

# Sentiment Analysis towards the 2024 Vice Presidential Candidate Debate Using the Support Vector Machine Algorithm

Raihan Rizieq Harahap<sup>1)</sup>, Mhd. Furqan<sup>2)</sup>

<sup>1,2)</sup>Department of Computer Science, Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara

<sup>1)</sup>[raihanrizieq160@gmail.com](mailto:raihanrizieq160@gmail.com), <sup>2)</sup>[mfurqan@uinsu.ac.id](mailto:mfurqan@uinsu.ac.id)

Submitted : Jul 18, 2024 | Accepted : Jul 23, 2024 | Published : Jul 27, 2024

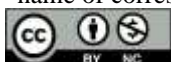
**Abstract:** In today's digital era, social media plays an important role in disseminating information and influencing public opinion. For instance, YouTube. At the 2024 Vice Presidential Debate, YouTube became a medium where people gave various comments. This study aimed to analyze public sentiment through comments on the 2024 Vice Presidential Debate on the Metro TV YouTube channel. This study used descriptive quantitative methods with the Support Vector Machine algorithm to identify various public comments. The results show that from the data experiment taken as many as 1012 data, 80% data training amounting to 809 data and 20% data testing amounting to 203 data is carried out. An accuracy of 82% was obtained with a precision value of 80%, a recall value of 87%, and an f1-score value of 83%. With a fairly high accuracy value, the support vector machine model can be said to be the right model to calculate the accuracy value in sentiment analysis.

**Keywords:** sentiment analysis, vice presidential debate, youtube, support vector machine.

## INTRODUCTION

The election of presidential and vice presidential candidates which is held every 5 years is part of the democratic process, especially in Indonesia. The 2024 general election for Presidential and Vice Presidential Candidates in Indonesia is one of the political events that is eagerly awaited by the public. This event is one part of the campaign process for presidential and vice presidential candidates, the implementation of which is regulated and designed by the General Election Commission (KPU). This debate event is a forum for candidates to show each other's ideas on an issue (Saptanti, 2020) The prospective candidates include pair 01, namely Anies Baswedan and Muhaimin Iskandar, pair 02, namely Prabowo Subianto and Gibran Rakabumi Raka, then pair 03, namely Ganjar Pranowo and Mohammad Mahfud Mahmodin. The debate for vice presidential candidates begins with the second debate on December 22 2023 with the theme of Economy (people's economy and digital economy), Finance, Tax Investment, Trade, APBN-APBD Management, Infrastructure, and Urban Affairs. The fourth debate will be on January 21 2024 with the theme Sustainable Development, Natural Resources, Environment, Energy, Food, Agrarian Affairs, Indigenous Peoples and Villages. Debate is one of the important stages in the presidential election series which allows candidates and representatives to convey their vision, mission and programs to the public. In the current digital era, one social media plays an important role in disseminating information and influencing public opinion, namely YouTube. Therefore, analyzing public sentiment towards the 2024 vice presidential candidate debate is a very relevant thing to study. YouTube is a social media that has content in the form of videos. Where, the video aims for entertainment, knowledge and so on. YouTube users can also create, view and share

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

videos for free. YouTube also provides a comments feature as a channel for viewers' responses (those who watch videos) to the video content they watch. Comments can be positive or negative (Mualfah et al., 2023). To overcome this problem, a sentiment analysis system is needed that can classify the sentiment of public opinion regarding the 2024 vice presidential candidate debate on Metro TV's YouTube channel social media. One of the methods or algorithms used is Support Vector Machine. There was previous research conducted by (Handayani & Zufria, 2023) entitled "Sentiment Analysis of the 2024 Indonesian Presidential Candidates on Twitter Using the SVM Algorithm", using the Support Vector Machine method. The total initial sentiment analyzed was 1,719 sentiments. With a distribution of 597 data for Anies Baswedan, 627 data for Ganjar Pranowo and 495 data for Prabowo Subianto. The analysis process carried out had an accuracy value of 75% for the sentiment of candidate 1, 86% for candidate 2, and 72% for candidate 3. So the average accuracy for the three candidates was 78%. In addition, related research was conducted by (Furqan et al., 2023) entitled "Support Vector Machine Algorithm for Analysis of Indonesian People's Sentiment towards the Corona Virus Pandemic on Social Media", conducted research to produce a model that can carry out sentiment classification using the SVM algorithm. In this research, it is found that the SVM algorithm can be used to carry out sentiment analysis on research topics regarding COVID-19 learning. The accuracy test results show the highest figure, reaching 98%. Furthermore, related research was conducted by (Sriani et al, 2023.) entitled "Sentiment Analysis of Policies for Providing Electric Motorcycle Subsidies Using the Support Vector Machine Method.

