JOURNAL OF INFORMATICS DEVELOPMENT

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

# Analysis of Twitter Sentiment on the Implementation of Regional Elections in Indonesia During *Covid-19* Using the *Support Vector Machine Method*

Yanuar Nurdiansyah[1], Tiara Amalinda Prabawanti[2]

Department of Informatics, Faculty of Computer Science, Universitas Jember, Indonesia[1,2]

Corresponding Author: Yanuar Nurdiansyah (yanuar_pssi@unej.ac.id)

ARTICLE INFO

Date of entry:
*22 October 2025*
Revision Date:
*25 October 2025*
Date Received:
*31 October 2025*

ABSTRACT

Sentiment analysis or opinion mining is a series of problem solving based on public opinion. The opinion is in the form of text or writing in the form of documents obtained from social media. Sentiment analysis serves to determine public opinion in responding to a policy, activity or issue that is happening and being discussed, one of which is on Twitter social media. Sentiment analysis in this study focuses on the activities of the 2020 regional elections during the Covid-19 pandemic which was held on 9 December 2020. Twitter social media works in real-time, so in retrieving research data using the Trending Topic feature to retrieve research datasets. The results of the dataset are then processed using text mining techniques and used as material for analysis to determine the public's response to the implementation of the elections during covid- 19 whether it tends to have a positive or negative sentiment, as well as knowing the opinion factors that often arise. The adoption of the Support Vector Machine (SVM) method for sentiment analysis was carried out by testing the composition of various datasets. From the test results using 4 scenarios of training data and test data, namely 90:10, 80:20, 70:30, 60:40, it is obtained that the SVM method can be implemented with an accuracy value of 87% in the data scenario of 80% training data and 20% test data. Variables that affect accuracy are the amount of data, the ratio of the number of training and test data and the ratio of the number of positive and negative data used.

Keywords: Sentiment Analysis, Covid Elections, Support Vector Machine

Chite this as : Nurdiansyah, Y. (2025). Analysis of Twitter Sentiment on the Implementation of Regional Elections in Indonesia During Covid-19 Using the Support Vector Machine Method. *Journal of Informatics Development*, *4*(1), 27–38. https://doi.org/10.30741/jid.v4i1.1757

## INTRODUCTION

WHO stated that since March 2020 the spread of covid-19 cases is a global pandemic, because it has spread around the world. The Indonesian government also designated the Covid-19 case as a national disaster because the number of victims is increasing rapidly. According to data from the Covid-19 Pandemic Handling Task Force (covid19.go.id, 2020) as of November 2020, there were 538,883 positive cases of Covid-19. The determination of the Covid-19 national disaster has an impact on all aspects of people's lives, namely the policy of restricting activities in facilities and

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

public places. The policy also has an impact on government activities that will be carried out, namely the Simultaneous Regional Elections which were originally held in September 2020 and then postponed to December 9, 2020. The implementation of the Regional Elections this time the government held democracy parties in 270 regions covering 9 provinces, 224 districts, and 37 cities in Indonesia, which means that almost 60% of regions will hold Regional Elections (Supriyadi, 2020). So that the implementation of the Regional Elections during Covid-19 needs special rules compared to the implementation of the Regional Elections in the previous year. This has drawn various criticisms from the public on social media. The urgency of the implementation of the Regional Elections is the center of attention in the midst of the Covid-19 transmission crisis because some people argue that the implementation of the Regional Elections will have the potential to transmit the virus in the community, but on the other hand, the Regional Elections must still be carried out in order to get regional leaders who can handle Covid-19 optimally.

The public's response to the implementation of the Regional Elections during Covid-19 was widely expressed on social media. Social media users in Indonesia have grown very rapidly since the Covid-19 pandemic, one of which is social media Twitter which will be used in this study. Data from Data Indonesia states that Twitter users in Indonesia were 10.64 million active users in 2020 (www.dataindonesia.id, 2020). Twitter is used as a form of public expression in responding to the phenomenon that is happening. Through Twitter, people can interact with each other, exchange their thoughts and opinions about the implementation of the Regional Elections during Covid-19. Twitter is a microblogging service website that limits the size of each post to 280 characters. Twitter is one of the social media used by the government to interact with the public. In addition, twitter was chosen because of the rapid flow of information obtained by users and the ease of interaction between fellow users (Emeraldien, F. Z., et al., 2019). Twitter provides real-time updates and has a Trending Topic feature to see the most talked about trends right now. This opinion through this tweet is used by the author to analyze public sentiment regarding the implementation of the Regional Elections in Indonesia during Covid-19.

