*Journal of Applied Engineering and Technological Science*
Vol 4(1) 2022 : 223-234

# *OPTIMIZATION OF SENTIMENT ANALYSIS OF PROGRAM SEMBAKO(BPNT) BASED ON TWITTER*

**Mohamad Noor[1*], Windu Gata[2], Risnandar[3], Fakhrudin[4], Anisah Novitarani[5]**
Nusa Mandiri University[1 2]
Ministry of Social Affairs - Republic of Indonesia[3 4 5]
14002440@nusamandiri.ac.id, windu@nusamandiri.ac.id, risnandar@kemensos.go.id, fakhrudin@kemsos.go.id, anisah.novitarani@kemsos.go.id

Received : 17 September 2022, Revised: 03 October 2022, Accepted : 03 October 2022
**Corresponding Author*

***ABSTRACT***
*Food Assistance Program (Program Sembako) is a development of the Non-Cash Food Assistance (BPNT) program which has been implemented by the Ministry of Social Affairs since 2017, namely of food assistance in the form of non-cash from the government which is given to Beneficiary Families (KPM) every month through an electronic account mechanism that is used only to buy food in food traders/e-warong in collaboration with banks. Twitter social media has now become one of the places to disseminate information about the Program Sembako/BPNT. This case study uses text mining techniques with the support vector machine (SVM), Naïve Bayes (NB) and K-Nearest Neighbor (k-NN) methods which aims to classify public sentiment towards the Program Sembako/BPNT on Twitter. The dataset used is tweets in Indonesian with the keywords "BPNT" and "Kartu Sembako" with a total dataset of 1,094 tweets. Text mining, transformation, tokenize, stemming and classification, etc. A useful technique for constructing sentiment classification and analysis. RapidMiner and Gataframework are also used to help create sentiment analysis to measure classification values. The results obtained by optimization using Particle Swam Optimization (PSO) using the support vector machine (SVM) algorithm and the accuracy value obtained is 78.02%, with a precision value of 78.73%, a recall value of 82.16%, and an AUC of 0.848.*
***Keywords :*** *Sentiment Analysis, Program Sembako(BPNT), Twitter.*

## 1. Introduction

The Ministry of Social Affairs of the Republic of Indonesia as stated in Presidential Regulation Number 110/2021 article 4 has the task of carrying out government affairs in the social sector to assist the President in administering state government, and specifically carrying out the functions of formulating, determining, and implementing policies in the field of social rehabilitation, social security, social empowerment, and social protection; determination of criteria and data for the poor, vulnerable groups, and underprivileged people; implementation of social rehabilitation standards; as well as the implementation of technical guidance and supervision of the implementation of the affairs of the Ministry of Social Affairs in the regions.

Program Sembako is a development of the Non-Cash Food Assistance (BPNT) program which is the result of the transformation of the Rice Welfare Program (Rastra) and has been implemented by the Government since 2017. The addition of the Basic Food Program index to IDR 200,000 per month provides an opportunity for Beneficiary Families (KPM) to spend aid funds to purchase types of food commodities with more complete nutritional content at the electronic Warung Gotong Royong (e-Warong). The Basic Food Program is a complementary program that supports the implementation of the Stunting Reduction Acceleration Program as an effort by the Government to reduce stunting rates in Indonesia(Kamia, 2022; Safitri, & Rodiyah, 2022).

Program Sembako is one of the efforts of the Ministry of Social Affairs in supporting the implementation of the National Non-Cash Movement in Indonesia so that it can raise awareness while increasing the use of non-cash among the public, businesspeople and government institutions. Like the BPNT Program, the Sembako Program is also distributed to KPM using the banking system through the Prosperous Family Card (KKS), which can then be used to purchase foodstuffs containing carbohydrates, protein, vitamins and minerals so that KPM can obtain more complete and balanced nutrition. KPM This basic food program is registered in the Social Welfare Integrated Data which is managed by the Social Welfare Data and Information Center-Ministry

of Social Affairs. KPMs are not allowed to have the status of Civil Servants, members of the Indonesian National Armed Forces, or members of the Indonesian National Police.

Sentiment analysis is the process of extracting, processing and understanding data in the form of unstructured text automatically in order to retrieve sentiment information contained in an opinion. Sentiment analysis can be applied to opinion in all fields such as economics, politics, social and law. This social media Twitter opens a window for researchers to study emotions, moods, and public opinion through sentiment analysis(Birjali et al., 2021; Yadav, A., & Vishwakarma).

Text Mining is the discovery of knowledge in the database in textual form (knowledge discovery in textual database or abbreviated as KDT), it can also be called extracting or searching data in the form of text, including interest in newly created knowledge, defined as part of the process. extracting or searching for previously unknown text data, so that it can be understood, has potential and practical patterns or knowledge from massive and unstructured text data collections or corpus(Hertina et al., 2021). Every opinion written through twitter can have a positive value or a negative value. So, it is necessary to carry out a sentiment analysis process to classify an opinion written(Zimbra et al., 2018; Marcec & Likic, 2022).