" The level of accuracy produced by the Support Vector Machine method in classifying sentiment can be said to be good. A dataset of 700 data with a training data and test data ratio of 8:2, namely 560 training data and 140 testing data, obtained accuracy values of 86.43%, precision of 83.33%, recall of 87.30%, and f1-score of 85.27%. The difference and novelty of this research from previous research is that this research was developed using the Support Vector Machine algorithm with linear dimension features. However, the social media used is no longer Twitter or X, but YouTube, more precisely the Metro TV channel. The researcher would create a system that can analyze public opinion from the YouTube channel Metro TV, to get public opinion sentiment contained in the YouTube comments feature in the 2024 vice presidential candidate debate video. Therefore, the researcher would classify comments in the analysis of public sentiment towards the 2024 Vice Presidential Debate to find out public perceptions of the candidates with issues discussed. The application of sentiment analysis was carried out to see whether public opinion regarding the 2024 Vice Presidential Debate case could contain positive or negative sentiment polarities. In the field of sentiment analysis, words will be extracted in the form of text into a vector and then labeled to be used as information in sentences (Furqan et al., 2023). The SVM (Support Vector Machine) algorithm in sentiment analysis of the 2024 Vice Presidential Candidate Debate can prove to be effective in classifying text based on positive and negative sentiment. One advantage is that this step can be completed relatively quickly. The data taken in this research was 1012 comment data.

## LITERATURE REVIEW

### Sentiment Analysis

Sentiment analysis is the ability of software to understand the attitude, perspective, or opinion of a writer or speaker, commonly referred to as opinion mining and semantic analysis. This process involves identifying positive and negative statements in a text and scoring them to indicate their level of positivity or negativity (Furqan et al., n.d.). Sentiment analysis is utilized to provide an overview of social media to determine whether the prevailing opinions are positive or negative (Nurian & Nurina Sari, 2023).

### Vice Presidential Candidate Debate

Being one of the few democracies in the world, Indonesia has long practiced pemilu as a kind of rakyat repression or guerilla warfare. According to PKPU Nomor 23 Tahun 2018 Pasal 1 Ayat 21, in the process of the tahapan pemilu, it is stated that the pemilu camp, also called the kampanye, is a group of people or another organization that is invited by the pemilu to ensure that the program, vision, and values of the pemilu are provided. Furthermore, verse 1 of chapter 23 states that the only campaign

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

strategy used by KPU is public debates or open debates between rival groups. Debate, being a type of contemporary rhetoric, is usually confined to two parties or more that prohibit communication in language and persistently work to undermine the dignity and reputation of other people or parties so that they can no longer exist (Iskandar, 2020).

### Labeling Textblobs

Labeling is the process whereby the results of the previous step are subjected to the polarity of the *ambil ulasan* in order to produce two categories: positive and negative labels (Mualfah et al., 2023). In this study, the labeling process is carried out automatically using textblobs. Textblob will reduce subjectivity and polarity. Polarity serves as a tool to examine the sense of ambiguity in a text. On the other hand, subjectivity serves the purpose of highlighting the significance of a given text. The text's value could be an opinion or a factual statement. When a text has more subjectivity, it can be described as an opinion; conversely, when a text has more polarity, it indicates a positive emotion (Baita & Cahyono, 2021).

### Support Vector Machine (Svm)

Support Vector Machine is a biased classification method that is applied to both linear and non-linear data that is represented by a hyperlane or *pembatas* (Raudya et al., 2023). The SVM method is one technique that is frequently used to classify data, particularly text data. The one benefit is that this procedure can be completed quite quickly. Sentiment analysis and opinion analysis are academic disciplines that examine how people's feelings, opinions, *sikaps*, *penilaian*, and *perasaan* are expressed in written language (Handayani & Zufria, 2023).

### Term Frequency-Inverse Document Frequency (TF-IDF)

The term frequency-inverse document frequency (TF-IDF) model is used to determine the relationship between the term and the data by applying a *bobot* to each word or phrase (Oktavia & Ramadahan, 2023). Once the words have been reduced to *angka*, the TF-IDF number is compared to the learning criteria that are considered in context and can be met by means of *mesin*-based learning methods. We may use *TfidfVectorizer* to do TF-IDF calculations using the Python Sklearn module (Rabbani et al., 2023).

### Confusion Matrix

The evaluation method used in the research uses a confusion matrix. Confusion Matrix, sometimes known as Error Matrix, is a technique used to measure accuracy for the classification or supervised learning process. In terms of arithmetic average, there are eight combinations of predicted and actual values. False Positive (FP) and False Negative (FN) are the two terms in question. True Positive (TP) data is positive information that is reliably predicted. The True Negative (TN) value is the total amount of negative data that can be reliably detected. False Positive (FP) error, also known as Type-1 error, is a negative data that is detected as positive data. False Negative (FN) error, also known as Type-2 error, is a result of True Positive error, where positive data is detected as negative data (Purbolaksono et al., 2021).

