Journal of Engineering, Electrical and Informatics

E-ISSN: 2809-8706
P-ISSN: 2810-0557

Research Article

Implementation of the Naive Bayes Algorithm and Support Vector Machine for Public Sentiment Analysis towards the Ratification of the Job Creation Bill on Twitter

Untung Surapati [1*], Sopan Adrianto [2], Erno Sumantri [3], Melinius Nopianto [4,]

[1] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki) Jakarta
[2] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki) Jakarta
[3] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki) Jakarta
[4] Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki) Jakarta
* Corresponding Author: e-mail : frencis@stikomcki.ac.id

Abstract: The test design of the Public Sentiment Analysis on the Ratification of the Job Creation Bill with the RapidMiner Studio application. The initial stage is to collect data in the form of tweets of Twitter users and then put it into a CSV file, the data obtained will be divided into training data and test data. Furthermore, the training data will be labeled consisting of 2 types of labels, namely Positive and Negative labels, then the data will be cleaned from unneeded words such as Mention or Hastag, then the data will go through several stages in the Preprocessing stage to convert raw data into data that is ready to be processed. Furthermore, each word will be weighted with the TF-IDF method. The final result of the comparison with these two test methods, namely the prediction of Public Sentiment Towards the Issue of Determining the Job Creation Bill based on data obtained from Twitter and implemented by the SVM (Support Vector Machine) method, showed an accuracy value of 96.52%. Of the 605 test data, 492 data were predicted as Negative Sentiment and 112 data as Positive Sentiment and the Naive Bayes Method showed an accuracy value of 49.67%. Of the 605 test data, 492 data were predicted as Negative Sentiment and 112 data as Positive Sentiment.

Keywords: Naive Bayes; RapidMiner; RUU Cipta Kerja; Sentiment Analysis; Support Vector Machine.

1. Introduction

The Job Creation Law (Omnibus Law) is a policy prepared by the Indonesian government with the aim of improving the investment climate, expanding job opportunities, and encouraging economic growth through simplification of various regulations that are considered to hinder business activities. From the discussion process to its ratification in 2020, this policy has generated various responses from the public. Some parties support the implementation of the Job Creation Law because it is considered to be able to increase investment competitiveness and open up wider job opportunities. However, some others expressed their rejection due to concerns about labor protection, environmental sustainability, and the potential for economic inequality that could arise due to the change in regulations [1], [2].

These differences of views are widely expressed through social media, especially Twitter, which is one of the main platforms for the public to express their opinions on various public policies. Twitter allows users to express their opinions quickly and openly, resulting in data that can be used to understand public perception of an issue. One form of public response to the ratification of the Job Creation Law is shown through various uploads and digital campaigns that use certain hashtags as a means of expressing support and rejection of the policy [3], [4].

Received: December 15, 2021
Revised: January 10, 2022
Accepted: January 25, 2022
Published: February 16, 2022
Curr. Ver.: February 28, 2022

Sentiment analysis is one of the approaches that can be used to identify and group public opinion into positive and negative sentiment categories based on textual data obtained from social media. Through sentiment analysis, the information contained in public opinion can be processed into useful knowledge to understand the tendency of public perception towards a government policy. In addition, the results of sentiment analysis can be used as evaluation material for policymakers in assessing the social impact caused by a regulation [5], [6].

Various previous studies have applied sentiment analysis methods to examine public opinion on the Job Creation Law. [2] conducted an analysis of public opinion sentiment regarding the Job Creation Law on Twitter, while [1] applied the Backpropagation and Term Frequency–Inverse Document Frequency (TF-IDF) algorithms to identify public sentiment. Another study conducted by [4] used the Support Vector Machine (SVM) method in analyzing sentiment related to the Omnibus Law, while [7] compared the performance of the Naive Bayes Classifier (NBC) and the Support Vector Machine (SVM) in the same case.

Although various studies have been conducted, the comparison of the implementation of the Naive Bayes algorithm and the Support Vector Machine in identifying public sentiment towards the ratification of the Job Creation Law is still an interesting topic to study. Both algorithms are classification methods that are widely used in sentiment analysis research because they have different characteristics and levels of accuracy on different types of text data [8], [9] Therefore, research is needed that can evaluate the ability of the two algorithms to classify positive and negative sentiments based on public opinion conveyed through Twitter.

This study aims to analyze the public's response to the ratification of the Job Creation Law through Twitter data using the Naive Bayes algorithm and the Support Vector Machine. In addition, this study aims to compare the ability of the two algorithms in identifying positive and negative sentiments so that they can provide an overview of more effective methods to be used in sentiment analysis on public policy issues. The results of the research are expected to contribute to understanding public perception of the Job Creation Law and become a reference for the development of sentiment analysis research on social media in the future.

2. Literature Review

Systematic Literature Review

Methodology Survey

This study uses the Systematic Literature Review (SLR) approach to identify previous studies that are relevant to sentiment analysis on social media Twitter using the Naive Bayes algorithm and the Support Vector Machine (SVM). The literature identification process was carried out using the PICOC (Population, Intervention, Comparison, Outcomes, and Context) framework to ensure the suitability of the research used as a reference.

Table 1. PICOC Research Framework.

Components	Description
Population	Sentiment analysis on Twitter social media using Support Vector Machine and Naive Bayes methods.
Intervention	Collection and processing of tweet data from Twitter for sentiment analysis purposes.
Comparison	-
Outcomes	Classification of positive and negative sentiment and the accuracy level of the classification model.
Context	The public dataset is obtained through Twitter's crawling process using RapidMiner Studio.

The PICOC framework is used to assist the literature search process in accordance with the focus of the research, namely the analysis of public sentiment towards the ratification of the Job Creation Bill on social media Twitter.

Survei Protocol

The literature search process is carried out by establishing a search protocol consisting of the year of publication, the type of publication, the search keyword, and the number of articles selected as research references.

Tabel 2. Review Survei Protocol.