The purpose of this study was to obtain a classification of positive and negative public responses from posting tweets about the Program Sembako/BPNT in order to provide information to leaders about sentiment analysis of the Ministry of Social's Program Sembako/BPNT sentiment on Twitter. This study proposes several methods to extract tweets and perform text classification of twitter data with the keywords "BPNT" and "Kartu Sembako" which have been obtained with a total of 1,094 tweets with the aim of knowing the characteristics of the content of the tweets and knowing the level of classification accuracy obtained. After examining the data in question, then conducting a text mining process to handle these problems using the Support Vector Machine (SVM)(Zhou et al., 2022; Sabanci et al., 2022; Zhou et al., 2022), Naïve Bayes (NB)(Mansour et al., 2022; Vu, 2022; Ahmad et al., 2022) and K-Nearest Neighbor (k-NN)(Dann et al., 2022; Cubillos et al., 2022; Gallego et al., 2022) algorithms using the Particle Swarm Optimization (PSO) technique to optimize accuracy(Shami et al., 2022; Fan et al., 2022).

## 2. Research Methods

The research methodology used in this experimental research using the Cross-Industry Standard Process for Data Mining method consists of six stages, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

In addition to analyzing sentiment and knowing its accuracy, there are several stages to get the best results. The steps taken are data collection and labeling, pre-processing, sentiment analysis and evaluation to obtain accuracy, precision, and recalls.

The entire sentiment analysis methodology that has been carried out can be seen in the following:

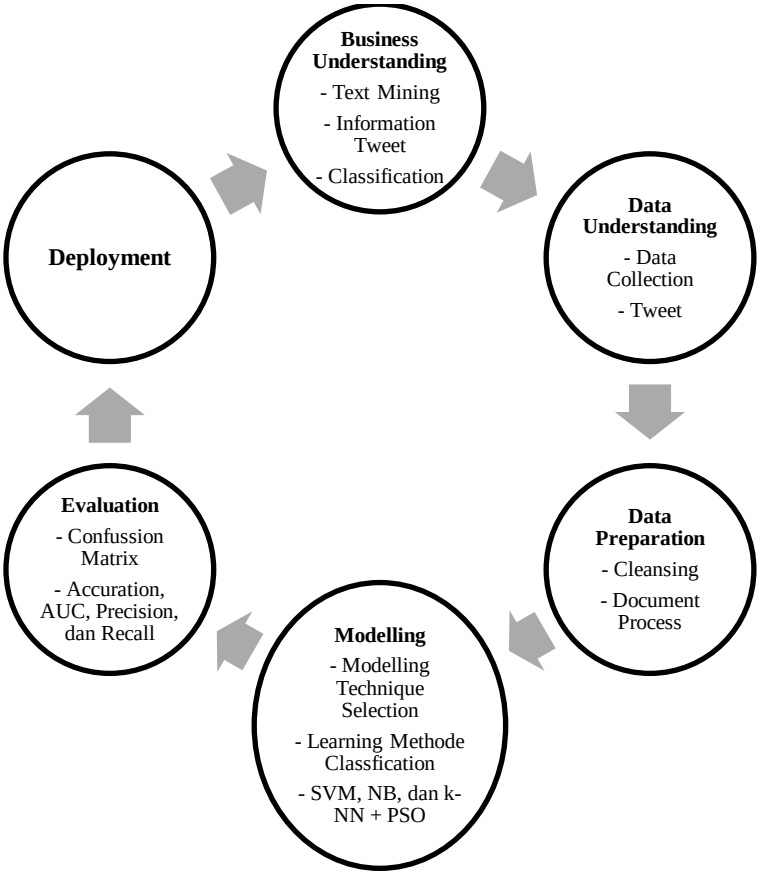


Fig 1. Research Process Model of CRIPS-DM

A. Business Understanding

Determination of project objectives and requirements in detail within the scope of the business or research unit comprehensively. In addition, it also translates goals and limitations into formulas for data mining problems and prepares initial strategies to achieve goals. This stage can be analogous to how we make an airplane without making a model of the airplane. The data is taken from twitter in the form of an open-source dataset and data collection using the text mining method and tweet classification model that will be used by the Support Vector Machine, Naïve Bayes and K-Nearest Neighbor algorithms and adding the Particle Swarm Optimization feature to improve the accuracy of the classification method used. in datasets.

B. Data Understanding

The data understanding stage requires an understanding of the data that will be used in the research process. The data used for this research experiment was collected through tweet data originating from twitter about the Program Sembako/BPNT with recent and popular options until June 17, 2022, using the Rapid Miner application called local data, then local data was labeled as the result of selecting the dataset. created in Microsoft Excel to be processed in data testing.

C. Data Preparation

Data Preparation uses tweet data from Twitter with the hashtags #bpnt and #kartusembako, which are taken with recent and popular options until June 17, 2022. The field used in this problem is only Text and a new field is added, namely status to be used as a class. Text mining is an important prologue to data mining which is proposed to analyze text and content using word relationship extraction.