### Evaluation and Validation

Evaluation is done to determine the results of a model that has been developed and aligned with existing or absent facts. It is necessary to do more research based on the completed study results in order to determine the accuracy of the results (Permata Aulia et al., 2021). Validation is carried out to understand the estimated model performance; in this instance, *k-fold* validation is used. *K-fold* validation is a single cross-validation model that works by storing a large number of *k-fold* data points and doing iteration/*perulangan* on a large number of *k-fold* data points as well (Arsi & Waluyo, 2021). *Akurasi* illustrates the model's success rate in a proper label or class prediction. estimating the extent to which a model can accurately classify data in a comprehensive manner. In other words, accuracy illustrates the model's performance in appropriate *kelas* prediction (Rabbani et al., 2023).

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

**METHOD**

The research method used is a quantitative descriptive method to analyze the content or text used in this study. This study discussed the analysis of public sentiment or opinion regarding the 2024 Vice Presidential Candidate Debate through comments on the Metro TV YouTube channel. This framework sequence provides an overview of the stages that will be implemented in solving the problems discussed. The data collection technique in this study was carried out to analyze and obtain data and information related to this study. The data taken is in the form of video comments on YouTube social media on the Metro TV channel. Retrieval of comment data was carried out using the scraping technique using Google Colaboratory, and the data that was successfully retrieved was 1,100 comments which were saved in the form of a .csv format.

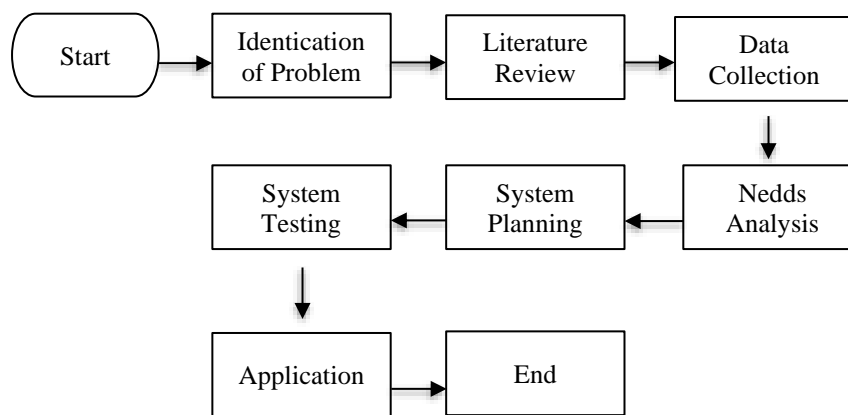


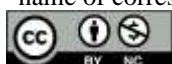
Figure 1 Research Framework

In this study, the researcher applied a data sharing ratio of 80:20, 70:30, and 60:40. From the total data obtained, it is used as training data to develop a sentiment analysis model. This training data has been automatically labeled using TextBlob, namely as positive or negative, according to the sentiment contained in the data. Meanwhile, the remaining data is used as test data to evaluate system performance. Implementation, a dataset consisting of 1100 data was collected from the Metro TV YouTube channel using the Web Scraping method with the help of Google Colaboratory. The dataset was then filtered to remove data that contained sarcasm, sarcasm or irony, leaving 1012 relevant data. Next, the dataset will undergo a preprocessing process to clean the data for further analysis. After the preprocessing process is complete, the prepared data will be labeled positive or negative using the TextBlob library for sentiment analysis.

Table 1 Example of Commentary Data from the Data Collection Process

Date and Time	User	Comment
2023-12-22 At 20:51	@asepnurly550	Sipaling lihat teks semuanya gk capres nya gk cawapres nya,, cmn no 2 doang yg jarang lihat teks semuanya dari hati yg disampaikan ole kata2
2023-12-22 At 20:49	@sinahbagus5720	Debat Cawapres milik Gibran RBR 70 % 02 dan 01 + 03 hanya 30% dibagi 2.
2024-01-22 At 02:06	@ellamasyitoh8502	Jujur pantas prabowo gandeng ms gibran,tau lh kali ini prabowo bisa ad kesempatan menang talak.
2024-01-22 At 02:06	@catursuriani9598	GIBRAN KERENNNNN!!!!!!!