Criteria	Description
Publication Year	2018–2023
Publication Type	Scientific Journal
Search String	Sentiment Analysis SVM, Naive Bayes Sentiment Analysis, Twitter Sentiment Analysis
Final Selected	21 Journal Articles

Based on the protocol, a number of studies were obtained that discussed the implementation of the Naive Bayes algorithm and Support Vector Machine in various cases of sentiment analysis, including the analysis of public opinion on government policies and the Job Creation Law.

Research Gap Analysis

Research on sentiment analysis on the Job Creation Law has been conducted by various researchers using different approaches and algorithms. [2] analyzed public opinion on the Job Creation Law through Twitter, while [1] applied the Backpropagation and TF-IDF algorithms to identify public sentiment. Other studies used the Support Vector Machine [4], Naive Bayes [10], or a combination of the two [7].

Although there have been many studies that discuss sentiment towards the Job Creation Law, there is still a need to compare the performance of the Naive Bayes algorithm and the Support Vector Machine on Twitter's dataset related to the passage of the Job Creation Bill. Therefore, this study focuses on the implementation and comparison of the two algorithms to find out the model that performs best in the classification of positive and negative sentiments.

In conducting a search from previous research sources, stages are carried out in obtaining research journals that have been published for which a Study Selection Strategy or Studies Selection Strategy is carried out, namely carrying out the following stages:

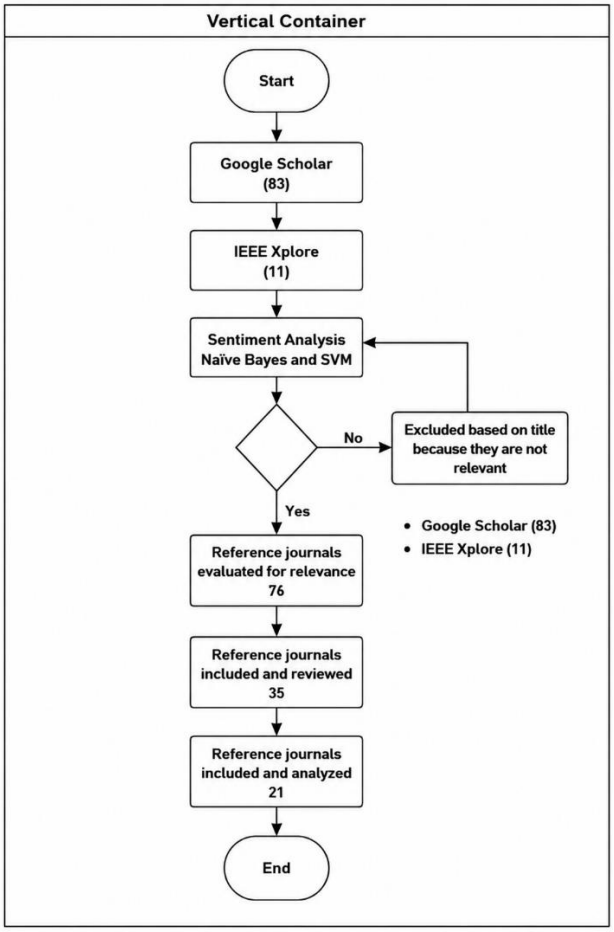

Figure 1. Studies Selection Strategy.

Theoretical Background

Sentiment Analysis

Sentiment analysis or opinion mining is a branch of text mining that aims to identify and classify the opinions contained in a text into positive, negative, or neutral categories. Sentiment analysis is widely used to understand public perception of a product, service, organization, or public policy [5], [6].

In general, the stages of sentiment analysis include data collection, preprocessing, feature transformation, feature selection, classification, and model evaluation. The evaluation stage typically uses accuracy, precision, recall, and F1-score metrics to measure the performance of the classification model [1].

Job Creation Bill

The Job Creation Law is a regulation that is prepared using the concept of Omnibus Law, which is a law-making approach that combines various legal provisions into one regulation. This policy aims to increase investment, expand employment, and simplify regulations that are considered to hinder economic growth. However, the passage of the Job Creation Law has generated various public responses, both in the form of support and rejection [2], [11].

Twitter as Social Media Platform

Twitter is one of the social media that is widely used to convey opinions, criticisms, and responses to various public issues. Twitter's characteristic of allowing users to express their opinions openly makes the platform a relevant data source for sentiment analysis research [3], [4].

User-generated data in the form of tweets can be used to understand public perceptions and responses to a policy, including the ratification of the Job Creation Law.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a classification algorithm that works by looking for an optimal hyperplane capable of separating data into different classes. SVMs are known to have good capabilities in handling high-dimensional data, including feature-extracted text data [4], [12].

Naive Bayes

Naive Bayes is a probability-based classification algorithm that uses Bayes' Theorem to determine the class of a data. This method is widely used in sentiment analysis because it is simple, fast, and effective in processing text data [10], [13].

Text Mining

Text mining is the process of extracting information from unstructured text data with the aim of finding certain patterns, concepts, or relationships that are useful for decision-making. In sentiment analysis, text mining plays a role in converting text data into numerical representations that can be processed by classification algorithms [5].

Twitter API

Twitter's Application Programming Interface (API) is a service that allows developers to automatically access Twitter data for research and application development purposes. In sentiment analysis research, the Twitter API is used to collect tweet data which is then processed and analyzed using a specific classification method [14].