D. Modelling

The method that will be proposed is the use of a classification method, namely Support Vector Machine, Naïve Bayes and K-Nearest Neighbor as a comparison. The three selection methods are compared to determine the best selection method to apply using the Support Vector Machine, Naïve Bayes, and K-classifiers. Nearest Neighbor because it is a popular machine learning technique for text classification and has good performance in many domains.

- Support Vector Machine (SVM)
  Classification using the Support Vector Machine is a machine learning technique that is quite popular for text classification and has good performance in many domains and can identify hyperplanes separately between two different classes so that the results are maximized and can also maximize the distance between the data closest to the hyperplane. Classification is done by looking for a hyperplane or decision boundary that separates one class from another. Support Vector Machine searches for hyperplane values using support vectors and margin values.
- Naive Bayes (NB)
- Naive Bayes is an algorithm used to find the highest probability value to classify test data in the most appropriate category (Nugroho, Agung, Dzulatkha 2020). The use of the Bayes theorem in the Naive Bayes algorithm is to combine the previous probabilities and conditional probabilities in a formula that can be used to calculate the probability of each possible classification.
- K-Nearest Neighbor (k-NN)
- KNN algorithm is also a method for classifying data based on prepared sample/training data. The purpose of this method is to classify a data based on the attribute match value on the training data (sample). Classification of a data is done by drawing as many as k neighbors on the training data, then voting is done to classify the data based on most values in the k training data.

The data must go through the preprocessing stage first in order to get the relevant words to be classified. The preprocessing result is Bags of Word (BoW) containing all the terms contained in the collection of documents resulting from preprocessing.


Fig 2. Bags of Word

After processing the initial data, the next step is to enter the data into the proposed method

- Particle Swarm Optimization (PSO)
  Aims for optimization by continuously calculating potential solutions using a quality reference. This algorithm optimizes the problem by moving the particles/prospective solutions in the problem space using certain functions for the position and velocity of the particles. The motion of the particle is affected by the best solution for that particle, and the best solution in general obtained from other particles.

The classification results are then validated for reliability. To carry out this validation, 10-fold cross validation is used. There are three measures used to assess the comparison of the three algorithms, namely Recall, Precision, Accuracy and AUC.

E. Evaluation

This study will use the X-Cross Validation technique using Rapidminer software to test the Support Vector Machine, Naïve Bayes and K-Nearest Neighbor algorithm approaches by adding the Particle Swarm Optimization feature to classify tweets from the Program Sembako/BPNT. A more detailed explanation for the testing of the algorithmic approach model will be explained in the next section.

F. Deployment

This stage is making conclusions from the research after getting the results of the accuracy of tweet classification from each of the Support Vector Machine (SVM), Naïve Bayes (NB) and K-Nearest Neighbor (k-NN) algorithm approaches by adding the Particle Swarm Optimization feature. (PSO) is used in local datasets and implemented in an application.

## 3. Results and Discussions

Based on the methodology described in the previous section, this section will explain the stages of implementing the research methodology.

A. Data Collection

The first stage in this process is data collection. The dataset used is a local dataset through tweet data originating from twitter about the Program Sembako/BPNT. Based on this, the tweet classification model approach will use the SVM, NB and k-NN algorithms and add the PSO feature to increase the accuracy of the classification method used in the local dataset.

B. Labelling Data

This stage is the process of understanding the data that will be used as material to be studied so that it can be carried out to the next stage, namely Preprocessing. Prepare tweet data taken from twitter. The data taken were 1,094 tweets, the process carried out was cleaning data from the many attributes by selecting the important attribute, namely the tweet text attribute plus one status attribute to create a label from the selected attribute.

After having two attributes that will be used for modeling then cleaning the data because in 1,094 tweet data there is retweet data which is the same tweet shared repeatedly after cleaning, 382 data are obtained which are then labeled Positive or Negative status, data labeling method with details 174 negative tweets and 208 positive tweets which also involved the participation of the Ministry of Social Affairs.

Table 1- Labelling Datasets

| *Text* | *Status* |
|---|---|
| @Pardimin9 @idextratime Dana BPNT juga gabakal cair bang | Negative |
| Nama Penerima BPNT Juni 2022 Ada di Sini, Login Segera ke https://t.co/wAG9qW18pG dan Dapatkan Rp2,4 Juta https://t.co/eR9OKQds5O | Positive |
| Ambil KTP! Cek Nama Daftar Penerima Bansos BPNT Rp200 Ribu di https://t.co/czw8Y0GN13 https://t.co/pc2lkspPdH | Positive |
| Bansos BPNT dan PKH Rp4, 4 Juta Tidak Cair? Simak Info Masalah Penerima https://t.co/WgNdBhqdf8 | Negative |
| Login di Link Ini untuk Daftar DTKS Online dan Cek Penerima BPNT Kartu Sembako Juni 2022, Dapatkan Bansos ini! https://t.co/NjD9FYkFiL | Positive |

Table 2 - Sentiment Polarity Tweets

| Tweets | Total | % |
|---|---|---|
| Negative | 174 | 45.55% |
| Positive | 208 | 54.45% |

Table 2 is a table of the results of the polarity percentage of the tweets obtained. Sentiment analysis on the polarity of tweets obtained is the number of positive tweets 208 and the number of negative tweets 174. The percentage of positive opinions is 54.45% and negative is 45.55%. From these results, it can be said that public opinion (netizens) on the Program Sembako/BPNT is dominated by positive opinions. Although some netizens do not agree with the program.