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

After being weighted using the TF-IDF method, the system divides the dataset into two parts: training dataset and testing dataset, with a ratio of 80:20, 70:30, and 60:40. Next, the system processes the training dataset and uses the dataset as a learning model to classify word-based sentiment. After the training process is complete, the validation stage is carried out by testing k-fold cross validation and evaluating the confusion matrix, one of the methods commonly used to calculate accuracy. Then according to (Sriani et al., 2023), the system uses a model that has been trained to predict sentiment on word-based testing datasets. It is hoped that the system will have a good level of prediction accuracy, so that it can provide accurate results in classifying sentiment on the testing dataset. This study aimed to build a model for classifying sentiment on user comment data on YouTube. The following is a system design flowchart that was created.

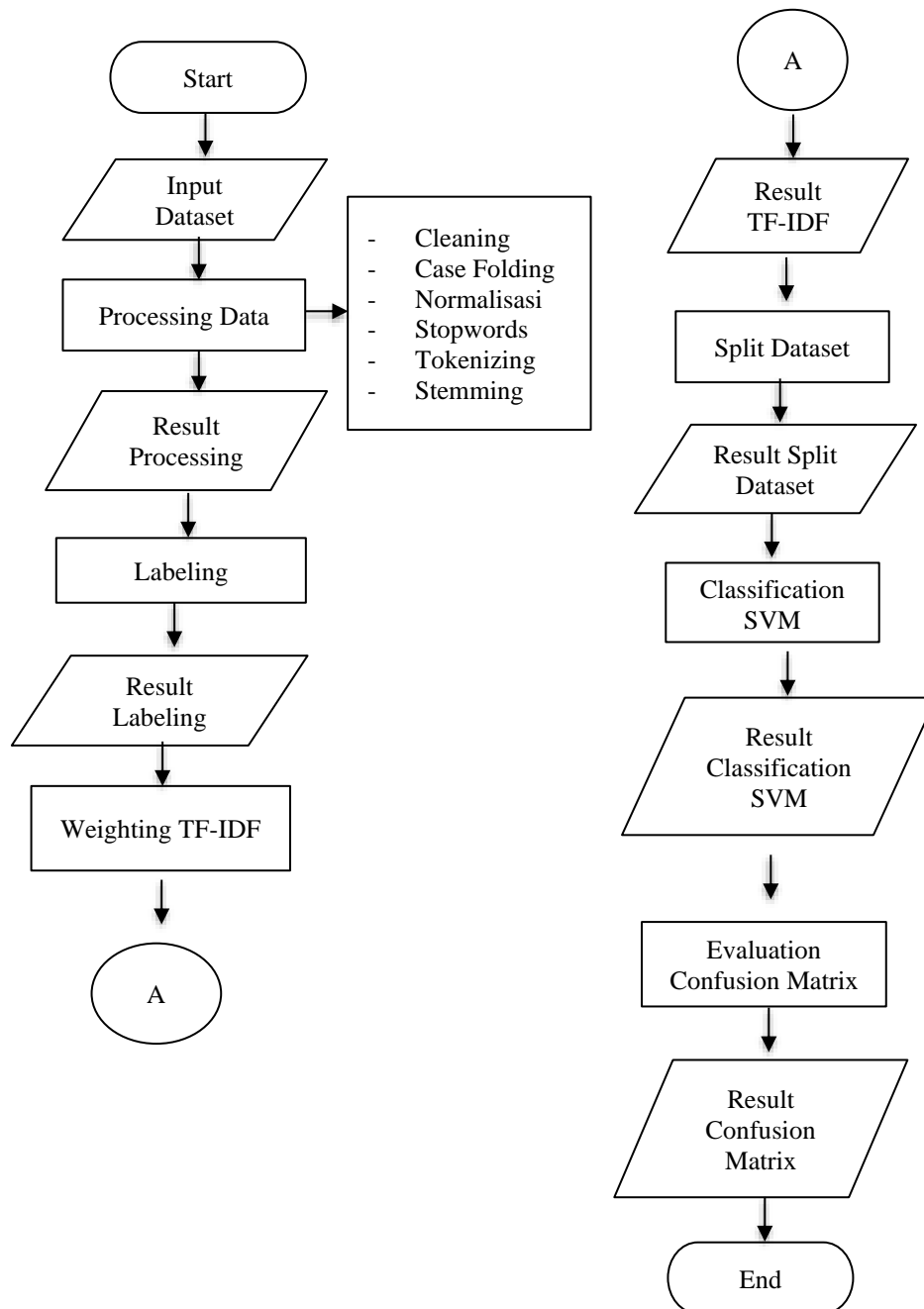


Figure 2 Support Vector Machine Sentiment Analysis Flowcharts

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Data analysis discussed the flow of data and the system used in sentiment analysis research on the 2024 vice presidential candidate debate. The data used in this study is in the form of public opinion on the Metro TV YouTube channel which is related to this study. The data used is obtained from crawling results which are saved and also processed in .csv format. Before entering Support Vector Machine training, it is necessary to preprocess the data. This stage includes case folding, filtering, tokenization, normalization, and stemming. All data obtained is given a label with the aim of learning about the dataset during the training data process at the classification stage. The classification method used is the Support Vector Machine algorithm. This algorithm is used to determine the evaluation of the system created in the opinion classification process of the data tested based on the training data and test data processes. Then the system will carry out a classification and will produce information on the results of the classification, as well as prediction results on existing test data in the form of a confusion matrix which functions in this research as an evaluation of system performance by showing the accuracy of the Support Vector Machine algorithm against the public opinion dataset in Metro TV YouTube channel.

