Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol 6, No.1, March 2023, pp. 8 – 17
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Comparison of Naïve Bayes, Support Vector Machine, and Decision Tree Algorithms in Analyzing Public Opinion Sentiments on COVID-19 Vaccination in Indonesia

[1]Rahmaddeni, [2]Firman Akbar
[1,2]Departement of Informatics Engineering, STMIK Amik Riau, Jalan Purwodadi, Pekanbaru, Riau
Email: [1]rahmaddeni@sar.ac.id, [2]2110031802067@sar.ac.id

Article Info
Article history:
Received Nov 23th, 2022
Revised Feb 14th, 2023
Accepted Mar 24th, 2023

Keyword:
Decision Tree
Naïve Bayes Classifier
Sentiment Analysis
Support Vector Machine
Twitter

ABSTRACT
The spread of COVID-19 in Indonesia has caused many negative impacts in various sectors. In order to control the spread of COVID-19, the government has taken steps to conduct vaccinations. This government action has generated public reactions expressed on Twitter social media, some in support and some against the vaccination. In this study, the responses expressed on Twitter were used as public data taken from the Drone Emprit Academic (DEA) portal with a total of 700 data. The data obtained was classified using Naïve Bayes, Support Vector Machine (SVM), and Decision Tree algorithms. The aim of this research is to provide an understanding to the public whether the COVID-19 vaccination tends towards positive, neutral or negative opinions by comparing the best accuracy levels produced by the three algorithms used, namely Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree. Validation testing was performed using the K-Fold Cross Validation method, AdaBoost feature selection, and TF-IDF Transformer feature extraction. The results of the methods used in this study using 50:50, 70:30, 80:20 and 90:10 data splitting showed an increase in accuracy in the 90:10 data splitting, with 82.86% accuracy for the SVM algorithm, 81.43% for Naïve Bayes and 78.57% for Decision Tree, and the processed data generated knowledge or information that public sentiment polarity tends towards the positive direction.

Corresponding Author:
Rahmaddeni,
Departement of Informatics Engineering,
STMIK Amik Riau,
Jalan Purwodadi, Pekanbaru, Riau.
Email: rahmaddeni@sar.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v6i1.19966

1. INTRODUCTION

The Corona Virus Disease 2019 (COVID-19) is a new disease reported in Wuhan, China, starting in December 2019 [1]. This virus is highly dangerous because it spreads rapidly worldwide due to its high transmission rate. According to sources from www.covid19.go.id as of August 6th, 2021, Indonesia alone has 3.6 million positive cases and more than 104 thousand deaths.

The Indonesian government has made efforts to suppress the spread of Corona Virus Disease 2019 (COVID-19) so that the negative impacts can be controlled, including by carrying out vaccination actions. Vaccines not only protect those who have been vaccinated, but also the wider community by reducing the spread of the disease within the population [2].

Information about vaccination and virus prevention methods has been posted on various social media [3]. Social media is one of the most common sources for communication, sharing documents, and data of large communities [4]. One social media platform frequently used by Indonesian citizens is Twitter, with 10,645,000 users in Indonesia at present [5]. The use of vaccines has sparked various reactions and opinions from different groups, ranging from constructive to contradictive and dismissive.

Based on the information available on Twitter social media, various analytical methods can be used to analyze public opinions. One of them is sentiment analysis classification. Sentiment analysis is a type of text mining that analyzes and classifies data obtained from the internet to determine its polarity [6]. Sentiment analysis is part of the supervised learning group classification algorithm. The grouping is done to determine whether the polarity of reviews is positive, neutral, or negative [7].

Through sentiment analysis, existing opinion polarities can be collected and used to predict the public mood or emotional image of netizens as negative, neutral, or positive. Previous studies have used several methods in sentiment analysis, including Naïve Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), and K-Nearest Neighbor (K-NN). This is based on three studies of classification methods used in sentiment analysis on social media such as studies [8–10].