Sentiment analysis is natural language processing in the form of data processing to determine polarity and study the sentiment of each word or sentence in a document. Sentiment analysis is able to extract and process unstructured data into structured data in order to obtain information in an opinion sentence to determine the tendency of the opinion towards an issue, whether it tends to be positive, central, or negative. (Rozi, I. F., et al., 2013). The grouping of sentiment classes can describe the emotions of the individual themselves such as sad, happy, angry, or disappointed.

Text mining is the process of retrieving text data from a document that aims to find words that represent the content of the document so that it can analyze the relationship between each document (Feldman, R. and Sanger J., 2006). Text mining can also find information in the form of topics, keywords, patterns, concepts or other attributes that come from textual databases. Text mining is useful for grouping text to find a pattern in a large collection of document text data in a structured format. Text mining requires several stages in preparing text so that it can be processed into a more structured text so that it is ready to be used for the next process. Research conducted by F. J. Damanik and D. B Setyohadi in 2020 on Analysis of Public Sentiment About Covid-19 in Indonesia on Twitter Using Multinomial Naïve Bayes and Support Vector Machine results that the SVM method performs better than the naïve Bayes algorithm method in sentiment analysis. So, the author wants to make a sentiment analysis research using the support vector machine method to classify positive and negative sentiments related to the implementation of the Regional Elections during Covid-19.

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

## METHOD

The type of research used in this study is quantitative. Quantitative research is research to find knowledge or information that presents data in the form of numbers. This study uses a tweet dataset in the form of non-numerical data and then processed into numerical data through a text mining process so that the results of this numerical processing are used to analyze public sentiment towards the implementation of the Regional Elections during Covid-19 using the support vector machine classification model.

The object of research used in this study was a tweet from social media Twitter which discussed the implementation of the Regional Elections during Covid-19. The data obtained in this study was obtained directly from Trending Topics on Twitter and the use of hashtags #Pilkada2020#PilkadaSerentak2020, #TundaPilkada and #PilkadaAmanPatuhiProkes. And the total number of datasets retrieved was 3035 tweet data.

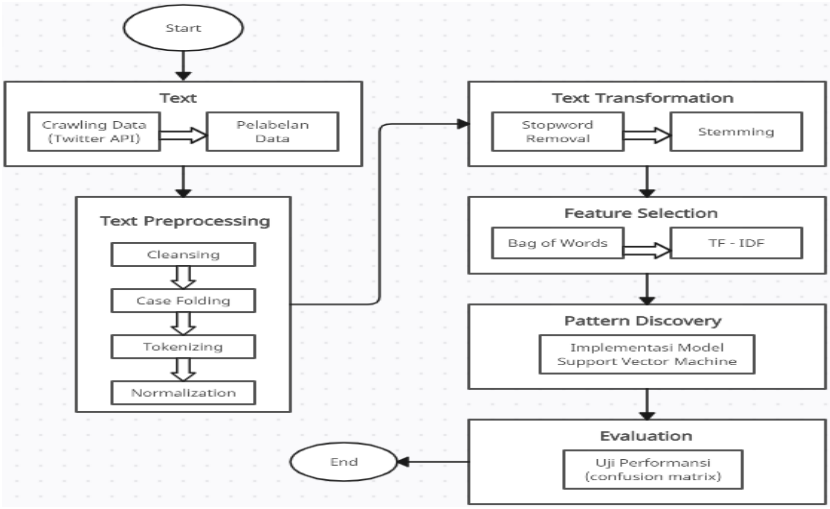


**Figure 1. Research Flow**

The first stage of text mining is text, where the data obtained is unstructured and semi-structured data in a collection of document text. At this stage, the author crawls Twitter data using the Twitter API tool by registering an account on the developer's twitter website page. After the dataset is obtained, the data is labeled into positive and negative groups with the help of linguists. Text preprocessing is the process of analyzing the correctness of the arrangement of a text in a document. In this process, the aim is to prepare the dataset into clean data that will be processed at the next stage (Feldman, R. and Sanger J., 2006). Text pre-processing needs to be done because the data obtained contains a lot of noise such as non-standard language, abbreviations, local languages, slang words, and even unclear words.