## RESULT

### Data Preprocessing

Data preprocessing is a process that really needs to be done to reduce noise to make it easier to test sentiment analysis. Therefore, the data must be cleaned to make the research process easier in processing the data. The flow of several stages in data preprocessing includes cleaning, case folding, normalization, stopwords, tokenizing and stemming.

#### Case Folding

Case folding is the first step in preprocessing in this research. In this section, the case folding function is to equate all letters in the dataset to lower case. Example of a word after case folding “Semakin” becomes “semakin,” the word has been successfully changed to lower case.

Table 2 Case Folding Results

Comment Data	Case Folding Results
@rendragalaxy9474: “Pak mahfud kembali jadi menkopolhukam. Cak imin jadi menteri agama.”	@rendragalaxy9474: “Pak mahfud kembali jadi menkopolhukam. Cak imin jadi menteri agama.”

#### Cleaning

Cleaning is a process of identification with the aim of eliminating or deleting data so that data is minimally disturbed. Things that are deleted include hashtags, user accounts, emoticons, links, email addresses and punctuation.

Table 3 Cleaning Results

Case Folding Data	Cleaning Results
@rendragalaxy9474: “Pak mahfud kembali jadi menkopolhukam. Cak imin jadi menteri agama.”	“pak mahfud kembali jadi menkopolhukam. cak imin jadi menteri agama.”

#### Normalization

Normalization is a process of changing, deleting and correcting non-standard words to become standard, such as the abbreviation “yg” being changed to “yang”, “udh” to “udah,” and “sm” to “sama.” These words are often used on social media in documents that are not detected in Indonesian standard.

\*name of corresponding author



Table 4 Normalization Results

Cleaning Data	Normalization Results
gibranlah netral dlm tutup calon wapres tidak pakai teks tdk banyak omongdibandingkan wakil capres moga brpasangan prabowa gibran menang hsl kerja bukti amanah pundak gibranlah netral dalam tutup calon wapres tidak pakai teks tidak banyak omongdibandingkan wakil capres moga brpasangan prabowa gibran menang hasil kerja bukti amanah pundak	gibranlah netral dlm tutup calon wapres tidak pakai teks tdk banyak omongdibandingkan wakil capres moga brpasangan prabowa gibran menang hsl kerja bukti amanah pundak gibranlah netral dalam tutup calon wapres tidak pakai teks tidak banyak omongdibandingkan wakil capres moga brpasangan prabowa gibran menang hasil kerja bukti amanah pundak

**Stopwords**

Stopwords is a step taken to delete words in a document that have no meaning, such as conjunctions and pronouns.

Table 5 Stopwords Results

Normalized Data	Stopwords Results
orang 'saya' cerdas pilih nomor 2 buah jatuh 'tidak' jauh 'di' pohonnya jokowi gibran lengkap mantap sempurna	orang cerdas pilih nomor 2 buah jatuh jauh pohon nya jokowi gibran lengkap mantap sempurna

**Tokenizing**

Tokenizing is the stage in which to separate the words in the document by making them into one word at a time.

Table 6 Tokenizing Results

Stopwords Data	Tokenizing Results
orang cerdas pilih nomor 2 buah jatuh jauh pohon nya jokowi gibran lengkap mantap sempurna	['orang', 'cerdas', 'pilih', 'nomor', '2', 'buah', 'jatuh', 'jauh', 'pohon', 'nya', 'jokowi', 'gibran', 'lengkap', 'mantap', 'sempurna']

**Stemming**

Stemming is the process of changing words in text into their basic form by removing the beginning or ending of words. The aim is to reduce variations in words that have the same root, making it easier to analyze the text.