3. Materials and Method

Research Data

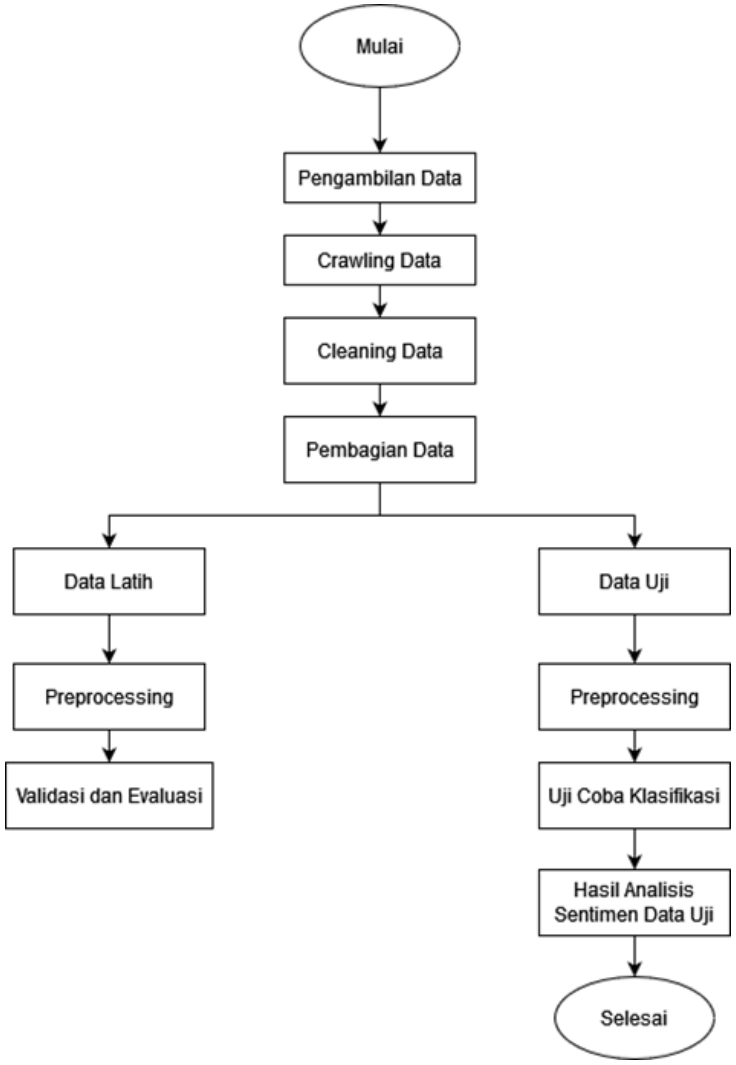

Figure 2. Research stages.

Research Stages

This research begins with a series of general stages in research consisting of stages of literature study to be used as a reference or reference in this research to strengthen the problem and as a theoretical basis in conducting research. Literature studies are obtained from reading journals, articles and websites on the internet so as to get a collection of references that are relevant to the problems in this study. After conducting a literature study, the researcher identifies the problem to be researched to become the object of research.

Data Collection Methods

The data collection process in this study will use public set data. The public set data is obtained by Crawling Tweet Data of Twitter users.

Research Classes

In this research class, the researcher created a system to conduct sentiment analysis, while in this study using the Naive Bayes algorithm and the Support Vector Machine. To make it easier in this study, the researcher used the RapidMiner Studio program to conduct Sentiment Analysis. The results of the sentiment analysis will be divided into 2, namely Positive Sentiment and Negative Sentiment.

Examples of Positive and Negative Sentiments used in this research class:

Table 3. Research Class.

No.	Classes	Tweet
1.	Positive Sentiment	**InstitutEcosocRights** @ecosocrights · May 8 Akan lebih baik kalau publik jg dishare ttg isu2 yg ada di **RUU** Kesehatan, sehingga tahu apa sebenarnya yg dipermasalahkan& apa yg akan berubah. Jangan sampai yg terjadi pd omnibus **cipta kerja** terulang di **RUU** Kesehatan. Kalau penguatan layanan primer siapa yg tdk mendukung.
2.	Sentimen Negatif	**ferizandra** @ferizandra · May 9 Replying to @KangSemproel and @DrEvaChaniago Jangan sampe **RUU** Kesehatan dibahas & disetujui DPR pada tengah malam di akhir minggu spt RUU Cipta Kerja...

Application of Methodology

The Naive Bayes method and SVM (Support Vector Machine) are two algorithms that are often used in the field of Machine Learning. The following is the application of the methodology using these two methods: The methodology used in this study uses the Support Vector Machine classification method. The software used in conducting sentiment analysis is RapidMiner Studio.

Data Collection

The data collection process in this study will use public set data. The public set data was obtained from the Tweets of Twitter users using the "Search Twitter" Operator in RapidMiner Studio with the keyword "Job Creation Bill" as many as 1000 data which were then filtered to remove duplicate tweets, then the data was converted into CSV form.

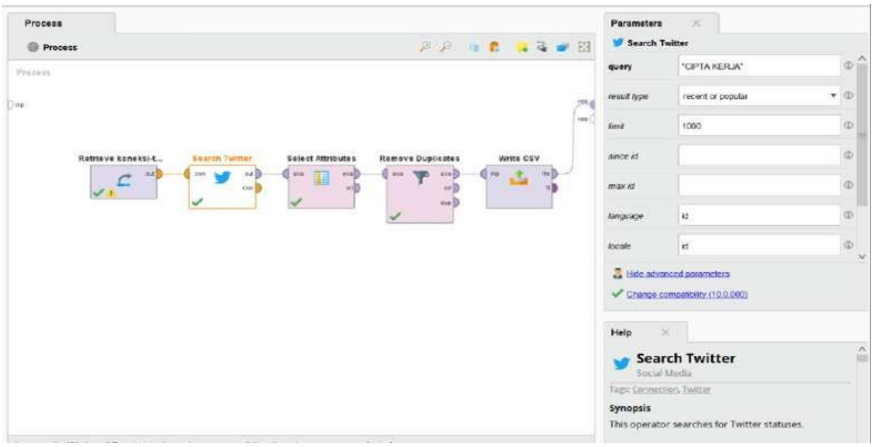

Figure 3. Data Collection Process.

In the process of collecting and processing Twitter data using RapidMiner, several operators are used to support the sentiment analysis stage. The Twitter Retrieve operator functions to retrieve public data from Twitter through an Application Programming Interface (API) connection that has been connected to the user's account. This operator allows researchers to determine search parameters such as keywords, time ranges, locations, and certain accounts so that the data obtained is more relevant to the research needs. Furthermore, the Twitter Search operator is used to filter data based on certain criteria, such as the date of upload, location, number of retweets, and number of followers. After the data is collected, a preprocessing stage is carried out which includes cleaning up unnecessary characters, normalizing text, and deleting stopwords. To optimize data quality, the Select Attributes operator is used in the feature selection process by selecting relevant attributes for analysis, while the Remove Duplicates operator is used to eliminate duplicate data so that the quality of the dataset is improved. The final result of the data processing process is then stored

using the Write CSV operator, which allows the data to be exported into Comma Separated Values (CSV) format for use in the later stages of analysis and modeling.