Cleansing is the process of cleaning datasets of characters that are not needed in the classification process such as symbols or punctuation, numbers, mention '@', hashtag '#', emoticons and url links in order to reduce noise and produce quality datasets. Case folding is the process of changing all the letters from 'a-z' in the dataset to lowercase or lowercase. Tokenizing is the process of breaking or cutting a sentence into units of words called tokens. The tokenizing stage involves the removal of numbers, punctuation marks or other characters that are still not lost and are considered to have no effect on text processing. Normalization is the process of identifying slang words or words alay, shortened or abbreviated words and extended words. The word alay or slang words was changed to a word that is in accordance with the Great Dictionary of the Indonesian Language (KBBI). This stage is assisted by using language dictionary tools. Text transformation is the process of changing

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

a word to its basic form and reducing the number of words in the dataset to get the expected representation of the dataset. The approach is with stopword removal and stemming.

Stopword removal is a stage of removing words that often appear in a sentence but do not have such important meanings as conjunctions (which, and, in from, etc.), tense and question words. This stage uses 2 algorithm methods, namely, stoplist (removing unimportant words) and wordlist, then the IDF value will be larger. The IDF value is calculated using the equation below:

$\mathbf{idf_t} = \boldsymbol{log_{10}}$

Information:

$$\frac{\boldsymbol{N}}{(\boldsymbol{df(t)})}$$

(save important words) that will be used in the next stage.

Stemming is the process of removing the suffix word from each word as a result of stopword removal. This process is used to return a word to its basic form or find the root word of each word using certain rules. In Indonesian texts, suffixes, prefixes and suffixes need to be eliminated (Agusta, L., 2009). There are several forms of stemming methods that are often used, one of which is stemming by Nazief-Adriani (Indonesia). Feature selection is the stage of selecting features (words). The goal of this stage is to minimize the amount of data and find the most important information from the processed dataset. In this process, the author uses word weighting features, namely Term Frequency and Inverse Document Frequency (TF-IDF). Word weighting can be in the form of a word, phrase or unit of indexing results in a sentence to find out the context of the sentence.

Term Frequency (TF) is the frequency with which a word appears in a sentence. The more often a word appears in a sentence (high TF), the greater the weight or value of its suitability. The equation of Term Frequency can be described in the equation below:

$$\boldsymbol{Wtft,d} = \{\mathbf{1} + \boldsymbol{log10}\, \boldsymbol{tft,d}, \boldsymbol{if}\ \boldsymbol{tft,d} > \mathbf{0}\}$$
$$\mathbf{0}, \boldsymbol{if}\ \boldsymbol{tf_{t,d}} = \mathbf{0}$$

Information:
$\mathbf{tf_{t,d}}$ : Frequency of the term (t) in document (d)

Document Frequency (DF) is the frequency of a document that contains words or terms that indicate how many words appear in many sentences. Inverse Document Frequency (IDF) is a calculation of the frequency with which a word appears in an entire sentence. In the IDF, the fewer words appear in the entire sentence.

**N** = Number of documents
$\mathbf{df_t}$ = Number of documents against term (t)

Term Frequency – Inverse Document Frequency (TF-IDF) is a combination of the formula for calculating TF with IDF by multiplying the value of TF by IDF as follows:
$\mathbf{W_{t,d} = Wtf_{t,d}\ x\ idf_t}$
Information:
$\mathbf{W_{t,d}}$ = Sentence weight (d) against word (t)
$\mathbf{Wtf_{t,d}}$ = Term Frequency (TF)
$\mathbf{idf_t}$ = Inverse Document Frequency (IDF)

Support vector machine is a method that can solve data classification and regression problems both linearly and non-linearly. A support vector machine is a classification algorithm between two data

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

classes and separates the classes through a solid line in the middle or a hyperplane line and this algorithm has the advantages of determining the distance of using the support vector so that it can speed up the computational process (Boser, Guyon, Vapnik, 1992). Therefore, the support vector machine can separate data linearly with 2 variables, namely x as a dataset and y as a label. This method is included in the supervised learning class, which is a class that uses the training stage using sequential training and the testing stage (Santoso, 2015). Support vector machine uses a linear model as a decision boundary with the following general shape:

$$\mathbf{y(x) = w^T f(x) + b}$$

Information:
**w** = parameter bobot **x** = vector input **f(x)** = function basis
**b** = bias value

Decision boundary (DB) is a linear model or hyperplane **y(x)** with parameters **w** and **b**, support vector machine using the concept of margin which is defined as the closest distance between DB and the trained data. By maximizing the margin, a certain DB will be obtained. The application of the SVM classification model that has been formed in the processing of training data to data testing is made to assess the accuracy of a prediction based on the dataset that has gone through the text mining process that has been made using 4 comparison scenarios between training data and data testing, namely 60:40, 70:30, 80:20, and 90:10.