The Naïve Bayes method is quite popular because of its simple, fast, and accurate model structure, making it widely used in big data analysis and other fields. Study [11] conducted a sentiment analysis of product reviews using the Naïve Bayes method and obtained an accuracy rate of 77.78%. In addition, study [12] achieved a high accuracy rate of 98%. Study [13] also conducted a sentiment analysis of restaurant reviews in Singapore and found an accuracy rate of 70%.

Decision Tree is another commonly used method with high accuracy rates, in addition to the Naïve Bayes method. After comparing several methods, Decision Tree was able to produce a high accuracy rate of 83.3% [14]. In another study, as mentioned earlier, Decision Tree was found to have an accuracy rate of 96.83%, indicating that Decision Tree is good and accurate [15], other studies have found that the Decision Tree method has a perfect accuracy rate of 100% [16].

Study [17] also achieved higher accuracy rates for the Decision Tree and Naïve Bayes methods based on the dataset used. The Decision Tree and Naïve Bayes methods are recommended in study [18] as methods that can provide more accurate and effective predictions of early disease detection. Similarly, study [19] discusses public opinions on the spread of COVID-19 in commuter train passengers by comparing the Decision Tree method with the Naïve Bayes method. In addition to Naïve Bayes and Decision Tree methods, Support Vector Machine is also often used in data analysis because it is a new type of method based on statistical learning theory and has high accuracy.

Study [20] conducted sentiment analysis using the Support Vector Machine method and obtained an accuracy rate of 93.65%. Study [21] also conducted sentiment analysis using the Support Vector Machine method and obtained an accuracy rate of 83%. Furthermore, study [22] conducted sentiment analysis using the Support Vector Machine method and obtained an accuracy rate of 96.26%.

Study [23] combined Support Vector Machine with Decision Tree and proved that the approach provided better classification results in terms of f-measure and accuracy compared to without the combination. Study [24] compared the evaluation of Decision Tree, K-NN, Naïve Bayes, and Support Vector Machine algorithms with the MWMOTE technique on the UCI Dataset, which resulted in the best classification algorithm being Decision Tree with an accuracy of 93.73% for imbalanced data and 96.30% for balanced data after being processed using the MWMOTE technique.

Various previous studies have shown that public sentiment analysis can be done using machine learning to understand what the public thinks about an issue and becomes a hot topic discussed on social media. Based on this context, a study was conducted to determine the algorithm that can produce the best classification from popular algorithms used, including Naïve Bayes, Support Vector Machine, and Decision Tree in analyzing public opinion sentiment on Twitter about Covid-19 vaccination in Indonesia.

2. RESEARCH METHOD

The research technique consists of a planned and systematic process to provide a solution to a problem. Figure 1 illustrates the methods that will be applied in this research. Using the sentiment dataset of Covid-19 vaccination in Indonesia from Drone Emprit Academic (DEA) is a good way to start the public sentiment classification analysis project on Covid-19 vaccination in Indonesia. Drone Emprit Academic (DEA) is a social media data analysis platform that collects data from various sources. The collected data is then analyzed and processed to provide useful information for users, such as information on social media trends and user behavior. Moreover, this information can be accessed free of charge by the general public. When using the sentiment dataset of public opinion on Covid-19 vaccination in Indonesia from Drone Emprit Academic (DEA), it is important to ensure that the dataset is relevant to the task at hand, and contains sufficient data to train and evaluate the model. The dataset includes information about tweets and sentiment labels indicating the feelings or opinions of netizens about Covid-19 vaccination in Indonesia. To process the data for use by machine learning algorithms, it may be necessary to clean the data from unnecessary characteristics such as punctuation, stopwords, and special characters, and encode the text to convert the text into numbers or vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method. After processing the data, machine learning algorithms, such as Naïve Bayes, Support Vector Machine, and Decision Tree C4.5, can be trained

with the dataset. It is important to evaluate the performance of the model using appropriate evaluation metrics, and to adjust the model by adjusting hyperparameters, or trying different algorithms if necessary.