C. Text Preprocessing

At this stage, namely preparing data to perform steps called text preprocessing, using two applications for preprocessing, first using the Gata Framework which is accessed via the link http://gataframework.com/textmining for the cleansing process and using RapidMiner for the Document Process. Preprocessing is needed before the classification process so that the dimensions of the vector space model become smaller by reducing the dimensions of the vector space model, the classification process will be faster. The purpose of this pre-processing is to homogenize and reduce the volume of words.

1) Cleansing Data

The following is a cleansing step using the Gata framework website to create a local data preprocessing design model. @Anotation removal, transform URL, regexp, transformation not, Indonesian stemming, and Indonesian stop word removal

a) @anotation removal

In this process, all annotations contained in the tweet will be removed and all capital letters will be converted to lowercase.

b) Transformation Remove URL

The results of @annotation removal will be continued to the Transformation Remove URL process, in this process the link or URL contained in the tweet will be removed.

c) Regxp

The results of the Transformation Remove URL are continued by the Tokenization (Regexp) process, where all words in each document are collected and punctuation removed, and if there are symbols, special characters or anything that is not a letter.

d) Transformation is not negative

From the results of Tokenization (Regexp) then the transformation process is not negative. In this process, words that have negative values will be converted into non-negative sentences.

e) Indonesian Stemming

The result of the not negative transformation will be a stemming process, namely changing the affixed words into basic words using Indonesian stemming for language tweets.

f) Indonesian Stopword Removal

The results of Indonesian stemming will be continued to the Indonesian stop word removal process, in this process irrelevant words will be removed, such as the words “tetapi”, “untuk”, “dengan”, which have no separate meaning if separated from other words and are not related to adjectives. related to sentiment.

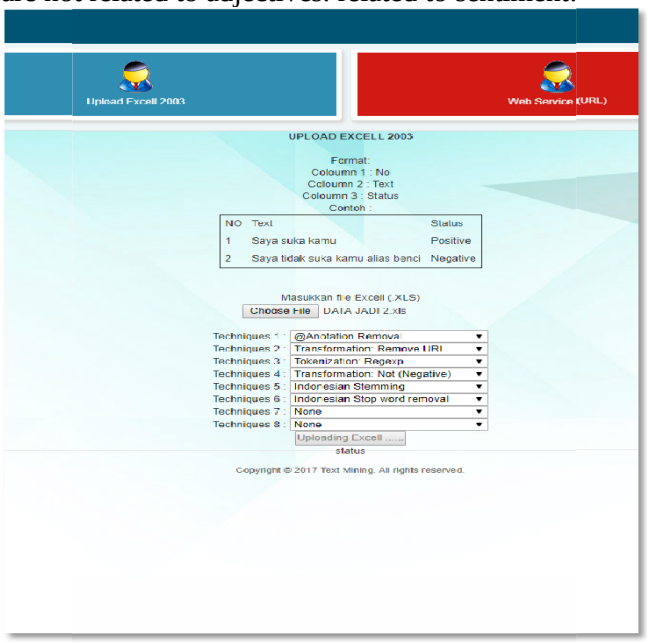

Fig 3. Preprocessing Data using Gataframework

2) Document Process
   a) Tokenization
      The results of Indonesian stemming are continued by the Tokenization process from RapidMiner, where all words in each document are collected and punctuation removed, and removed if there are symbols, special characters or anything that is not letters and breaks sentences into words.
   b) Transform Cases
      The next stage is transformed cases, which aims to homogenize the shape of the letters into lower cases because Twitter comments themselves have various letter forms.
   c) Filter Stopwords (Dictionary)
      For the next process is the stop word filter stage (dictionary), in this process irrelevant words will be removed, such as words but, for, with, which have no separate meaning if separated from other words and are not related to adjectives related to sentiment.
   d) Filter Tokens (by Length)
      The results of the stop word filter process (dictionary) are followed by the Filter Tokens (by Length) stage, in this process words that have a character length of less than 4 and more than 25 will be deleted, such as the words “di”, “ada”, “oleh”, which are words that does not have its own meaning if it is separated from other words and is not related to adjectives related to sentiment.

