Coaching Sentiment	Class
['China', 'free', 'tax', 'drive', 'electric']	Negative
['thailand', 'guyur', 'incentives', 'discounts', 'taxes', 'driving', 'electricity']	Negative
['package', 'incentive', 'free', 'tax', 'vehicle', 'electricity', 'import', 'sell']	Positive

Table 10
Sample Test Data

Test Sentiment
['tax', 'drive', 'expensive', 'if', 'use', 'car', 'electricity', 'free', 'tax']

Table 11
DF Value on Training Data

Term	TF			DF
	D1	D2	D3	
China	1	0	0	1
free	1	0	1	2
tax	1	1	1	3
Ride	1	1	1	3
electricity	1	1	1	3
Thailand	0	1	0	1
Guyur	0	1	0	1
incentive	0	1	1	2
discount	0	1	0	1
parcel	0	0	1	1
Import	0	0	1	1
saleable	0	0	1	1

Once the TF (Term Frequency) value is obtained, the next step is to calculate the IDF (Inverse Document

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

Frequency) value. Here are the calculations:

IDF = log () + 1= $\frac{D+1}{df+1}$log () + 1= ln(2)+1= 1.693$\frac{3+1}{1+1}$

Once the TF and IDF values are known, they can be calculated for the TF-IDF values. Where TF-IDF has the following similarities.

$W = TF \ x \ IDF$ = 1 x 1.693 = 1.693

Table 12
TF-IDF Values on Training Data

Term	**TF-IDF**		
	D1	**D2**	**D3**
China	1.693	0	0
free	1.288	0	1.288
tax	1	1	1
Ride	1	1	1
electricity	1	1	1
Thailand	0	1.693	0
Guyur	0	1.693	0
incentive	0	1.288	1.288
discount	0	1.693	0
parcel	0	0	1.693
Import	0	0	1.693
saleable	0	0	1.693

Then, the TF-IDF values are normalized to equalize the range of each data. The following is an example of the calculation in the first normalization data:

Wnorm = $\frac{TF(t,d)}{\sqrt{\sum i(TF(t,d))^2}}$

= 1.693 / (1.693√² + 1.2882 + 1.2882 + 12 + 12 + 12 + 12 + 12 + 12 + 12 + 12 + 12 + 1.6932 + 1.2882 + 1.2882 + 1.6932 + 1.6932 + 1.6932 + 1.6932)

= 0.3276

And the results are as shown in the table below.

Table 13
Data Normalization

No.	**D1**	**D2**	**D3**
1	0.328	0	0
2	0.249	0	0.249
3	0.0002	0.0002	0.0002
4	0.0002	0.0002	0.0002
5	0.0002	0.0002	0.0002
6	0	0.328	0
7	0	0.328	0
8	0	0.249	0.249
9	0	0.328	0
10	0	0	0.328
11	0	0	0.328
12	0	0	0.328

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

4.5 SVM Classification

In the classification, the kernel used is a linear kernel type because the existing data is linear data. Here's the equation used to calculate the linear kernel. After calculating the TF-IDF value in the previous table, a comparison was made using the A x AT matrix. The following is an example representation for three data.

Table 14
Linear Kernel Sample

	A1	A2	A3
B1	K(A1, B1)	K(A2, B1)	K(A3, B1)
B2	K(A1, B2)	K(A2, B2)	K(A3, B2)
B3	K(A1, B3)	K(A2, B3)	K(A3, B3)

The results of the linear kernel calculation from the sample data are presented in the table below.

Table 15
Linear Kernel Calculation

No.	**1**	**2**	**...**	**11**	**12**
1	0.108	0.082	...	0	0
2	0.082	0.124	...	0.082	0.082
3	0	0	...	0	0
4	0	0	...	0	0
5	0	0	...	0	0
6	0	0	...	0	0
7	0	0	...	0	0
8	0	0.062	...	0.062	0.062
9	0	0	...	0	0
10	0	0.082	...	0.108	0.108
11	0	0.082	...	0.108	0.108
12	0	0.082	...	0.108	0.108

Example calculation for column 1 row 1.

$$K\ (x, y)\ =\ x * y$$

$$K\ (x, y)\ =\ (t1d1\ *\ t1d1\ +\ t1d2\ *\ t1d2\ +\ t1d3\ *\ t1d3)$$

$$=\ (0.328\ *\ 0.328\ +\ 0\ *\ 0\ +\ 0\ *\ 0)$$

$$D11\ \ =\ 0.107584$$

Once the kernel value is obtained, the next step is to calculate the Hessian matrix. Before the calculation is carried out, several parameters will be determined first, such as αi, C, γ, λ and the maximum number of iterations. The following is an explanation of the parameters that will be used in the calculation process of the Hessian matrix.