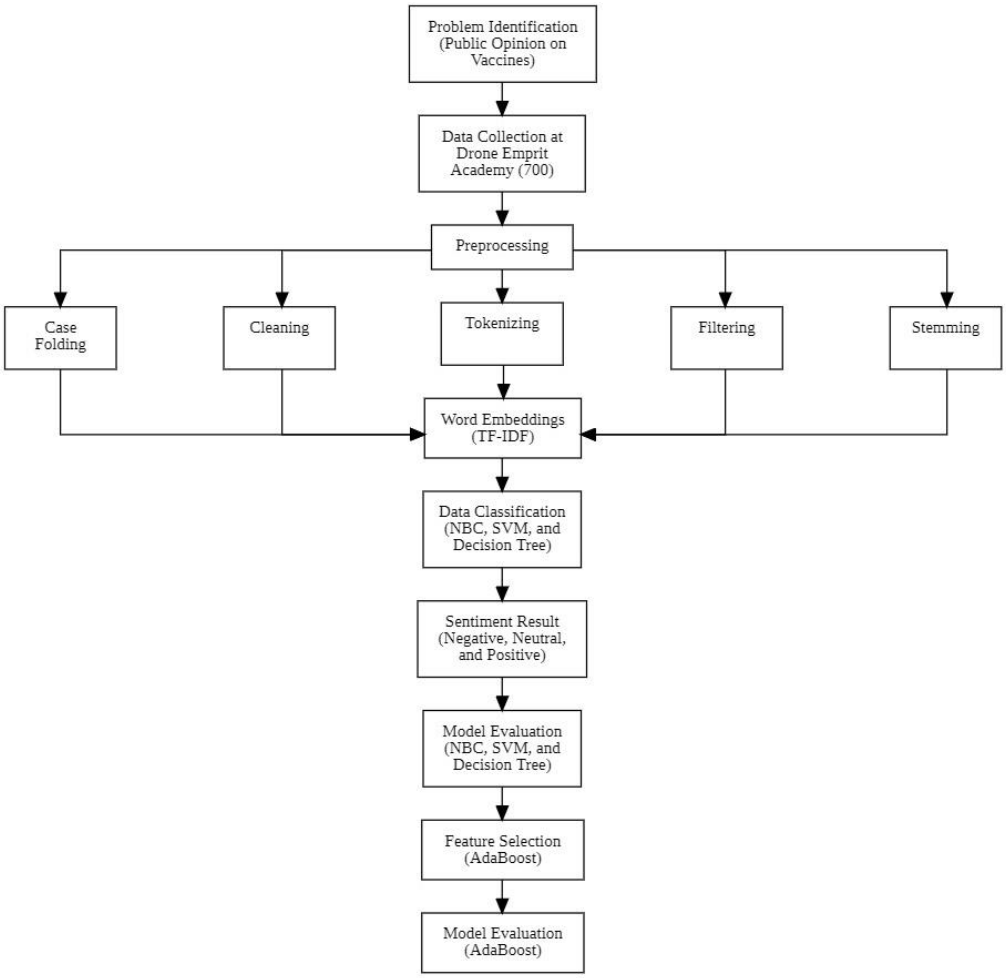

Figure 1. Research Methodology

2.1. Problem Identification

The problem in this research is about public opinion on Covid-19 vaccination in Indonesia on Twitter social media. The aim of this research is to generate a sentiment polarity that indicates public opinion on Covid-19 vaccination in Indonesia.

2.2. Data Collection

Data collection was conducted by retrieving public opinions from the Drone Emprit Academic (dea.uii.ac.id) website. The search was focused on vaccination against COVID-19 in Indonesia using the keywords "vaccine Pfizer and Moderna". The collected data consisted of 700 public comments related to this study.