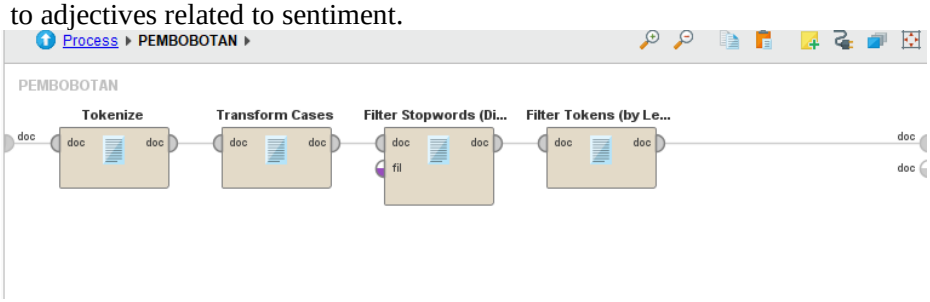


Fig 4. Preprocessing Data using Rapidminer

D. Modelling
   In the next stage, the dataset that has been created in the previous stage is used as input for the classification algorithm, which is used as a training and testing dataset. In this study, three types of algorithms will be used for comparison, namely Support Vector Machine (SVM), Nave Bayes (NB) and K-Nearest Neighbor (k-NN)
   1) Process Modelling
      The sequence of each modeling process for each algorithm is read excel, nominal to text, process document, multiply, cross validation which includes the algorithm used, apply model and performance. Furthermore, to add the PSO feature in which there is cross validation, apply models and the performance of each algorithm used so that the results are more optimal.

   Figure 6 is the process modeling of the Rapidminner application using the SVM method with the order of reading excel, nominal to text, process document, multiply, cross validation which includes the SVM algorithm, apply model and performance, while in Figure 7 is the process modeling of the RapidMiner application using the method SVM is added with PSO which includes cross validation of the SVM algorithm, apply model and performance.

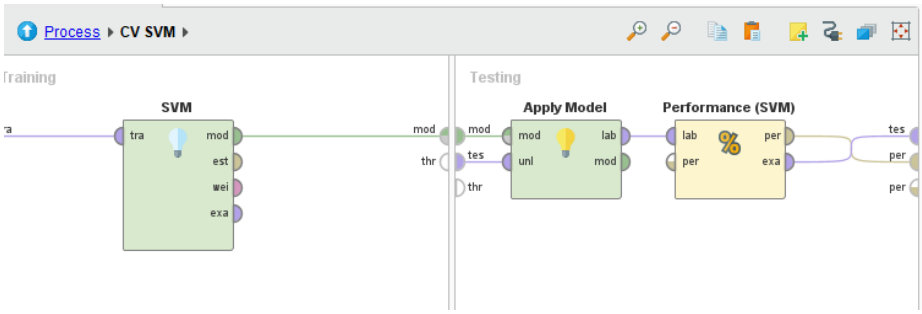


Fig 5. Preprocessing Modelling SVM Method

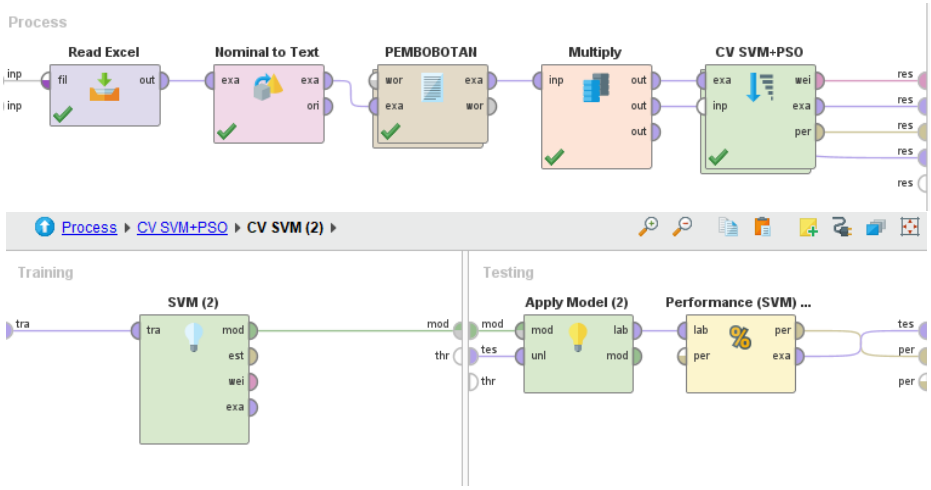


Fig 6. Preprocessing Modelling SVM+PSO

E. Evaluation of Test Results and Model Validation

The results of the model testing carried out are to classify the tweets of the Program Sembako/BPNT in Indonesia into Positive and Negative as local data using the SVM and SVM+PSO algorithms. NB and NB+PSO, k-NN and k-NN+PSO. At this evaluation stage, the performance of the classification process that has been carried out will be tested with the parameter's accuracy, precision, and recall.

Evaluation of this system analysis is done by calculating the level of accuracy of a method in analyzing opinions, k-fold cross validation is chosen to test the accuracy of the Naive Bayes method and Support Vector Machine. Dividing data into two parts, namely training data and testing data is the principle of cross-validation. In this study using 10-fold cross validation. The data will be divided into 10 parts, part 1 will be the training data and part 2-10 will be the testing data. Furthermore, a cross process is carried out, where the training data is used as testing data and vice versa, this process is repeated 10 times.