Data Labeling

A total of 418 data will be labeled manually which will later be used as training data. To reduce subjective assessment, the labeling process was carried out with the help of the author's wife. The labeling of the data is divided into 2 labels, namely the "Positive" label and the "Negative" label.

Dataset Labeling Criteria

The dataset labeling process is carried out manually by the researcher to produce training data used in the sentiment classification process. The labeling is carried out based on the interpretation of the content of the tweet related to the ratification of the Job Creation Bill. Each tweet is categorized into two classes of sentiment, namely positive and negative. Positive sentiment is given to tweets that contain positive words, such as "good", "competent", "amazing", or other expressions that show support for the Job Creation Bill policy. In addition, tweets that indicate approval or support for the process and implementation of the Job Creation Bill are also classified as positive sentiments. In contrast, negative sentiment is given to tweets that contain negative-toned words, such as "bad," "embarrassing," "terrible," and other expressions that indicate dissatisfaction or rejection. The negative category is also given to tweets that contain opinions of rejection or disapproval of the process and implementation of the Job Creation Bill. This labeling aims to produce a dataset that can be used as a basis for learning the Naive Bayes model and the Support Vector Machine in conducting sentiment classification.

Cleaning Data

The Data Cleaning Stage is a stage to clean the Tweet of unnecessary words such as the hashtag character "#", mention "@", removing url-urls or symbols that are not needed in the sentiment analysis process.

Data Preprocessing

The preprocessing stage aims to convert the raw data from the collection of tweets into data that is ready to be used in the sentiment classification process. This process begins with tokenizing, which is breaking the text into word units or tokens, for example, the sentence "working on a thesis" is changed to "medium", "working", and "thesis" tokens. Next, case folding is carried out to change all characters to lowercase (lowercase) so that the data format becomes uniform. The next stage is stemming, which is changing the adjective word into a root word, for example the word "do" to "work". After that, a token by length filter is carried out to remove words that are less than two characters long or more than twenty-five characters to reduce noise in the data. Finally, a stopword filter is carried out to eliminate common words that do not have a significant contribution to the classification process, such as "me", "me", "he", and "them". Through this preprocessing stage, the quality of text data is improved so that it can improve the performance of the Naive Bayes model and the Support Vector Machine in conducting sentiment analysis.

Testing Plans

The test design of the Public Sentiment Analysis on the Ratification of the Job Creation Bill with the RapidMiner Studio application. The initial stage is to collect data in the form of tweets of Twitter users and then put it into a CSV file, the data obtained will be divided into training data and test data. Furthermore, the training data will be labeled consisting of 2 types of labels, namely Positive and Negative labels, then the data will be cleaned from unneeded words such as Mention or Hastag, then the data will go through several stages in the Preprocessing stage to convert raw data into data that is ready to be processed. Furthermore, each word will be weighted with the TF-IDF method. Next, classification modeling will be made using the Naive Bayes Algorithm and Support Vector Machine so that the model can be applied to the test data.

Research Roadmap

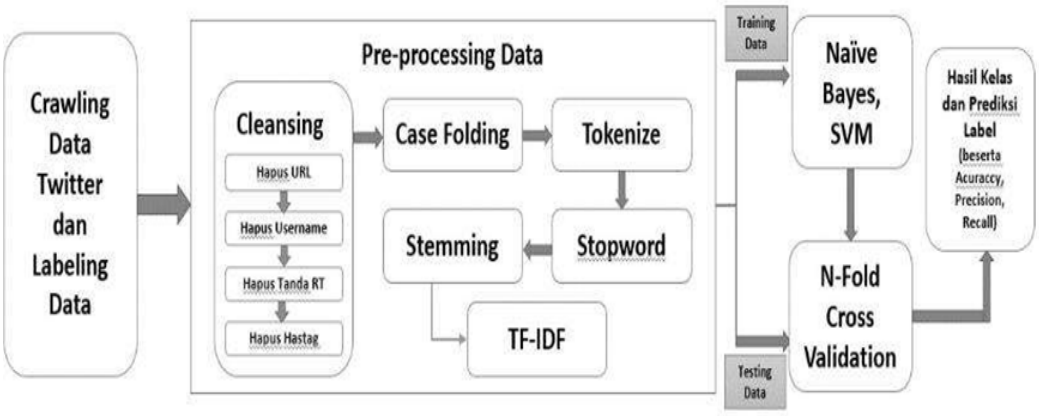

Figure 4. Research Roadmap.

Time Line Penelitian

Table 4. Research Timeline.

NO.	Sequence of Activities	Implementation Time				
		Maret	April	Mei	Juni	Juli
1.	Collect Data					
2.	Processing Data					
3.	Rancangan Testing					
4.	Testing					
5.	Implementation					

4. Results and Discussion

Research Tools

The tools and materials used in this study include: a.) Microsoft Excel is used to label sentiment from the results of data crawling on social media Twitter to be used in training data, b.) RapidminerStudio is a tool or tools used to implement the text mining method used in this study, c.) TF-IDF (Term Frequency - Inverse Document Frequency) algorithm as an algorithm to calculate the weight of each word in the training data and test data. d.) Naive Bayes Algorithm and SVM (Support Vector Machine) as sentiment classification algorithms in this study.

Implementation and testing

Stages of Data Collection

Data is collected from Twitter's social media platform using a data collection tool called an API. The data collected includes user comments on different Job Creation Bill Determination Tweets. The data taken were 1000 comments taken from the popular Job Creation Bill Recruitment Tweet. The data collection process in this study will use public set data. The public set data was obtained from the Tweets of Twitter users using the "Search Twitter" Operator in RapidMiner Studio with the keyword "Job Creation Bill" as many as 1000 data which were then filtered to remove duplicate tweets, then the data was converted into CSV form. This process resulted in 605 tweets.