2.3. Preprocessing

The collected data will undergo preprocessing first. Preprocessing is used to transform unstructured textual data into structured data [25]. The preprocessing process that will be performed in this study consists of 5 steps, namely case folding, cleaning, tokenizing, filtering, and stemming. Case folding is the first preprocessing step that aims to convert all document texts into a standard form (lowercase) [25]. Cleaning is the process of removing numbers, word separators such as commas (,), periods (.), and other punctuation marks. Word cleaning aims to reduce noise [26]. Tokenizing consists of cutting the input string. In this process, some characters (such as punctuation) are removed and spaces are used as separators to separate sentences into sets of words [27]. Filtering is the process of removing meaningless or unimportant words. Stemming is the process of finding the stem (base word) generated from stopword removal (filtering) [28].

2.4 Word Embeddings

Word Embeddings is the process of transforming words into numerical form (word vectors). Word Embeddings is done using Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF is a method that aims to weight the relationship between terms and documents/comments. TF-IDF also evaluates the importance of a word in a document. TF-IDF calculation uses a library in Sklearn python, which is TfidfVectorizer(). The formula for word weighting is as follows:

$$W_{ij} = tf_{ij} \times \log\left(\frac{n}{df}\right) \quad (1)$$

Where wij is the term weight (tj) for the document (dj), tfij is the number of occurrences of term (tj) in (dj), n is the total number of documents, and dfj is the number of documents containing the term.

2.5 Klasifikasi Data

Before classifying the data, there are several things to know about data classification, including:

1. Splitting Data
   Splitting data is used to divide the data into two parts, namely the training data and the testing data. In this study, the research team conducted four experiments as follows:
   Step 0 : 50% of the data used as training data and 50% as testing data.
   Step 1 : 70% of the data used as training data and 30% as testing data.
   Step 2 : 80% of the data used as training data and 20% as testing data.
   Step 3 : 90% of the data used as training data and 10% as testing data.

2. Training Data
   Training data is used to train the system in this study. The purpose of the training data is to train the Naive Bayes, SVM, and Decision Tree classification methods so that they can learn to classify comments as negative, neutral, or positive.

3. Testing Data
   After the Naive Bayes, SVM, and Decision Tree methods are trained, the next step is to evaluate the performance of these methods using testing data. Testing data is used to test the system in this study. The purpose of testing data is to test the Naive Bayes, SVM, and Decision Tree classification methods by inputting new data, then the methods will classify this new data correctly as negative, neutral, or positive.

2.6 Sentiment Result

After sentiment analysis, the results will show data indicating negative, neutral, and positive opinions. Next, the data will be visualized in the form of diagrams to generate graphs of each opinion.

2.7 Model Evaluation

To evaluate the performance of Naïve Bayes, SVM, and Decision Tree methods, it is necessary to test their models. The test results are displayed in the form of a confusion matrix table. At the same time, the accuracy value of the model is obtained by dividing the number of correct data in the classification results by the total data, as shown in the following equation.

$$accuracy = \frac{XX+YY+ZZ}{XX+XY+XZ+YX+YY+YZ+ZX+ZY+ZZ} \quad (2)$$

The accuracy value is calculated by dividing the number of correctly classified instances by the total number of instances. Additionally, model performance evaluation is conducted by examining the accuracy value using the confusion matrix, as well as precision and recall scores for each model. After testing the test data, a list of classes from the test data, called predicted classes, is generated. The predicted classes are then compared to the actual classes of the test data, which were previously unknown. This allows for the calculation of accuracy, precision, recall, and f1-score values. The following are the formulas used to evaluate the model:

$$precision = \frac{TP}{FP+TP} \quad (3)$$

$$accuracy = \frac{TN+TP}{FN+FP+TN+TP} \quad (4)$$

$$recall = \frac{TP}{FN+TP} \quad (5)$$

If described in a confusion matrix, it contains the results of the model testing against the dataset in the form of a table consisting of true and false classes.

Table 1. Confusion Matrix

True Class	Predict Class	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (TP)	True Negative (TN)

Explanation:
TP (true positive) : positive data that are correctly predicted as positive.
TN (true negative) : negative data that are correctly predicted as negative.
FP (false positive) : negative data that are incorrectly predicted as positive.
FN (false negative) : positive data that are incorrectly predicted as negative.