In the process of 10-fold cross validation, a confusion matrix is created which consists of True Positives (TP) which is the number of positive class data classified as positive class and True Negatives (TN) which is the number of negative class data classified as negative class. While False Positives (FP) is the number of negative class data classified as positive class, and False Negatives (FN) is the number of positive class data classified as negative class. The average value of accuracy, precision, recall can be obtained from the table(Widowati dan Sadikin 2021).

| Row No. | id | STATUS | prediction($... | confidence(... | confidence(... | aapa | adil | afkir | ajar | ajuk | akses |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 1534992970... | POSITIF | NEGATIF | 0.359 | 0.641 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 1534638084... | NEGATIF | NEGATIF | 0.497 | 0.503 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 1534524597... | NEGATIF | NEGATIF | 0.332 | 0.668 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 1534484461... | NEGATIF | POSITIF | 0.501 | 0.499 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 1534463648... | NEGATIF | NEGATIF | 0.492 | 0.508 | 0 | 0 | 0 | 0 | 0 | 0 |

Fig 7. Prediction Results from RapidMiner

And to find out the best model for each model, here are the results of testing the model on RapidMiner :

Table 3 - Accuration

| Model | SVM | NB | k-NN | SVM + PSO | NB + PSO | k-NN + PSO |
|---|---|---|---|---|---|---|
| Accuracy | 70.69% | 72.26% | 76.69% | 78.02% | 76.17% | 77.75% |
| AUC | 0.834 | 0.505 | 0.820 | 0.848 | 0.552 | 0.791 |

Table 4 - Precision and Recall

| Algorithm | Precision | Recall | Prediction |
|---|---|---|---|
| SVM | 71.53% | 71.53% | Negative |
| SVM | 70.17% | 70.17% | Positive |
| SVM+PSO | 77.44% | 77.44% | Negative |
| SVM+PSO | 78.44% | 78.44% | Positive |
| NB | 71.53% | 71.53% | Negative |
| NB | 70.17% | 70.17% | Positive |
| NB+PSO | 77.44% | 77.44% | Negative |
| NB+PSO | 78.44% | 78.44% | Positive |
| k-NN | 78.91% | 78.91% | Negative |
| k-NN | 75.32% | 75.32% | Positive |
| k-NN+PSO | 72.36% | 72.36% | Negative |
| k-NN+PSO | 83.61% | 83.61% | Positive |

Table 5 - Confusion Matrix

| Algorithm | Prediction | True Negative | True Positive |
|---|---|---|---|
| SVM | Negative | 103 | 41 |
| SVM | Positive | 71 | 167 |
| SVM+PSO | Negative | 127 | 37 |
| SVM+PSO | Positive | 47 | 171 |
| NB | Negative | 104 | 36 |
| NB | Positive | 70 | 172 |
| NB+PSO | Negative | 108 | 25 |
| NB+PSO | Positive | 66 | 183 |
| k-NN | Negative | 116 | 31 |
| k-NN | Positive | 58 | 177 |
| k-NN+PSO | Negative | 144 | 55 |
| k-NN+PSO | Positive | 30 | 153 |

Figure 7 shows the results of sentiment prediction in the RapidMiner application. Compares the predefined sentiment label with the predicted result label from the process. Table 3 is the result of the accuracy of each algorithm, table 4 is the result of precision and recall, table 5 is the result of the confusion matrix.

Based on these results, testing with the SVM algorithm model without PSO features obtained an accuracy of 70.69% and the AUC result is 0.834. Then the results of the SVM added with the PSO feature, the accuracy is 78.02% and the AUC result is 0.848. It can be concluded that the PSO feature is very influential in increasing the accuracy, but the use of the PSO feature is not significant in increasing the AUC of each algorithm model.

F. Deployment

Based on the evaluation results of the SVM, NB, and k-NN algorithm model testing process without features as well as SVM, NB, and k-NN using PSO features, the highest model testing results from all algorithm testing results are PSO feature-based SVM. Therefore, the weights that will be used in modeling the application are based on the test results of the PSO feature based SVM algorithm