Finally, the evaluation to calculate the performance of the classification model uses a confusion matrix to calculate the accuracy, precision, recall and f-measure values of the classification results that have been made in the previous process.

## RESULTS AND DISCUSSION

The data collection in this study was taken using the twitter data crawling process using Twitter API tools. Twitter API (Application Programming Interface) is a feature on Twitter's social media that is useful for creating or developing programs according to the needs of each user. Through the Twitter API, writers can dig up the data and information needed in real-time with a large amount and analyze it. Data crawling is done by registering the author's twitter account on the twitter website developer (https://developer.twitter.com/en). Twitter accounts that have been registered on the developer's twitter website will get Consumer keys in the form of API Keys and API Secret Keys, as well as Authentication Tokens in the form of Access Tokens and Access Token Secrets. The keywords used in this research process were obtained directly from Trending Topics on Twitter, namely #Pilkada2020, #PilkadaSerentak2020, #TundaPilkada and #PilkadaAmanPatuhiProkes taken in the period from November 9, 2020 to January 9 2021 and validated by linguists. The stage of labeling the tweet dataset is necessary in sentiment analysis because each sentence has a different sentiment, which can be positive or negative sentiment, even though it is in one document. The labeling of the dataset was carried out manually by the author and then validated by the Bachelor of Indonesian Language and Literature Education, University of Jember because they are considered to have more in-depth expertise in the meaning of words, sentences, language and the intent of the word arrangement so that it can provide appropriate labeling. The dataset that has gone through the labeling process produces a tweet dataset of 2000 data, each consisting of 2 sentiment classes, namely positive sentiment amounting to 1051 tweets and negative sentiment amounting to 949 tweets. The following is a sample of the datasets that have passed the data labeling process.

**Table 1. Sample Dataset Results**

| No | Tweet | Label |
|---|---|---|
| 1. | Ayo Patuhi Prokes saat Ke TPS demi terwujudnya Pilkada Serentak 2020 Sehat aman dan | Positif |

| | | |
|---|---|---|
| | damai #Pilkada2020 https://t.co/YLpfyrlj2V | |
| 2. | Pesta demokrasi atau pesta Covid? (Rakyat bertaruh nyawa, sementara Negara pesta-pora pilkada!). Masih ada waktu untuk #TundaPilkada https://t.co/YLpfyr lj2V | Negatif |

In the text preprocessing stage, dataset processing begins with the cleansing stage to clean the dataset from URL links, username (@), hashtag (#), numbers, symbols, punctuation and whitespace. The results of the cleansing process can be seen in the table below:

**Table 2. Cleaning Result**

| | | |
|---|---|---|
| 2. | Pesta demokrasi atau pesta Covid? (Rakyat bertaruh nyawa, sementara Negara pesta-pora pilkada!). Masih ada waktu untuk #TundaPilkada | Pesta demokrasi atau pesta Covid Rakyat bertaruh Nyawa sementara Negara pesta pora pilkada Masih ada waktu untuk |
| 2. | pesta demokrasi atau pesta covid rakyat bertaruh nyawa sementara negara pesta pora pilkada masih ada waktu untuk | ['pesta', 'demokrasi', 'atau', 'pesta', 'covid', 'rakyat', 'bertaruh', 'nyawa', 'sementara', 'negara', 'pesta', 'pora', 'pilkada', 'masih', 'ada', 'waktu', 'untuk'] |

The dataset that has been obtained from the data crawling process, is then labeled by the author

After the cleansing stage, the next is the casefolding stage. The casefolding stage is the stage of changing all the letters in the dataset from a-z to lowercase or lowercase. The results of the casefolding process can be seen in the table below:

**Tabel 3. Casefolding Result**

| No | Cleansing | Case Folding |
|---|---|---|
| 1. | Ayo Patuhi Prokes saat Ke TPS demi terwujudnya Pilkada Serentak Sehat aman dan damai | ayo patuhi prokes saat ke tps demi terwujudnya pilkada serentak sehat aman dan damai |
| 2. | Pesta demokrasi atau pesta Covid Rakyat bertaruh nyawa sementara Negara pesta pora pilkada Masih ada waktu untuk | pesta demokrasi atau pesta covid rakyat bertaruh nyawa sementara negara pesta pora pilkada masih ada waktu untuk |

After the casefolding stage, the next tokenizing stage. The tokenizing stage is the stage of cutting sentences into fragments of words or tokens to find patterns before undergoing the next stage. The results of the tokenizing process can be seen in the table below:

**Table 4. Tokenizing Result**

| No | Case Folding | Tokenizing |
|---|---|---|
| 1. | ayo patuhi prokes saat ke tps demi terwujudnya pilkada serentak sehat aman dan damai | ['ayo', 'patuhi', 'prokes', 'saat', 'ke', 'tps', 'demi', 'terwujudnya', 'pilkada', 'serentak', |

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

| | |
|---|---|
| | 'sehat', 'aman', 'dan', 'damai'] |

After the tokenizing stage, the next normalization stage. The normalization stage is the stage of changing non-standard words or slang words into root words in accordance with KBBI. At this stage, it is also changed to an abbreviated word or an extended word to a word that matches the form of the root word. The results of the normalization process can be seen in the table below:

**Table 5. Normalization Result**

| No | *Tokenizing* | *Normalization* |
|---|---|---|
| 1. | ['ayo', 'patuhi', 'prokes', 'saat', 'ke', 'tps', 'demi', 'terwujudnya', 'pilkada', 'serentak', 'sehat', 'aman', 'dan', 'damai'] | ['ayo', 'patuhi', 'prokes', 'saat', 'ke', 'tps', 'demi', 'terwujudnya', 'pilkada', 'serentak', 'sehat', 'aman', 'dan', 'damai'] |
| 2. | ['pesta', 'demokrasi', 'atau', 'pesta', 'covid', 'rakyat', 'bertaruh', 'nyawa', 'sementara', 'negara', 'pesta', 'pora', 'pilkada', 'masih', 'ada', 'waktu', 'untuk'] | ['pesta', 'demokrasi', 'atau', 'pesta', 'covid', 'rakyat', 'bertaruh', 'nyawa', 'sementara', 'negara', 'pesta', 'pora', 'pilkada', 'masih', 'ada', 'waktu', 'untuk'] |
| 3. | ['semua', 'warga', 'negara', 'berhak', 'memberikan', 'suaranya', 'baik', 'mereka', 'yg', 'disabilitas', 'ataupun', 'yg', 'terpapar', 'covid'] | ['semua', 'warga', 'negara', 'berhak', 'memberikan', 'suaranya', 'baik', 'mereka', 'yang', 'disabilitas', 'ataupun', 'yang', 'terpapar', 'covid'] |

After the normalization stage, the next stopword removal stage. The stopword removal stage is the stage of removing words that often appear in documents but do not have such important meanings as conjunctions (which, and, in, from, etc.), tense and question words. This stage uses 2 methods, namely, stoplist and wordlist. The results of the stopword removal process can be seen in the table below:

**Table 6. Stopword Removal Result**

| No | Normalization | Stopword |
|---|---|---|
| 1. | ['ayo', 'patuhi', 'prokes', 'saat', 'ke', 'tps', 'demi', 'terwujudnya', 'pilkada', 'serentak', 'sehat', 'aman', 'dan', 'damai'] | ['ayo', 'patuhi', 'prokes', 'tps', 'terwujudnya', 'pilkada', 'serentak', 'sehat', 'aman', 'damai'] |
| 2. | ['pesta', 'demokrasi', 'atau', 'pesta', 'covid', 'rakyat', 'bertaruh', 'nyawa', 'sementara', 'negara', 'pesta', 'pora', 'pilkada', 'masih', 'ada', 'waktu', 'untuk'] | ['pesta', 'demokrasi', 'pesta', 'covid', 'rakyat', 'bertaruh', 'nyawa', 'negara', 'pesta', 'pora', 'pilkada'] |

After the stopword removal stage, the stemming stage is followed. The stemming stage is the stage of removing suffixes, inserts, prefixes, and suffix prefixes from the dataset. The results of the voting process can be seen in the table below:

**Table 7. Stemming Result**

| No. | *Stopword* | *Stemming* |
|---|---|---|
| 1. | ['ayo', 'patuhi', 'prokes', 'tps', 'terwujudnya', 'pilkada', 'serentak', | ayo patuh prokes tps wujud pilkada serentak sehat aman damai |

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

| | | |
|---|---|---|
| | 'sehat', 'aman', 'damai'] | |
| 2. | ['pesta', 'demokrasi', 'pesta', 'covid', 'rakyat', 'bertaruh', 'nyawa', 'negara', 'pesta', 'pora', 'pilkada'] | pesta demokrasi pesta covid rakyat taruh nyawa negara pesta pora pilkada |