Based on the formula above, precision, recall, and f1-score values are obtained. In this study, python program is used to calculate the precision, recall, and f1-score values. After obtaining the results, the classification method performance in each class is seen from the precision, recall, and f1-score values for each class. The values for precision, recall, and f1-score range from zero to one, and the higher the value, the better and more accurate the model is. The following table summarizes the evaluation results.

Table 2. Evaluasi Model

Classification Types	Precision	Recall	F1-Score
Positive	?	?	?
Negative	?	?	?
Neutral	?	?	?

Therefore, it can be concluded that the evaluation results of the model can be seen from the precision, recall, and f1-score values of each class.

2.8 Feature Extraction and Feature Selection

Feature extraction is the main core of diagnosis, classification, clustering, detection, and recognition. In this study, the features used are divided into two groups; tweet features and class features. The feature selection used is AdaBoost and the feature extraction used is the TF-IDF transformation. The implementation process is done using Python3 program with Jupyter Notebook software to implement AdaBoost to improve the accuracy of the methods used.

Testing is carried out to determine whether AdaBoost is successful in improving the accuracy of Naïve Bayes, SVM, and Decision Tree methods. In this test, four experiments were conducted. The experiments were carried out using AdaBoost-based Naïve Bayes, SVM, and Decision Tree methods. The first experiment was carried out by dividing the test data by 50% and the training data by 50%. The second experiment was carried out by dividing the test data by 30% and the training data by 70%. The third experiment was carried out by dividing the test data by 20% and the training data by 80%. The fourth experiment was carried out by dividing the test data by 10% and the training data by 90%. The results of this experiment will show the performance of Naïve Bayes, Support Vector Machine, and Decision Tree methods based on AdaBoost.

2.9 AdaBoost Evaluation

To evaluate the performance of Naïve Bayes, Support Vector Machine, and Decision Tree methods based on AdaBoost, the models were evaluated. The classification results were presented as a confusion matrix table. The evaluation of the model also provided the accuracy level value. The accuracy level value of the model was calculated by dividing the number of correctly classified data by the total data.

Moreover, the model evaluation process also yielded the values of precision, recall, and f1-score. Python3 was used to calculate the precision, recall, and f1-score values in this study. After obtaining the results, the performance of the classification method for each class could be observed from the precision, recall, and f1-score values for each class. The values of precision, recall, and f1-score range from zero to one, where the higher the value, the better the model performance.

3. RESULTS AND ANALYSIS

The dataset used in this study is a public data sourced from Drone Emprit Academic (DEA) accessed for 3 months (August-October 2021). The dataset consists of 700 tweets on social media Twitter about vaccines.

Before the data is analyzed, a data preprocessing stage is conducted which includes removing null or empty data, converting all text in the document into a consistent letter format, converting all sentences in the document into units of words, removing irrelevant words, URLs, or symbols, removing words with prefixes and suffixes, and transforming the data to fit the algorithm's requirements.

Table 3. The result of text representation using tf-idf

Document index	Word index	Weight
0	795	0.3697612450197021
0	1729	0.3697612450197021
0	1875	0.36095470381707717
0	1340	0.35318678395874953
0	1292	0.11544010027963734
0	1494	0.09161880126751445
0	2076	0.13227904828145906
0	2039	0.5770000466860385
0	396	0.3152408901204811
1	63	0.34937995202670064
1	183	0.25238352128791863
1	80	0.3035117037257312
1	960	0.311856610503672
1	387	0.27000017791326614
1	717	0.3035117037257312
1	600	0.3217265403460804
1	1198	0.3035117037257312
1	517	0.2229550411156791
1	935	0.18631195864170447
1	2078	0.35618795959403693
1	2110	0.21052708869751274
1	1494	0.07520360459769337
1	2076	0.05428940952016844
2	1382	0.4483043690068737
…	…	…
699	1332	0.6596854824792299
699	1857	0.45539520557669205
669	948	0.5256600323661662
699	1292	0.1872634712812819
699	2076	0.21457910811724853

After preprocessing the data into the appropriate format, the dataset consisted of 700 data points. Next, we implemented the Naive Bayes, Support Vector Machine, and Decision Tree models, and tested them based on several stages of data splitting, using feature extraction and feature selection.