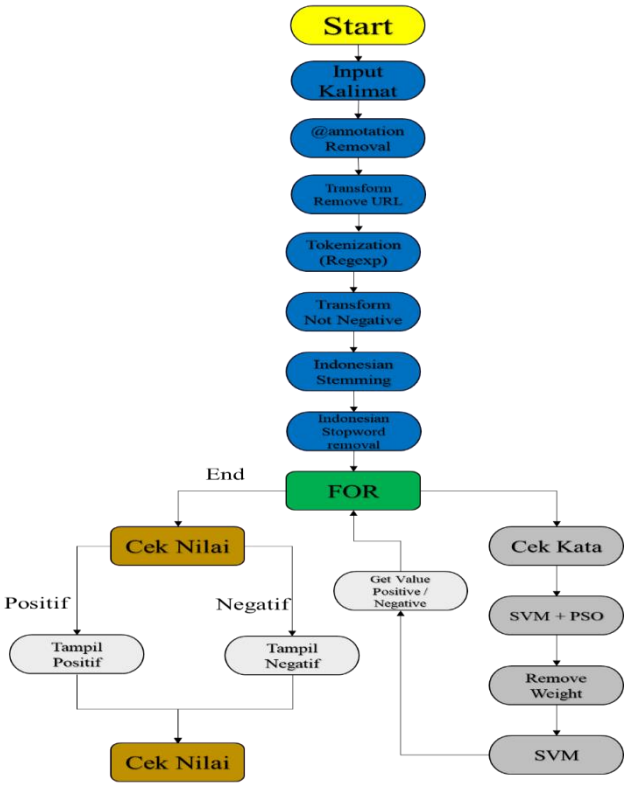


Fig 8. Deployment Flowchart

The way to deploy can't directly use the dataset from PSO, because the data from PSO only has a character weighting value of 0 to 1. After getting the dataset weight from PSO, the character value 0 will be entered into the stopword dictionary. Attributes with weight 0 will be added to the process document and entered in the stopword dictionary.

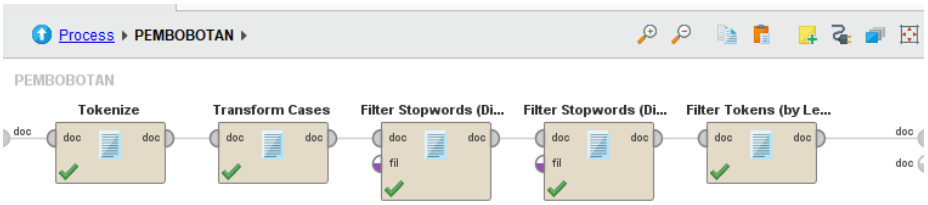


Fig 9. Document Process added with stopword dictionary filter from PSO weight results

After being entered into the filter stopword dictionary, the dataset is reprocessed using the SVM algorithm to obtain the weighted results. Next, enter the weight results into the database to create a deployed website. Then test the deployment results by entering text in the web application to determine whether the sentence is positive or negative.