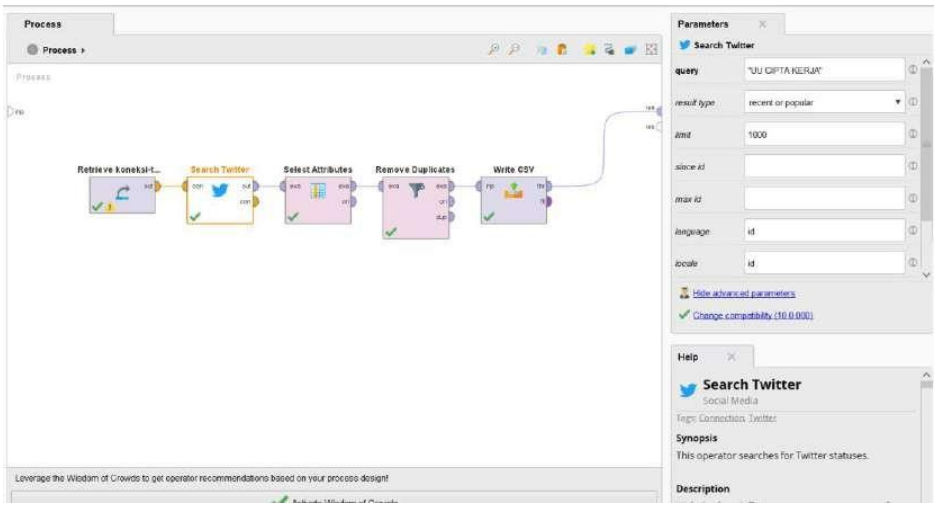

Figure 5. Data crawling process.

Data Labeling Levels

A total of 605 data will be labeled manually which will later be used as training data. To reduce subjective assessment, the labeling process was carried out with the help of the author's wife. Data labeling is divided into 2 labels, namely the "Positive" label and the "Negative" label.

	A	B	C	D	E	F	G
1	Text	SENTIMEN	COUNT	300			
2	Ajukan Gugatan UU Cipta Kerja, Jumhur Hidayat: MK Harus Hentikan Petualangan Presiden https://t.co/HwMCNRHqz8	Positive					
3	Kaum buruh masih terus berjuang lewat berbagai lini agar UU Cipta Kerja dicabut. Selain lewat jalur hukum, aksi besar-besaran yang melibatkan ratusan ribu buruh di 38 provinsi pun disiapkan secara bergelombang	Positive					
4	bertujuan untuk penguatan UMKM	Negative					
5	implementasi UU Cipta Kerja dipastikan akan meningkatkan kesejahteraan rakyat dalam upaya pemulihan ekonomi nasional.	Positive					

Figure 6. Levels of Data Labeling.

Data Cleaning Stage

This stage aims to clean the Tweet of unnecessary words such as hashtag characters "#", mention "@", discard url-urls or symbols that are not needed in the sentiment analysis process.

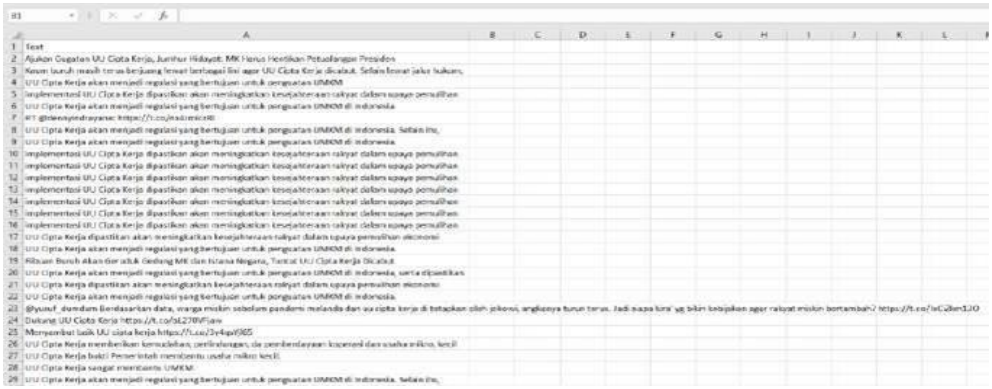

Figure 7. Before Cleaning Data.

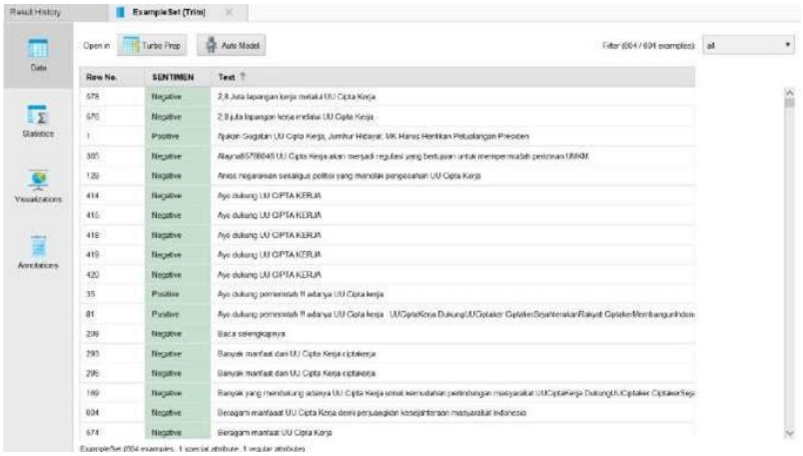

Figure 8. After the Data Cleansing Process.

Data Preprocessing

The preprocessing stage is carried out to convert the raw data from Twitter crawls into data that is ready to be used in the sentiment classification process. This stage begins with tokenizing, which is the process of breaking down tweet text into tokens or single words so that it is easier to analyze. Next, case folding is carried out to change all characters to lowercase letters (lowercase) to standardize the data format. After that, a stemming process is carried out to change the suffixing word into the root word, so that variations of words that have the same meaning can be represented in a uniform form. The next stage is the filter tokens by length, which aims to remove words with a character count of less than two letters or more than twenty-five letters to reduce noise in the dataset. In addition, stopwords filters are carried out to eliminate common words that do not have a significant contribution to the classification process, such as pronouns and conjunctions. The result of this series of preprocessing processes is a cleaner, structured, and ready-to-use dataset at the feature weighting stage as well as the classification process using the Naive Bayes algorithm and the Support Vector Machine.