Word weighting in this study consists of 2 stages, the first is the weighting of the word Bag of Words and the second is Term Frequency- Inverse Document Frequency (TF-IDF). Bag of Words is a form of simple feature extraction depiction of an unstructured text into numbers. The second stage of word weighting is TF-IDF. This stage is carried out to find out how relevant a term is in the dataset by providing word weighting after going through the text preprocessing stage which will then be used in the classification stage. The TF-IDF stage begins with determining the Term Frequency (TF) value. The following is a sample of the calculation of the TF value:

**Table 8. Results of TF and DF Value Calculation**

| Token | tf | | df | D/df |
|---|---|---|---|---|
| | D1 | D2 | | |
| daerah | 1 | 0 | 1 | 2 |
| zona | 1 | 0 | 1 | 2 |
| merah | 1 | 0 | 1 | 2 |
| selenggara | 1 | 0 | 1 | 2 |
| pilkada | 1 | 1 | 2 | 1 |
| tambah | 1 | 0 | 1 | 2 |
| ayo | 0 | 1 | 1 | 2 |
| patuh | 0 | 1 | 1 | 2 |
| prokes | 0 | 1 | 1 | 2 |

Each word in the sample table above will be assessed as the word weight of each document, so a word will be worth 1 when it appears in a document. Adding a word weight will add to the weight of the D value of a word. If all words in the document have been valued, then the document frequency (df) value is calculated. The df value is the sum of the weight of words in the entire document. The value df will be the divisor of the total document (D) to be analyzed. The number of documents used is as many as 2 documents so that D/df or 2 is divided by the df value of each word. After completing the calculation of the TF and df values, the inverse document frequency (IDF) value and W value will be calculated.

**Table 9. Results of Inverse Document Frequency (IDF) and W Value Calculation**

| **IDF=(log D/df)** | **W** | |
|---|---|---|
| | **D1** | **D2** |
| 0,301029996 | 0,301029996 | 0 |
| 0,301029996 | 0,301029996 | 0 |
| 0,301029996 | 0,301029996 | 0 |
| 0,301029996 | 0,301029996 | 0 |
| 0 | 0 | 0 |
| 0,301029996 | 0,301029996 | 0 |
| 0,301029996 | 0 | 0,301029996 |
| 0,301029996 | 0 | 0,301029996 |
| 0,301029996 | 0 | 0,301029996 |

The IDF value is the result of the D/df log. Next, the value of the TF-IDF weight or commonly called the W value is calculated. Finally, add up the total of the total value of W for each existing document.

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

In the model classification process, the first stage is to split the dataset. The dataset sharing process was carried out in 4 scenarios, namely using a scenario of 60% data training and 40% data testing, the second is 70% training data and 30% data testing, the third is 80% training data and 20% data testing and the fourth is 90% training data and 10% data testing. After the training and data testing process is carried out, an evaluation test is carried out using a confusion matrix to determine the accuracy value of the model classification process that has been processed previously. In the implementation of the calculation of the confusion matrix value, the author used the calculation of accuracy, precision, recall, and f-measure values. The results of the evaluation test stage can be seen in table 10 below:

**Tabel 10. Confuion Matrix Result**

| Information | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 |
|---|---|---|---|---|
| Data Training | 60% | 70% | 80% | 90% |
| Data Testing | 40% | 30% | 20% | 10% |
| True Positif (TP) | 372 | 264 | 171 | 94 |
| True Negative (TN) | 319 | 250 | 176 | 78 |
| False Positif (FP) | 14 | 14 | 10 | 6 |
| False Negative (FN) | 95 | 72 | 43 | 22 |
| Akurasi | 86% | 86% | 87% | 86% |
| Presisi | 96% | 95% | 95% | 93% |
| Recall | 77% | 78% | 80% | 78% |
| F1-measure | 85% | 85% | 87% | 85% |

Of the four scenarios that have been carried out in the model classification and evaluation test process, the best results were obtained in the third scenario, namely with 80% training data and 20% data testing with an accuracy of 87%.

**Sentiment Analysis Results**

This study was conducted using a dataset of 2000 tweet data that has been labeled sentiment by the author and validated by linguists, Figure 2 below is a comparison of positive and negative responses before the text mining process and before being modeled with the support vector machine method.