Table 16
Parameter Values

Parameters	value
αi	0
C	1
γ	0.1
λ	0.5
Iteration	3

The process of calculating the Hessian matrix begins by initiating the value of α = 0, then continues by performing the calculation using the following equation. Below this is the result of the calculation of the Hessian matrix.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

Table 17
Hessian Matrix Calculation

No.	1	2	...	11	12
1	0.358	0.332	...	0.25	0.25
2	0.332	0.374	...	0.332	0.332
3	0.25	0.25	...	0.25	0.25
4	0.25	0.25	...	0.25	0.25
5	0.25	0.25	...	0.25	0.25
6	0.25	0.25	...	0.25	0.25
7	0.25	0.25	...	0.25	0.25
8	-0.25	-0.312	...	-0.312	-0.312
9	-0.25	-0.25	...	-0.25	-0.25
10	-0.25	-0.332	...	-0.358	-0.358
11	-0.25	-0.332	...	-0.358	-0.358
12	-0.25	-0.332	...	-0.358	-0.358

Here's an example of calculating the Hessian matrix for column 1 row 1
$D11 = yiyj\ (K\ (xi, xj) + \lambda 2) = 1 * 1(0.108) + 0.25 = 0.358$

The next step after calculating the value of the hessian matrix is to perform the calculation of the seqential training using the following formula. At the beginning of the calculation, the iteration starts with the 0th iteration because the initial α value is still 0, the results of the calculation are shown in the following table.

No.	**E_i**
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0

Example calculation for column 1 row 1.

$$Ei = \sum_{j=1}^{i} \alpha jDij = 0 * 0.358 = 0$$

The next step is to calculate δα to obtain the value of α. The equation used to obtain δα in the 0th iteration is as follows.

Table 18
Δαi calculation

No.	δα
1	0.1
2	0.1
3	0.1

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

4	0.1
5	0.1
6	0.1
7	0.1
8	0.1
9	0.1
10	0.1
11	0.1
12	0.1

Example calculation for the second data.
$\delta\alpha i \;=\; \min\{max[\,\gamma\,(1 - Ei - \alpha i], \mathrm{C} - \alpha i\}$
$= \min\{\max[0.1(1-0), -0]\,.\,1-0\}$
$=\; 0.1$

Table 19
1st Iteration δαi calculation

No.	α
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1
6	0.1
7	0.1
8	0.1
9	0.1
10	0.1
11	0.1
12	0.1

Example calculation for the second data.
$\alpha i \;=\; \alpha j \;+\; \delta\alpha j$
$=\; 0 \;+\; 0.1$
$=\; 0.1$

Do the process *Sequential Training* Until reaching the maximum number of iterations, which in this study is set as many as 3 iterations where to obtain the αj value that is used in looking for support vectors. The results of the calculation of αi values in the 3rd iteration are as follows.

Table 20
α value

No.	α
1	0.1074
2	0.1122
3	0.0750
4	0.0750
5	0.0750
6	0.1074
7	0.1074
8	0.1122
9	0.1074

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

10	0.1074
11	0.1074
12	0.1074

After completing the previous calculations, a process is carried out to find the support vector on each document. Based on the latest αi score, the largest score from each class is selected.

Table 21
Determining Support Vector

No.	D1	D2	D3	α	Class
1	0.328	0	0	0.1074	1
2	0.249	0	0.249	0.1122	1
3	0.0002	0.0002	0.0002	0.0750	1
4	0.0002	0.0002	0.0002	0.0750	1
5	0.0002	0.0002	0.0002	0.0750	1
6	0	0.328	0	0.1074	1
7	0	0.328	0	0.1074	1
8	0	0.249	0.249	0.1122	-1
9	0	0.328	0	0.1074	-1
10	0	0	0.328	0.1074	-1
11	0	0	0.328	0.1074	-1
12	0	0	0.328	0.1074	-1

The next step is to compute the kernel function for each class using the highest α value of each class. Manual calculation *K* (*xi, x*+) is based on the largest alpha value of the positive class, which is 0.1122, while the value of *K* (*xi, x* -) for the negative class is also taken from the highest alpha, which is 0.1122. These values are seen from the Hessian matrix in the 2nd and 8th columns, so that the kernel values for each class are obtained as follows.