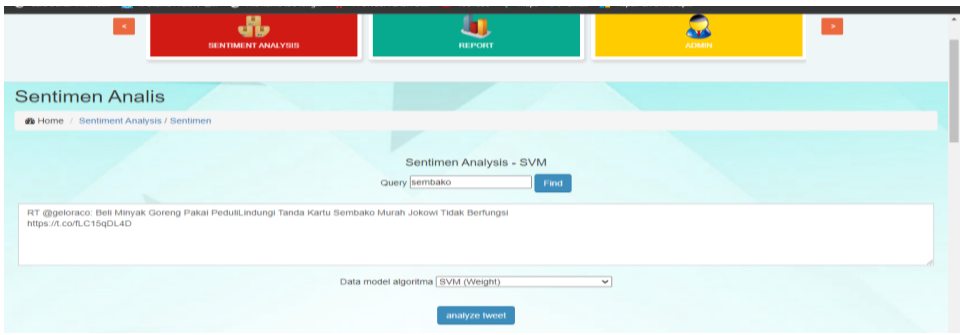


Fig 10. Input text into the deployed application

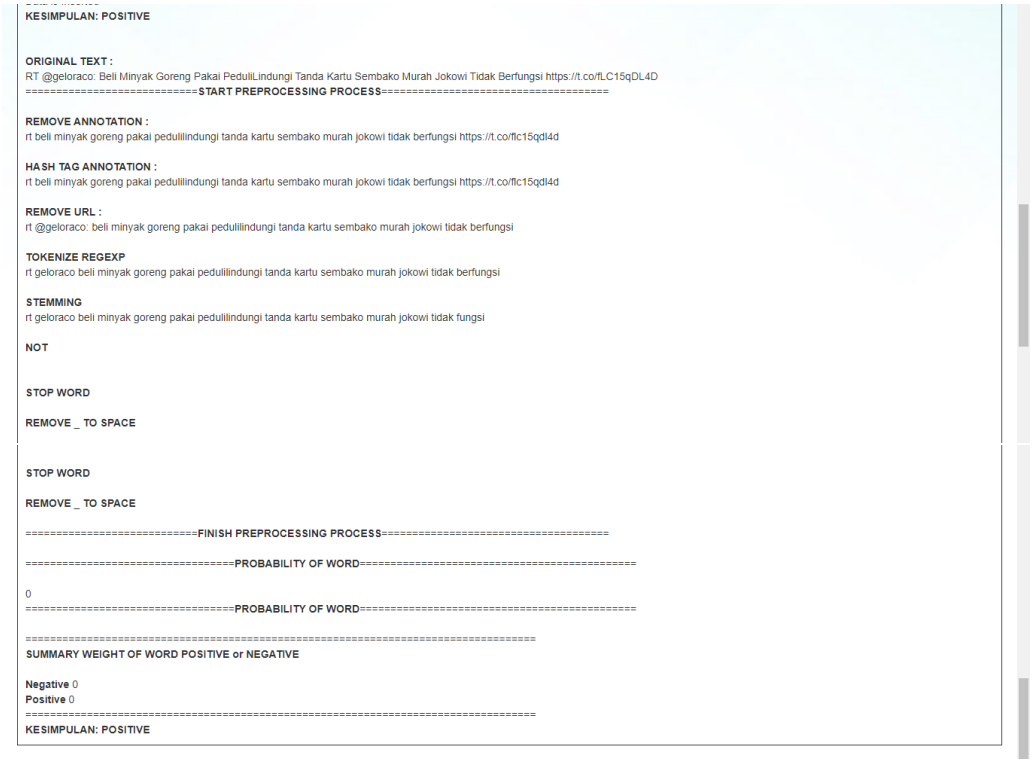


Fig 11. The Process Of Cleaning Sentences And Weighting Words In The Application

Figure 11 is the cleansing process into standard words, then weighted with the SVM algorithm plus the PSO feature in order to determine whether the result of the sentence is positive or negative which is carried out through the gataframework website.

## 5. Conclusion

The proposed use of the PSO feature in the classification algorithm has proven to be very influential in increasing the accuracy of the SVM algorithm to 78.02% of the tweet data processing as much as 382 data. The proposed use of the PSO feature in the classification algorithm is also proven to be very influential in increasing the accuracy of the NB algorithm to 76.17% of the tweet data processing as much as 382 data. While the use of the PSO feature on the k-NN algorithm is 77.75% of the tweet data processing as much as 382 data.

It is proven that the use of appropriate features in the SVM algorithm is superior to the NB and k-NN algorithms. However, the PSO feature added to the NB and k-NN algorithms does not significantly increase the accuracy of the algorithm.

## References

Ahmad, F., Tang, X. W., Qiu, J. N., Wróblewski, P., Ahmad, M., & Jamil, I. (2022). Prediction of slope stability using Tree Augmented Naive-Bayes classifier: Modeling and performance evaluation. *Math. Biosci. Eng*, *19*, 4526-4546.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Cubillos, M., Wøhlk, S., & Wulff, J. N. (2022). A bi-objective k-nearest-neighbors-based imputation method for multilevel data. *Expert Systems with Applications*, 117298.

Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D., & Marioni, J. C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nature Biotechnology*, *40*(2), 245-253.

Fan, Y., Wang, P., Heidari, A. A., Chen, H., & Mafarja, M. (2022). Random reselection particle swarm optimization for optimal design of solar photovoltaic modules. *Energy*, *239*, 121865.

Gallego, A. J., Rico-Juan, J. R., & Valero-Mas, J. J. (2022). Efficient k-nearest neighbor search based on clustering and adaptive k values. *Pattern Recognition, 122*, 108356.

Hertina, H., Nurwahid, M., Haswir, H., Sayuti, H., Darwis, A., Rahman, M., ... & Hamzah, M. L. (2021). Data mining applied about polygamy using sentiment analysis on Twitters in Indonesian perception. *Bulletin of Electrical Engineering and Informatics, 10*(4), 2231-2236.

Kania, I. (2022). Evaluation of the Non-Cash Food Assistance Program in Sadang Village, Sucinaraja District, Garut Regency. *ijd-demos, 4*(2).

Mansour, N. A., Saleh, A. I., Badawy, M., & Ali, H. A. (2022). Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. *Journal of ambient intelligence and humanized computing, 13*(1), 41-73.

Marcec, R., & Likic, R. (2022). Using twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal, 98*(1161), 544-550.

Sabanci, K., Aslan, M. F., Ropelewska, E., & Unlersen, M. F. (2022). A convolutional neural network-based comparative study for pepper seed classification: Analysis of selected deep features with support vector machine. *Journal of Food Process Engineering, 45*(6), e13955.

Safitri, N. R. Y., & Rodiyah, I. (2022). Implementation of the Non-Cash Food Assistance Program in Sidoarjo Regency. *Indonesian Journal of Public Policy Review, 20*, 10-21070.

Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle swarm optimization: A comprehensive survey. *IEEE Access*.

Vu, D. H. (2022). Privacy-preserving Naive Bayes classification in semi-fully distributed data model. *Computers & Security, 115*, 102630.

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review, 53*(6), 4335-4385.

Zhou, J., Zhu, S., Qiu, Y., Armaghani, D. J., Zhou, A., & Yong, W. (2022). Predicting tunnel squeezing using support vector machine optimized by whale optimization algorithm. *Acta Geotechnica, 17*(4), 1343-1366.

Zhou, W., Jiang, H., Cheng, Y., Pei, L., & Ding, S. (2022). Predicting seasonal patterns of energy production: a grey seasonal trend least squares support vector machine. *Expert Systems with Applications*, 118874.

Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems (TMIS), 9*(2), 1-29.