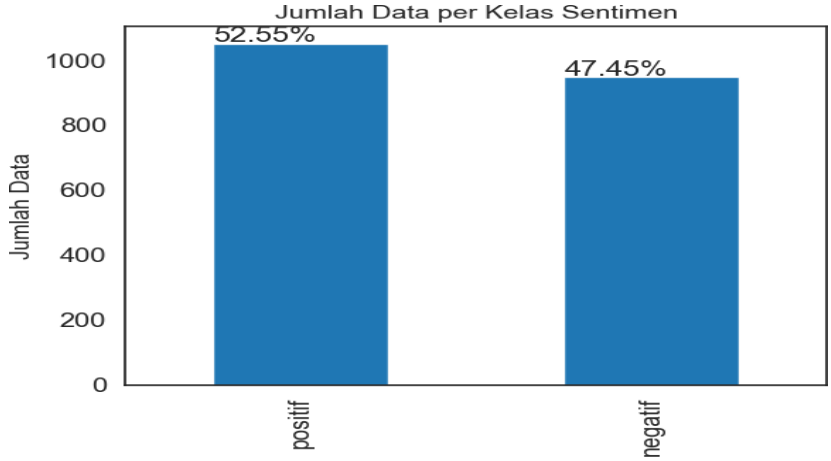


**Figure 2. Comparison of Sentiment Before Modeling**

In figure 2 above, it can be seen that the positive sentiment is 1051 (52.55%) and negative sentiment as many as 949 (47.45%). So that the public response is more supportive of the implementation of the simultaneous Regional Elections during Covid-19 with a difference of 5% with the response of the public who do not support and want to postpone the implementation of the Regional Elections. Furthermore, in the graph image below, a comparison of positive and negative sentiments is aimed after passing the text mining process and modeling is carried out using a support vector machine.

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

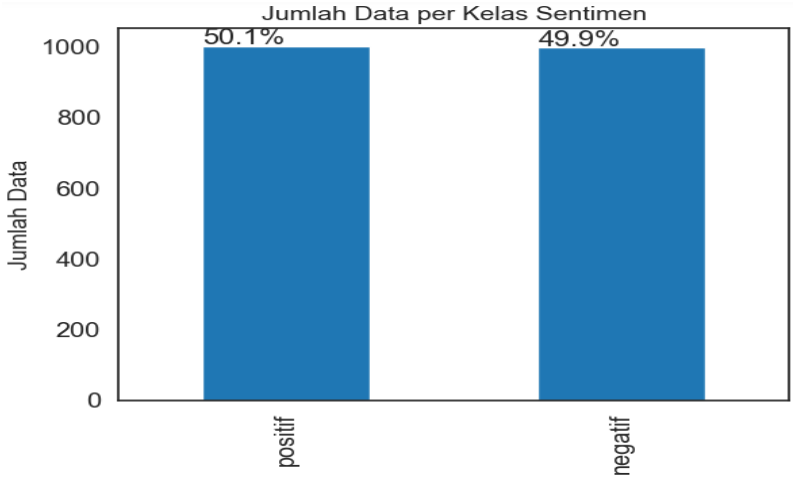


**Figure 3. Comparison of Sentiment After Modeling**

After going through the text mining process and SVM modeling, the positive sentiment generated was 1004 (50.23%) and sentiment negative by 995 (49.77%). It can be seen that positive sentiment is still greater than negative sentiment and the comparison between positive and negative sentiment is not very different and almost balanced. So it can be concluded that 50.23% of people on social media twitter support the implementation of simultaneous regional elections during Covid-19 with the implementation of strict health protocols. On the other hand, 49.77% of the public also want the postponement of the Regional Elections until the condition of the Covid-19 pandemic gradually improves. Wordcloud is an image that displays a list of words used in a document text, generally the more often the word is used in a document, the larger the size of the word in the image (Widodo, 2021). The following is a wordcloud of the results of sentiment analysis in this study.


**Figure 4. Wordcloud Positive Sentiment**

In figure 4 above is a wordcloud with a positive sentiment, it can be seen that the words that are often used by twitter users are "healthy protocol", "protocol compliance", "protocol application", "support the regional elections", "protocol discipline" and many other words of support for the implementation of the regional elections during covid-19 by implementing strict health protocols for both polling station officers and the public who vote. So it can be concluded that the community's response to the implementation of the Regional Elections is quite good on the condition that they

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

carry out strict health protocols, wear masks, carry out vaccines, and officers conduct rapid or PCR tests. The community also needs to carry out the appeal given by the government by maintaining distance, washing hands and doing 3M.


**Figure 5. Wordcloud Negative Sentiment**

While in figure 5 is wordcloud which has a negative sentiment, it can be seen that the words that are often used by twitter users are the words "pandemic", "people", "politics", "covid", "pilkada", "postpone the regional elections", "cluster", "choose orders". The difference between the percentage of positive and negative sentiment is not too far, so it can be concluded that the public also still feels afraid and anxious about the implementation of the Regional Elections which were carried out when the Covid-19 curve was still high. Many people also want the postponement of the regional elections until Covid-19 subsides and the transmission can be controlled. One of them is the evenly distributed Covid-19 vaccination in Indonesia and the decline in the death rate due to Covid-19.