Table 7 Stemming Results

Tokenizing Data	Stemming Results
"genz", "butuh", "program", "kredit", "usaha", "anak", "muda", "bunga", "didik", "mudah", "beasiswa", "ekonomi", "digital", "industri", "kreatif", "umkamu", "karena", "semua", "naik", "perlu", "bikin", "mumet", "all in"	program kredit usaha anak muda bunga butuh cuma mudah banyakin program beasiswa tingkat ekonomi digital industri kreatif umkamu karna kalo semua naik perlu program kredit bikin mumet all in

**Labeling Data**

Previously obtained data will be labeled automatically using *TextBlob*. Before entering the data labeling stage, the author will first translate the text using the *Google Translate library*. Data labeling is calculated using *TextBlob* to obtain *polarity* and *subjectivity values*. A *polarity* value  $\leq 0$  will be

\*name of corresponding author



labeled as negative sentiment, for *polarity* > 0 it will be labeled as a positive value. Labeling data resulted in negative sentiment of 547 and positive 465 from 1013 datasets.

	textdisplay	tweet_english	Subjectivity	Polarity	Analysis
0	hal mas gbrn letrang vibes nya kaya hal do	see mas gbrn calm his rich vibes see...	0.799000	0.337500	Positive
1	itu membuaiku merinding lmao	that's what gave me goosebumps lmao	1.000000	0.600000	Positive
2	menang no urut 2	win no sequences 2	0.400000	0.800000	Positive
3	eh pak mahfud kok kayak guru sd mencocorai ana...	uh sr mahfud how come it seems like the teach...	0.000000	0.000000	Negative
4	pasang serasi langsung sele gerje pasang kok k...	pairs compatible directly sele gerje pairs kok...	0.400000	0.100000	Positive
...					
1007	paslon no 1 cuma teori ngambang jesus hancur b...	paslon no 1 is just a floating theory clearly...	0.383333	0.100000	Positive
1008	suara mirip banget pak Jokowi	sounds very similar to Pak Jokowi	0.520000	0.900000	Negative
1009	lelap gbrn ilegal lewat jalur mk, parital pand...	still illegal gbrn through the path of his c...	0.058333	-0.208333	Negative
1010	farris kzmkn paling atas	should be the top comment	0.900000	0.500000	Positive
1011	soal sokloda special kucoba melek kota susah...	about sokloda special it's trying to liberate...	0.767857	0.054286	Positive

Figure 3 Labeling Results Data

## DISCUSSIONS

### Feature Extraction

After the data has been labeled, feature extraction is carried out using *TF-IDF*, which is a weighting that is done to get the value of a word that has been successfully extracted, as in the example of Figure 3.2, where these three modules are very useful in processing and representing text for tasks such as *text classification*, sentiment analysis, or document clustering.

```
# TF-IDF
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer, HashingVectorizer
cvec=CountVectorizer()
tvec=TfidfVectorizer()
hvec=HashingVectorizer()
```

Figure 4 TF-IDF Feature Extraction

Then divide the data, here the researcher uses data division to find out the Support Vector Machine model as in the example in Table 3.7 below.

Data	Data Sharing
2	80:20
3	70:30
4	60:40

### Support Vector Machine Classification

The classification process is carried out using a support vector machine algorithm and linear kernel. By using 'scikit-learn' to create an SVM model then 'clf1' which is an SVM model with a linear kernel, the program results can be seen in Figure 3.3 below.

```
from sklearn.svm import SVC

clf1 = SVC(kernel="linear")
```

Figure 5 Support Vector Machine Algorithm

\*name of corresponding author





**Confusion Matrix Evaluation and Results**

*K-fold cross validation* test with a number of folds of 5 from the validation results can produce an average accuracy as in the example in Table 3.8 below.

Table 9 K-Fold Cross Validation

Data	Accuracy	K-Fold Average
80:20	0.82	0.79
70:30	0.80	0.76
60:40	0.79	0.77

From the table above, it can be concluded that the best results are shown by 80:20 data with *k-fold cross validation* of 80%; followed by data from 70:30 results of 79% and 60:40 of 78%.