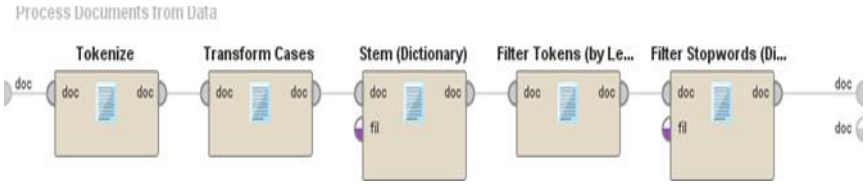

Figure 9. Preprocessing process.

Figure 9. shows the preprocessing stages applied in this study using RapidMiner, starting from the tokenizing process, case folding, stemming, filter tokens by length, to filter stopwords to produce text data that is ready to be processed at the sentiment analysis stage.

Word Weighting Stage

At this stage, the preprocessing results will be processed so that each word has a weight (value). The word weighting used by the author is the TF-IDF algorithm. Term Frequency-Inverse Document Frequency or TF-IDF is an algorithmic method that is useful for calculating the weight of each commonly used word. This method is also known to be efficient, easy and has accurate results. This method will calculate the Term Frequency (TF) and Inverse Document Frequency (IDF) values on each token (word) in each document in the corpus. In simple terms, the TF-IDF method is used to find out how often a word appears in a document.

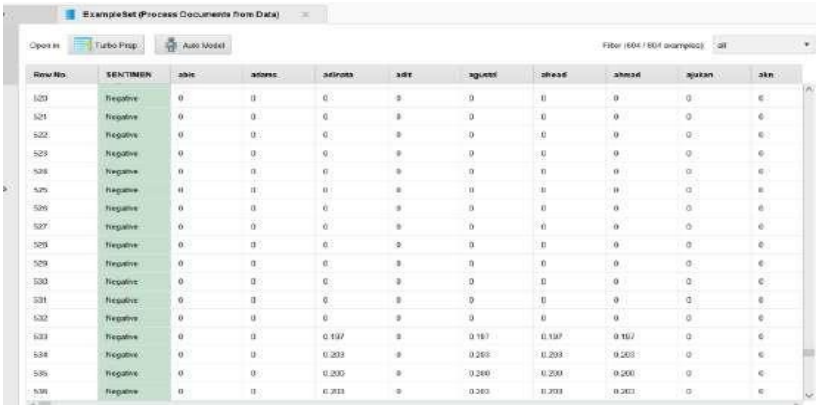

Figure 10. Word Weighting Results with TF-IDF.

Model Creation Stage

The output of this stage is a classification model with the Naive Bayes method and Support Vector Machine and Data Train which will later be used in the sentiment analysis process.

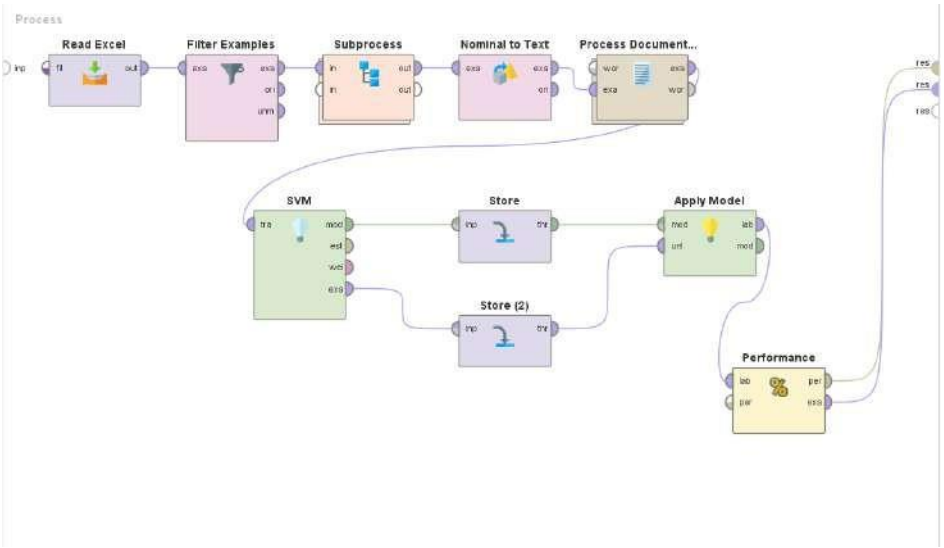

Figure 11. Model Creation of Support Vector Machine.

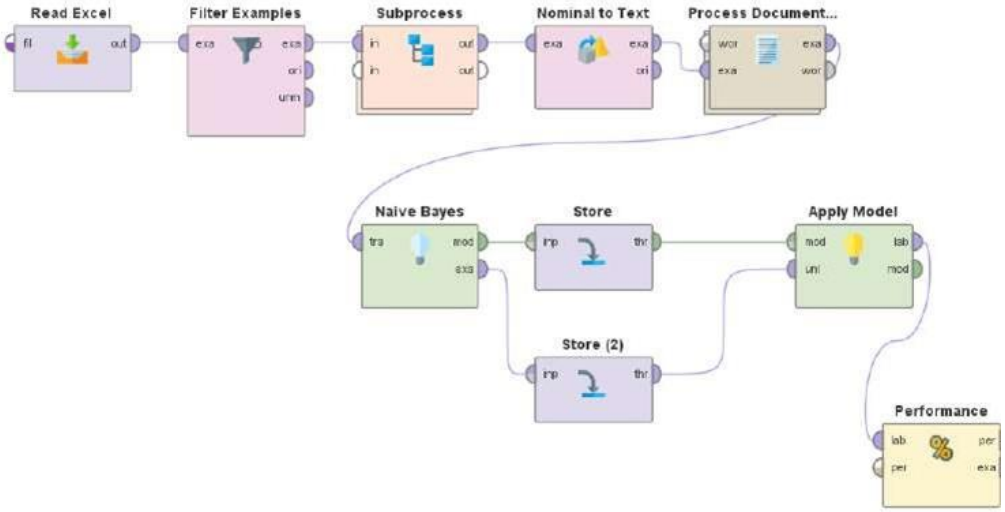

Figure 12. Creation of the Naive Bayes Model.