$$
\begin{aligned}
K\,(xi, x+) &= \sum \alpha i y i D i \\
&= (0.1074 * 1 * 0.332) + (0.1122 * 1 * 0.374) + (0.075 * 1 * 0.25) + (0.075 * 1 * 0.25) \\
&\quad + (0.075 * 1 * 0.25) + (0.1074 * 1 * 0.25) + (0.1074 * 1 * 0.25) \\
&\quad + (0.1122 * -1 * 0.312) + (0.1074 * -1 * 0.25) + (0.1074 * -1 * 0.332) \\
&\quad + (0.1074 * -1 * 0.332) + (0.1074 * -1 * 0.332) \\
&= 0.0584676
\end{aligned}
$$

$$
\begin{aligned}
K\,(xi, x-) &= \sum \alpha i y i D i \\
&= (0.1074 * 1 * 0.332) + (0.1122 * 1 * 0.374) + (0.075 * 1 * 0.25) + (0.075 * 1 \\
&\quad * 0.25) + (0.075 * 1 * 0.25) + (0.1074 * 1 * 0.25) + (0.1074 * 1 * 0.25) \\
&\quad + (0.1122 * -1 * 0.312) + (0.1074 * -1 * -0.25) + (0.1074 * -1 * 0.332) \\
&\quad + (0.1074 * -1 * -0.332) + (0.1074 * -1 * -0.332) \\
&= -0.05846
\end{aligned}
$$

After the values *of K* (*xi, x*+) and *K* (*xi, x* -) are calculated, the next step is to determine the bias value, which is as follows.

$$b = -\frac{1}{2}\,[w.x+ + w.x-] = -\frac{1}{2}\,[0.0584676 + (-0.05846) = -\frac{1}{2}\,(0.000008) = -0.000004$$

Once the bias value is obtained, the process proceeds to the testing stage. The first step in this stage is to calculate the TF-IDF value on the test data used as a sample.

Table 22
TF-IDF Value Normalization Test Data

No.	*Term*	TF	DF	IDF	TF-IDF	*Norm*

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

1	tax	2	1	2.098	4.196	0.4019
2	Ride	1	0	2.791	2.791	0.2673
3	expensive	1	0	2.791	2.791	0.2673
4	Kalo	1	0	2.791	2.791	0.2673
5	USE	1	0	2.791	2.791	0.2673
6	car	1	0	2.791	2.791	0.2673
7	electricity	1	0	2.791	2.791	0.2673
8	free	1	0	2.791	2.791	0.2673

The next step is to calculate the kernel for each test data using the previously obtained training data. The results of the kernel calculation between the training data and the test data are displayed as follows.

Table 23
Kernel Calculation of Training Data on Test Data

No.	tax	Ride	expensive	Kalo	USE	car	electricity	free
1	0.13	0.088	0.088	0.088	0.088	0.088	0.088	0.088
2	0.1	0.067	0.067	0.067	0.067	0.067	0.067	0.067
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0.088	0.088	0.088	0.088	0.088	0.088	0.088
7	0	0.088	0.088	0.088	0.088	0.088	0.088	0.088
8	0.1	0.067	0.067	0.067	0.067	0.067	0.067	0.067
9	0	0.09	0.09	0.09	0.09	0.09	0.09	0.09
10	0.13	0.09	0.09	0.09	0.09	0.09	0.09	0.09
11	0.13	0.09	0.09	0.09	0.09	0.09	0.09	0.09
12	0.13	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Example calculation in column 2 row 1.

$$K\ (x, y)\ =\ (x2 * y1)\ +\ (x2 * y2)\ +\ (x2 * y3)\ =\ (0.2673 * 0)\ +\ (0.2673 * 0.328)\ +\ (0.2673 * 0 =\ 0.13$$

Once the kernel value is obtained, the next step is to calculate the weight of the test data. The calculation of the test data weights is done using the following equation, and the results will be obtained for column 1 row 1 as shown below.

$$w.x\ =\ \alpha iyi\ K\ (xi, xj)$$
$$=\ 0.1074\ *\ 1\ *\ 0.13$$
$$=\ 0.014$$

Table 24 Weighting of Test Data Terms

No.	Tax	Ride	expensive	Kalo	USE	car	electricity	free
1	0.014	0.01	0.01	0.01	0.01	0.01	0.01	0.01
2	0.011	0.008	0.008	0.008	0.008	0.008	0.008	0.008
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01
7	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01
8	-0.011	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008
9	0	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
10	-0.014	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
11	-0.014	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
12	-0.014	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