Confusion matrix testing 80:20

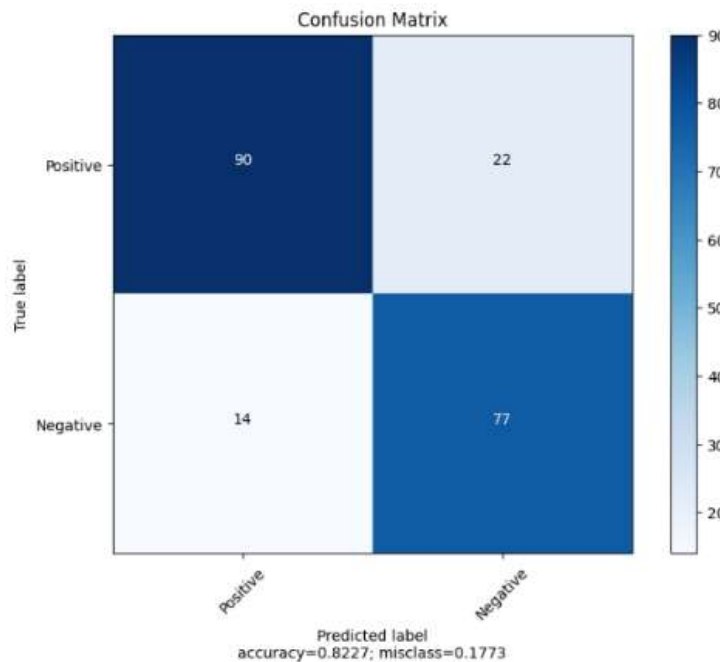


Figure 6 Confusion Matrix data 80:20

The results of the confusion matrix on 80% of the training data divided by 20% of the test data can be obtained precision, recall, f1-score, and support in table 3.9 and the accuracy obtained is 82%.

Table 10 Data 80:20

Data	Labeling	Precision	Recall	F1-score	Support
80:20	Negative	0.80	0.87	0.83	104
	Positive	0.85	0.78	0.81	99

\*name of corresponding author



Confusion matrix testing 70:30

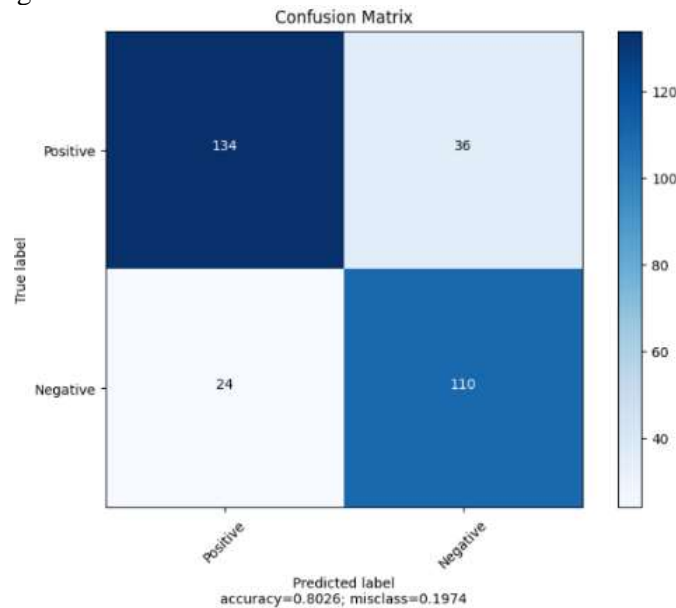


Figure 7 Confusion Matrix data 70:30

The results of the confusion matrix on 70% of the training data divided by 30% of the test data can be obtained precision, recall, f1-score, and support in table 3.10 and the accuracy obtained is 80%.

Table 11 Data 70:30

Data	Labeling	Precision	Recall	F1-score	Support
70:30	Negative	0.79	0.85	0.82	158
	Positive	0.82	0.75	0.79	146

Confusion matrix testing 60:40

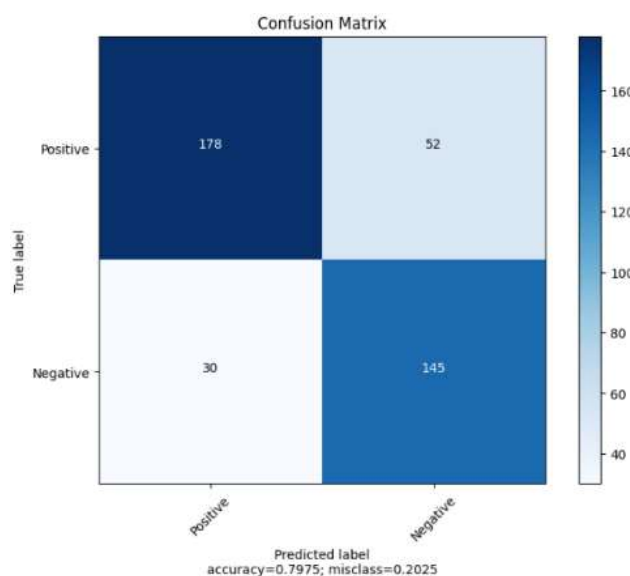


Figure 8 Confusion Matrix data 60:40

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The results of the confusion matrix on 60% of the training data divided by 40% of the test data can be obtained precision, recall, f1-score, and support in table 3.11 and the accuracy obtained is 79%.