The results of this study on four data splitting ratios, namely 50:50, 70:30, 80:20, and 90:10, using feature selection are presented in the following table and comparison graph.

Tabel 4. Comparison of Test Results of Methods based on Accuracy with Feature Selection

Splitting Data	Without K-Fold Cross Validation			With K-Fold Cross Validation			Accuracy Improvement with Adaboost		
	Naïve Bayes	SVM	Decision Tree	Naïve Bayes	SVM	Decision Tree	Naïve Bayes	SVM	Decision Tree
50 : 50	0,7	0,69	0,71	0,69	0,7	0,72	0,7	0,7	0,69
70 : 30	0,72	0,71	0,74	0,68	0,72	0,64	0,72	0,72	0,71
80 : 20	0,73	0,71	0,76	0,7	0,73	0,78	0,74	0,73	0,71
90 : 10	0,71	0,72	0,77	0,71	0,81	0,66	0,76	0,71	0,72

In the table and graph above, the comparison results are shown using feature selection, where the highest accuracy obtained is 0.81 with a 90:10 data splitting ratio using the AdaBoost Naïve Bayes method. The following are the results of this study for four data splitting ratios of 50:50, 70:30, 80:20, and 90:10 without using feature extraction, presented in the following table and comparison graph.

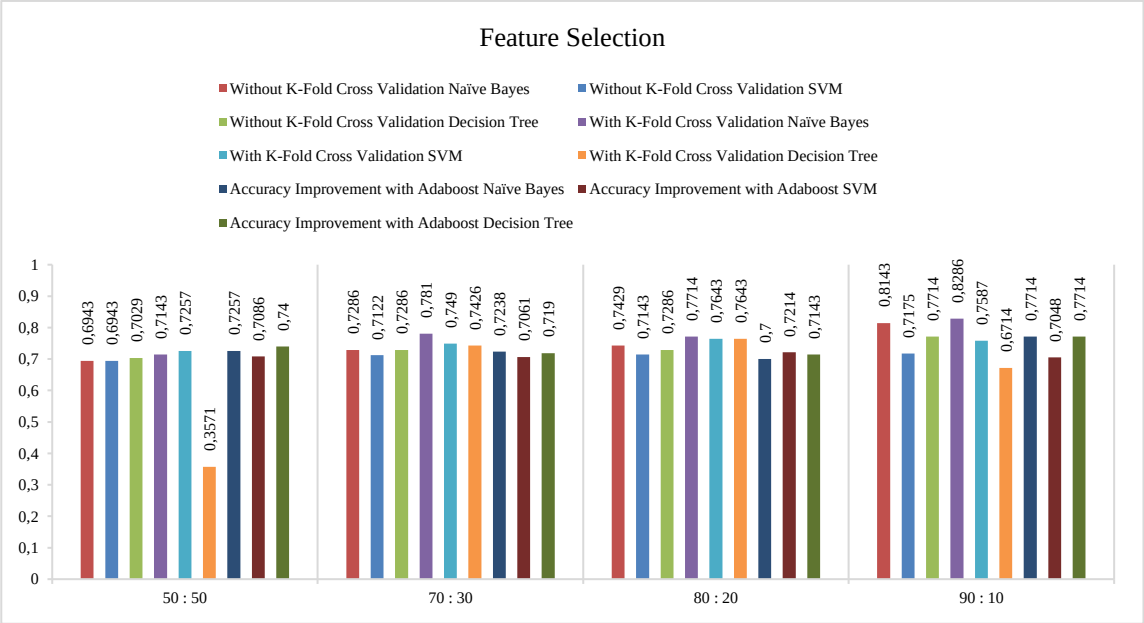

Figure 2. Graph of Comparison Results of Method Testing based on Accuracy with Feature Selection