∑	-0.003	-0.096	-0.096	-0.096	-0.096	-0.096	-0.096	-0.096

Next, the sum of the test data weights is used to calculate the value of the function f(x). Based on the hyperplane obtained, the data will be of positive value if $w.x + b = 1$ and will be negative if $w.x + b = -1$.
Test data:['tax', 'vehicle', 'expensive', 'if', 'use', 'car', 'electricity', 'free', 'tax']

$$
\begin{aligned}
f(x) &= w.x + b \\
&= \sum \alpha iyi\ K\ (xi, xj) + b \\
&= ((-0.003) + (-0.096) + (-0.096) + (-0.096) + \cdots + (-0.096)) + (-0.000004) \\
&= -0.675
\end{aligned}
$$

$$
\begin{aligned}
Fungsi\ Klasifikasi &= sign\ (-675.000004) \\
&= -1
\end{aligned}
$$

After the testing process was carried out on the test data, it was found that the classification function produced a positive value of 1. Therefore, the data is classified into the category of class 1, which is a negative class. In the training data, there are positive and negative sentiment classes, where SVM will learn the characteristics of the words of each class. Here is a wordcloud of positive and negative sentiment.

Fig 5. Wordcloud Positive

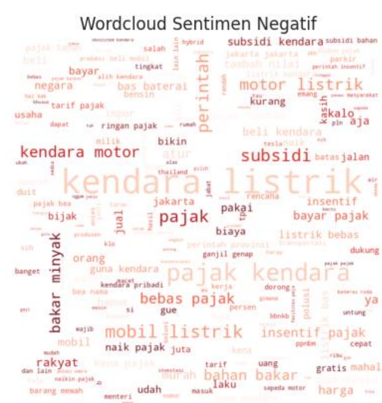

Fig 6. Wordcloud Negatives

4.6 Confusion Matrix

The results of the classification of test data in the form of sentiment classes will be compared with the actual class, so that the accuracy, precision, recall, and f1-score values of the model used against the dataset can be calculated. The

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

following is a visualization of the confusion matrix.

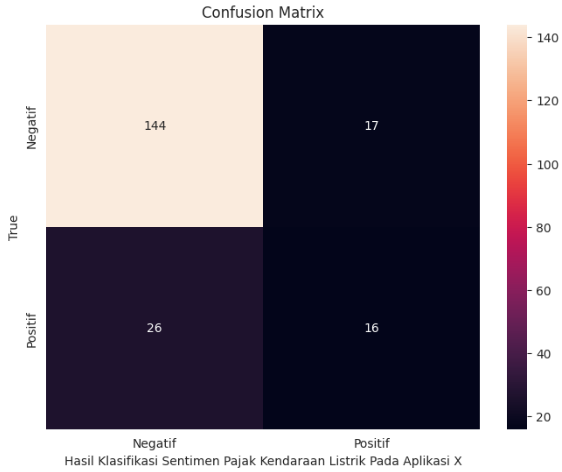

Fig 7. Visualization of the Confussion Matrix

From the table above, the value can be calculated *accuracy, precision, recall* and *f1-score* by using the following equation.

$Accuracy = \frac{160}{144+17+26+16}$ x 100% = 79%

$Precision$ = x 100% = 85%$\frac{144}{144+26}$

$Recall = \frac{16}{16+26}$ x 100% =89%

$F1\text{-}Score = \frac{2 \times 85 \times 89}{85+89}$ x 100% = 87%

Table 25 Electric Vehicle Tax Classification Report

Accuracy: 0.7881773399014779				
	Precision	Recall	F1-Score	Support
NEGATIF	0.85	0.89	0.87	161
POSITIF	0.48	0.38	0.43	42
accuracy			0.79	203
macro avg	0.67	0.64	0.65	203
weighted avg	0.77	0.79	0.78	203

DISCUSSIONS

The results of this study show that the majority of sentiment on platform X (Twitter) towards electric vehicle taxes is negative, reaching around 76% or 771 comment data, while positive sentiment is only around 24% or 243 comment data. The accuracy of the Support Vector Machine (SVM) method in sentiment classification reached 79% with 85%

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

precision and 89% recall, which indicates the effectiveness of this method in analyzing public opinion. This finding is in line with previous research by Pranata et al. (2023), which also shows the dominance of negative sentiment towards public policy on social media, especially related to dissatisfaction with operational costs and electric vehicle infrastructure. While SVM has proven to be effective, there are still opportunities to improve accuracy with larger datasets and better preprocessing techniques. The study also suggests the potential of merging SVM with other algorithms, such as Naive Bayes, to improve performance. The main limitation of this study is the amount of data used, so further research with a larger sample is needed to enrich the results and scope of the findings.