Test Data Preparation

The test data preparation stage is carried out to ensure that the data used in the test process meets the needs of the classification model. This process begins with filtering data that does not yet have sentiment labels so that only relevant data is used in testing. Next, a cleaning process is carried out to remove characters, symbols, and words that are not needed so that the quality of the data becomes better. After the cleaning process is complete, the test data is processed through a preprocessing stage that includes tokenizing, case folding, stemming, and stopwords removal. The final stage is word weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method to convert text data into numerical representations that can be processed by Naive Bayes algorithms and Support Vector Machine. The result of this stage is a test dataset that is ready to be used for the model classification and evaluation process.

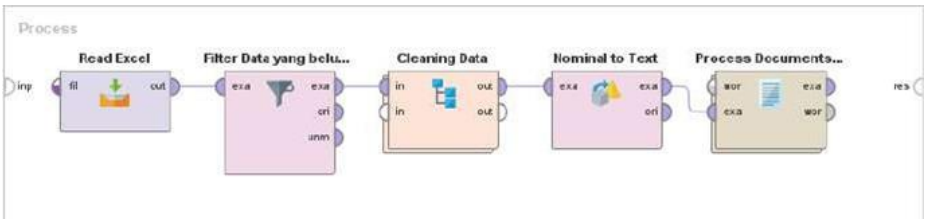

Figure 13. Preparation of test data.

Figure 13. shows the flow of test data preparation carried out in this study, starting from the process of data filtering, cleaning, preprocessing, to word weighting using the *TF-IDF* method.

Union Level/Data Merge

At this stage, the test data and the training data will be combined into one data.

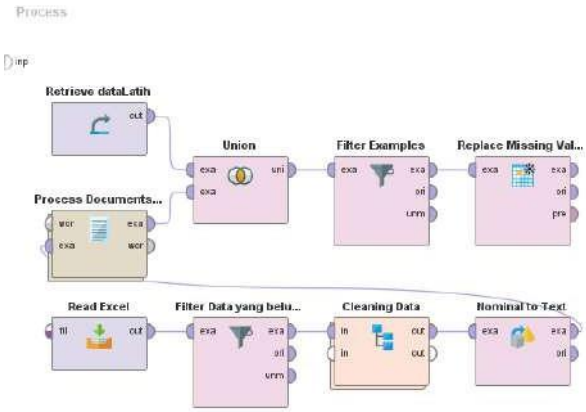

Figure 14. Data merge process.

Testing Stage

At this stage, a pre-built model will be applied to predict the sentiment on the test data.

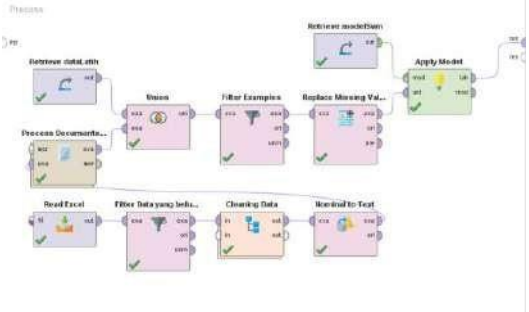

Figure 15. Testing Stage.

Testing

The Accuracy Value of Support Vector Machine and Naïve Bayees

The results of the calculation of the accuracy of the training data using the Support Vector Machine method, obtained an Accuracy value of 96.52%, a Positive Recall value of 82.14%, a Negative Recall value of 98.80%, a Positive Precision value of 98.92%, and a Negative Precision value of 96.09%. Of the 602 training data, the author labeled as 112 data as Positive Sentiment and 492 data as Negative Sentiment. For the predicted results of Negative Sentiment, there are 492 predicted Negative data and 1 predicted Positive data.

PerformanceVector

```
PerformanceVector:
accuracy: 96.52%
ConfusionMatrix:
True:   Positive        Negative
Positive:       92      1
Negative:       20      491
precision: 96.09% (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       92      1
Negative:       20      491
recall: 99.80% (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       92      1
Negative:       20      491
AUC (optimistic): 0.927 (positive class: Negative)
AUC: 0.923 (positive class: Negative)
AUC (pessimistic): 0.919 (positive class: Negative)
```

Figure 16. Testing Results of Support Vector Machine Model.

The results of the calculation of the accuracy of the training data using the Naive Bayes method, obtained an Accuracy value of 49.67%, a Positive Recall value of 100.00%, a Negative Recall value of 38.21%, a Positive Precision value of 26.92%, a Negative Precision value of 100.00%. Of the 602 training data, the author labeled as 112 data as Positive Sentiment and 492 data as Negative Sentiment. For the predicted results of Negative Sentiment, there are 492 predicted Negative data and 1 predicted Positive data.

PerformanceVector

```
PerformanceVector:
accuracy: 49.67%
ConfusionMatrix:
True:   Positive        Negative
Positive:       112     304
Negative:       0       188
precision: 100.00% (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       112     304
Negative:       0       188
recall: 38.21% (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       112     304
Negative:       0       188
AUC (optimistic): 0.995 (positive class: Negative)
AUC: 0.728 (positive class: Negative)
AUC (pessimistic): 0.461 (positive class: Negative)
```

Figure 17. Results of the Test of the Naive Bayes Model.

Comparison of Naive Bayes Algoritama and Support Vector Machine Accuracy Results

The result of the Implementation that has been carried out, the comparison of the accuracy level between the Naive Bayes method and the Support Vector Machine.

Table 4. Comparison of the Accuracy Results of NB and SVM Methods.

No	Metode	Nilai Akurasi
1	*Naive Bayes*	49.67%
2	*Support Vector Machine*	96.52%

Sentiment Analysis Results

From the sentiment analysis process with 482 test data, predictions of 335 Negative Sentiments and 147 Positive Sentiments were produced, and the author presented the data in a Pie Chart.