Table 12 Data 60:40

Data	Labeling	Precision	Recall	F1-score	Support
60:40	Negative	0.77	0.86	0.81	208
	Positive	0.83	0.74	0.78	197

The Support Vector Machine test with a linear kernel, it has good results on 80:20 data with an accuracy value of 82% which is quite a high result. With the evaluation results on the Negative label, the precision value is 80%, the recall value is 87%, and the f1-score value is 83%. Meanwhile, the results on the Positive label have a precision value of 85%, a recall value of 78%, and an f1-score value of 81%. Meanwhile, the lowest results were found in 60:40 data which had an accuracy of 79% with the Negative label having a precision value of 77%, a recall value of 86%, and an f1-score value of 81%. Then the result is a Positive label with a precision value of 83%, a recall value of 74%, and an f1-score of 78%.

## CONCLUSION

In this research, we succeeded in conducting experiments with 1012 data which had gone through the preprocessing stage and had been labeled using texblob, distributed evaluations with k-fold cross validation so that we could get a sentiment analysis of the 2024 vice presidential candidate public debate using the Support Vector Machine algorithm, able to carry out classification. well on public comments on Metro TV channel YouTube videos, with negative sentiment results of 543 and positive sentiment results of 469. From the test results, the Supporting Vector Machine model has good performance on 80% training data 20% testing data with an accuracy value of 82% and the data with the weakest performance is on 70% training data 30% testing data and 60% training data 40% testing data namely with an accuracy value of 80%.

## REFERENCES

- Arsi, P., & Waluyo, R. (2021). *ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)*. 8(1), 147–156. <https://doi.org/10.25126/jtiik.202183944>
- Baita, A., & Cahyono, N. (n.d.). *ANALISIS SENTIMEN MENGENAI VAKSIN SINOVAQ MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)*.
- Furqan, M., Fakhri, A., Nasir, A., Pahang, M., Abdullah, A.-S., Id, M. A., & My, A. E. (n.d.). BIG DATA APPROACH TO SENTIMENT ANALYSIS IN MACHINE LEARNING-BASED MICROBLOGS: PERSPECTIVES OF RELIGIOUS MODERATION PUBLIC POLICY IN INDONESIA. In *Journal of Applied Engineering and Technological Science* (Vol. 5, Issue 2).
- Furqan, M., Ikhsan, M., & Aini, R. (2023). Algoritma Support Vector Machine Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Pandemi Virus Corona Di Media Sosial. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, 4(4), 908–915. [https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/241%0Afiles/5265/Furqan et al. - 2023 - Algoritma Support Vector Machine Untuk Analisis Se.pdf](https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/241%0Afiles/5265/Furqan%20et%20al.%20-%202023%20-%20Algoritma%20Support%20Vector%20Machine%20Untuk%20Analisis%20Se.pdf)
- 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>

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Iskandar, F. (2020). Analisis Wacana Politik Debat Publik Calon Presiden dan Calon Wakil Presiden Republik Indonesia. *Journal of Education, Humaniora and Social Sciences (JEHSS)*, 3(1), 31–38. <https://doi.org/10.34007/jehss.v3i1.177>
- Mualfah, D., Ramadhoni, Gunawan, R., & Mulyadipa Suratno, D. (2023). Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM. *Jurnal Fasilkom*, 13(01), 72–80. <https://doi.org/10.37859/jf.v13i01.4920>
- Nurian, A., & Nurina Sari, B. (n.d.). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3), 2830–7062. <https://doi.org/10.23960/jitet.v11i3%20s1.3348>
- Oktavia, D., & Ramadahan, Y. R. (2023). Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Media Online*, 4(1), 407–417. <https://doi.org/10.30865/klik.v4i1.1040>
- Permata Aulia, T. M., Arifin, N., & Mayasari, R. (2021). Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *SINTECH (Science and Information Technology) Journal*, 4(2), 139–145. <https://doi.org/10.31598/sintechjournal.v4i2.762>
- Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 393–399. <https://doi.org/10.29207/resti.v5i2.3008>
- Rabbani, S., Safitri, D., Rahmadhani, N., Sani, A. A. F., & Anam, M. K. (2023). Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 153–160. <https://doi.org/10.57152/malcom.v3i2.897>
- Raudya, T., Irawan, B., & Faqih, A. (2023). PERBANDINGAN IMPLEMENTASI 2 ALGORITMA SVM DAN K-NN DALAM PENGKLASIFIKASIAN KEPUASAN PENGGUNA SMART E-LEARNING. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 3).
- Saptanti, E. I. (2020). Analisis Manajemen Impresi Ma'ruf Amin dalam Debat Pilpres 2019. *Ultimacomm: Jurnal Ilmu Komunikasi*, 12(2), 262–284. <https://doi.org/10.31937/ultimacomm.v12i2.1573>
- Sriani, Suardi, & Frianda Gultom, I. (n.d.). *Analisis Sentimen Kebijakan Pemberian Subsidi Motor Listrik Menggunakan Metode Support Vector Machine*.