5. CONCLUSION

This study addresses the problem of how public sentiment on platform X regarding electric vehicle taxes can be classified efficiently using the Support Vector Machine (SVM) method. By applying transformers for automatic data labeling and preprocessing on 1,014 tweets, the model was able to achieve 79% accuracy, 85% precision, 89% recall, and 87% f1-score. highlighting public skepticism toward the electric vehicle tax, particularly its perceived unfairness for lower middle-class citizens, who find electric vehicles more expensive than traditional fuel-powered cars.

6. REFERENCES

Ananta, A., Alvin Hidayat, D., Early Al Husni, D., Fauzi Ramadhan, I., Triyono, I., Al Qossam, I., & Rohmah, R. D. (2024). Peningkatan Kesadaran dalam Penggunaan Kendaraan Listrik di Lingkungan Universitas Negeri Semarang Melalui Kampanye Energi Bersih Sitasi. *Jurnal Angka*, *1*(1), 120–134. http://jurnalilmiah.org/journal/index.php/angka

Audrey Ramadhina, & Fatma Ulfatun Najicha. (2022). Regulasi Kendaraan Listrik di Indonesia Sebagai Upaya Pengurangan Emisi Gas. *Jurnal Hukum To-Ra : Hukum Untuk Mengatur Dan Melindungi Masyarakat*, *8*(2), 201–208. https://doi.org/10.55809/tora.v8i2.126

Azizi Hakim, M., Heriana, E., Ekoprianto, A., Sukmara, S., & Susanto, D. (2022). ANALISIS BANGUN RANGKA MOBIL. *TECHNOMA*, *02*(01).

Destitus, C., Wella, & Suryasari. (2020). Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia. *ULTIMA InfoSys*, *XI*(2), 107–111.

Furqan, M., Sriani, & Mayang Sari, S. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.COM*, *21*(1), 52–61.

Ginabila, & Fauzi, A. (2023). Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine. *J u r n a l I l m i a h I L K O M I N F O - J u r n a l I l m u K o m p u t e r d a n I n f o r m a t i k A*, *6*(2), 111–122.

Handayani, A., & Zufria, I. (2023). Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM. *Journal of Information System Research (JOSH)*, *5*(1), 53–63. https://doi.org/10.47065/josh.v5i1.4379

Haranto, F. F., & Sari, B. W. (2019). IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET. *Jurnal Pilar Nusa Mandiri*, *15*(2), 171–176. https://doi.org/10.33480/pilar.v15i2.699

Hasugian, A. H., Fakhriza, M., & Zukhoiriyah, D. (2023). Analisis Sentimen Pada Review Pengguna E-CommerceMenggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD* , *6*(1), 98–107. https://ojs.trigunadharma.ac.id/index.php/jsk/index

Ikhsan, M., Armansyah, A., & Tamba, A. A. (2022). Implementasi Jaringan Syaraf Tiruan Backpropagation Pada Klasifikasi Grade Teh Hitam. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2). https://api.semanticscholar.org/CorpusID:259618265

Kartino, A., M. Khairul Anam, Rahmaddeni, & Junadhi. (2021). Analisis Akun Twitter Berpengaruh terkait Covid-19 menggunakan Social Network Analysis. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 697–704. https://doi.org/10.29207/resti.v5i4.3160

Kurniawan, R., Halim, A., & Melisa, H. (2023). Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine). *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(2), 903–912. https://doi.org/10.30865/klik.v4i2.1246

Pranata, B., Susanti, Erlinda, S., & Asnal, H. (2023). Support Vector Machine untuk Sentiment Analysis Bakal Calon Presiden Republik Indonesia 2024. *Indonesian Journal of Computer Science*, *12*(3), 1335–1349.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4739

Submitted : Sep 28, 2024
Accepted : Oct 14, 2024
Published : Oct 16, 2024

Prianjani, D., & Sutopo, D. W. (2018). STUDI KOMPARASI PENELITIAN STANDAR KENDARAAN LISTRIK DUNIA DENGAN STANDAR KENDARAAN LISTRIK INDONESIA. *Prosiding SNST Ke-9*. 13

Septiani, D., & Isabela, I. (2022). ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS. *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia*, *1*(2), 81–88.

Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *4*(3), 650–658. https://doi.org/10.30865/mib.v4i3.2181

Veza, I., Afzal, A., Mujtaba, M., Hoang, A. T., Balasubramanian, D., Sekar, M., Fattah, I. M. R., Soudagar, M. E. M., El-Seesy, A. I., Djamari, D. W., Hananto, A. L., Putra, N. R., & Tamaldin, N. (2022). Review of artificial neural networks for gasoline, diesel and homogeneous charge compression ignition engine. *Alexandria Engineering Journal*. https://api.semanticscholar.org/CorpusID:246698256

* Corresponding author