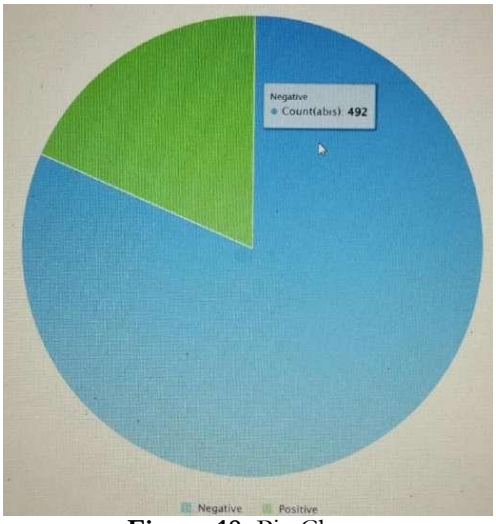

Figure 18. Pie Chart.

Figure 19. Wordcloud.

Final Test Results

The final result of the comparison with these two test methods, namely the prediction of Public Sentiment Towards the Issue of Determining the Job Creation Bill based on data obtained from Twitter and implemented by the SVM (Support Vector Machine) method, showed an accuracy value of 96.52%. Of the 605 test data, 492 data were predicted as Negative Sentiment and 112 data as Positive Sentiment and Naive Bayes Method showed an accuracy value of 49.67%. Of the 605 test data, 492 data were predicted as Negative Sentiment and 112 data as Positive Sentiment.

5. Conclusion

Based on the discussion, system implementation, and test results that have been carried out, it can be concluded that sentiment analysis can be done well using RapidMiner Studio software through the application of Support Vector Machine (SVM) and Naive Bayes algorithms. The analysis process is carried out through several stages, namely crawling data from Twitter social media, data labeling, data cleaning, preprocessing, feature extraction, and sentiment classification process. These stages allow the data obtained from social media to be processed into structured information that is ready for analysis.

The test results showed that there was a significant difference in performance between the two algorithms used. Based on the evaluation of 605 test data, the Support Vector Machine algorithm obtained an accuracy rate of 96.52%, while the Naive Bayes algorithm only obtained an accuracy rate of 49.67%. These results show that the Support Vector Machine method has a better ability to classify sentiment data compared to the Naive Bayes method in this study.

In addition, the results of sentiment classification showed that as many as 492 data were categorized as negative sentiment and 112 data were categorized as positive sentiment. The dominance of negative sentiment indicates that most Twitter users give a less supportive response to the enactment of the Job Creation Bill. Thus, it can be concluded that public opinion recorded through Twitter social media during the study period tended to show rejection of the policy.

Overall, this study proves that the Support Vector Machine algorithm is a more effective method of conducting sentiment analysis on social media data compared to Naive Bayes. The results of this study also show that sentiment analysis can be used as an approach to understand public perceptions and opinions on a public policy based on data obtained from social media.

References

[1] M. K. Sandryan, B. Rahayudi, and D. E. Ratnawati, "Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritma Backpropagation dan Term Frequency-Inverse Document Frequency," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5349–5355, 2021.

[2] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjeee.v3i2.10885.

[3] O. Fanny and H. Suroyo, "Analisis Sentimen Pengguna Media Sosial Terhadap Omnibus Law Berdasarkan Hashtag di Twitter," *Sist. J. Sist. Inf.*, vol. 11, no. 1, 2022, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[4] S. F. Pane, A. Owen, and C. Prianto, "Analisis Sentimen UU Omnibus Law pada Twitter Menggunakan Metode Support Vector Machine," *J. Telekomun. dan Komput.*, vol. 11, no. 2, p. 130, 2021, doi: 10.22441/incomtech.v11i2.10874.

[5] M. Faiq, A. Putro, and E. B. Setiawan, "Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media Sosial Twitter," *e-Proceeding Eng.*, vol. 9, no. 1, pp. 54–66, 2022.

[6] L. N. Pradany and C. Fatichah, "Analisa Sentimen Kebijakan Pemerintah Pada Konten Twitter Berbahasa Indonesia Menggunakan SVM dan K-Medoid Clustering," *SCAN - J. Teknol. Inf. dan Komun.*, vol. 11, no. 1, pp. 59–66, 2016.

[7] A. Ndruru, "Analisis Sentimen UU Cipta Kerja Melalui Omnibus Law Menggunakan Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM)," *Pelita Inform.*, vol. 10, pp. 85–90, 2022, [Online]. Available: https://www.ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/view/3768

[8] E. Indrayuni, "Komparasi Algoritma Naive Bayes dan Support Vector Machine untuk Analisa Sentimen Review Film," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 175, 2018, doi: 10.33480/pilar.v14i2.918.

[9] D. Rusdiaman and D. Rosiyadi, "Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naive Bayes Classifier dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 4, no. 2, pp. 230–235, 2019.

[10] A. Andreas and Y. D. Prabowo, "Analisis Sentimen Opini Publik dalam Bahasa Indonesia Terhadap UU Cipta Kerja Menggunakan Naive Bayes," *Kalbisiana J. Mhs. Inst. Teknol. dan Bisnis Kalbis*, vol. 8, no. 2, pp. 2146–2161, 2022.

[11] D. A. Wulandari, "Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Terhadap RUU Cipta Kerja Menggunakan Metode Klasifikasi Naive Bayes," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.

[12] T. S. Sabrila, Y. Azhar, and C. S. K. Aditya, "Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 10–19, 2022, doi: 10.14421/jiska.2022.7.1.10-19.

[13] S. Nurul, J. Fitriyyah, N. Safriadi, and E. Esyudha, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 279–285, 2019, doi: 10.26418/jp.v5i3.34368.

[14] M. D. D. Sreya and E. B. Setiawan, "Penggunaan Metode GloVe untuk Ekspansi Fitur pada Analisis Sentimen Twitter dengan Naive Bayes dan Support Vector Machine," *e-Proceeding Eng.*, vol. 9, no. 3, pp. 2008–2015, 2022.