## CONCLUSION

The conclusion of this study is to analyze the sentiment of the implementation of regional elections in Indonesia during Covid-19 using social media twitter with a vector machine support algorithm, a dataset used of 2000 tweet data through the text mining process, namely through the cleansing, casefolding, tokenizing, and normalization stages. After that, it goes through the stopword removal and stemming stage. After the dataset is clean, then calculate the weight of words in the document using 2 feature extractions, namely Bag of Words and TF-IDF. Furthermore, the dataset is processed using a support vector machine algorithm which is tested using performance testing from the model that has been formed using a confusion matrix.

The implementation of the support vector machine method will be optimal if the training data used has a large number of datasets. After conducting a performance test by conducting 4 test scenarios, the first is 60% training data and 40% data testing, the second scenario is 70% training data and 30% data testing, the third is 80% training data and 20% data testing and the fourth scenario is 90% l training data and 10% data testing. This will result in different confusion matrix values, accuracy levels, precision values, recall values and f1-measures. Of all the scenarios that have been carried out, the highest accuracy score was obtained, which is 87% in the second scenario, namely 80% data training and 20% data testing.

The results of the sentiment analysis classification resulted in a higher positive sentiment classification of 50.1% than the negative sentiment class of 49.9%. From the results of dataset

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

processing using text mining and the SVM method, it can be seen that the community's response to the implementation of the simultaneous Regional Elections during Covid-19 is almost balanced. So it can be concluded that the Indonesian people consider that the implementation of simultaneous regional elections during Covid-19 needs to be carried out because the Regional Elections will produce regional leaders who can handle the pandemic optimally, but with the note that the implementation of regional elections needs to implement good health protocols according to the recommendations of the Ministry of Health and WHO. However, the public also still feels anxious and afraid because the implementation of simultaneous regional elections during Covid-19 will open up the potential for coronavirus transmission in the community.

## REFERENCES

Agusta, Ledy. 2009. Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. Bali, Indonesia. Konferensi Nasional Sistem dan Informasi, KNS & I09-036

Boser, B. E., Guyon, I. M., Vapnik, V.N. 1992. A Training Algorithm for Optimal Margin Classifiers. Pittsburgh. Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92), 144-152.

Covid19.go.id. 2020. Peta Sebaran Covid-19 di Indonesia dan Analisis Data Covid-19 di Indonesia[https://covid19.go.id/]. Diakses pada Januari 2021

Damanik, F. J., D. B. Setyohadi. 2021. Analysis of Public Sentiment About Covid-19 In Indonesia on Twitter Using Multinomial Naïve Baye And Support Vector Machine. Briston, Britania Raya. IOP Publishing

Dataindonesia.id. 2020. Pengguna Twitter di Indonesia Capai 10,64 Juta pada 2020[https://dataindonesia.id/]. Diakses pada Januari 2021.

Emeraldien, F. Z., Sunarsono, R. J., & Alit, R. (2019). Twitter Sebagai Platform Komunikasi Politik di Indonesia. XIV.

Feldman, R., & Sanger, J. (2006). The Text Mining Handbook: Advanced Unstructured Data. New York: Cambridge University Press.

Rozi, I. F., Pramono, S. H., & Dahlan E. A. 2013. Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. Malang. Jurnal EECCIS Vol. 6, No. 1.

Santoso, S. (2015). SPSS20 Pengolahan Data Statistik di Era Informasi. Jakarta. PT. Alex Media Komputindo, Kelompok Gramedia.

Supriyadi. 2020. Menakar Nilai Keadilan Penyelenggaraan Pilkada 2020 di tengah Pandemi Covid-19. Palu. Kanun Jurnal Ilmu Hukum Vol. 22, No. 3, pp. 93-514.

Widodo. 2021. "Membuat Wordcloud Dengan Python" https://sites.unpad.ac.id/ membuat-wordcloud-d

Trismanto, T. (2018). Ambiguitas dalam bahasa Indonesia. *Bangun Rekaprima*, *4*(1), 42–48.

Wulandari, D. A., Widagdo, A., Shafira, H., Maulida, A., Wardani, R. W., Sarnita, S., & Rahmawati, A. (2025). Analisis Penerapan Strategi Pembelajaran Bilingual pada Video Praktik Pembelajaran di Sekolah Dasar. *Jurnal Inovasi Penelitian Ilmu Pendidikan Indonesia*, 177–183.