Table 5. Comparison of Test Results of Methods based on Accuracy without Feature Extraction

Splitting Data	*Naïve Bayes*	*Naïve Bayes + K-Fold*	*Naïve Bayes + Adaboost*	*SVM*	*SVM + K-Fold*	*SVM + Adaboost*	*Decision Tree*	*Decision Tree + K-Fold*	*Decision Tree + Adaboost*
50 : 50	0,6886	0,6886	0,7	0,7057	0,7343	0,7171	0,6829	0,69	0,72
70 : 30	0,7143	0,7061	0,7238	0,7762	0,7510	0,6381	0,7286	0,7	0,73
80 : 20	0,7214	0,7143	0,7286	0,7643	0,7643	0,7643	0,7143	0,7196	0,7143
90 : 10	0,8	0,7175	0,8143	0,8286	0,7619	0,6571	0,7857	0,7016	0,7857

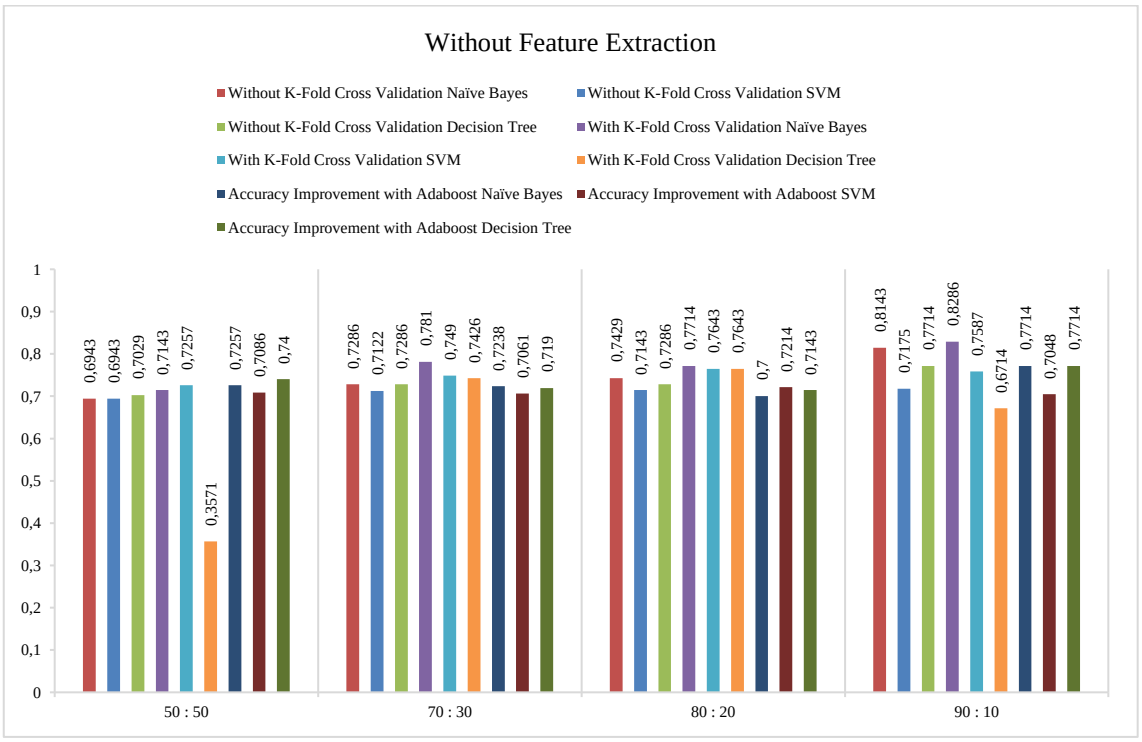

Figure 3. Graph of Comparison Results of Method Testing based on Accuracy without Feature Extraction

The table and graph above show the comparison results using feature extraction, where the highest accuracy achieved is 0.8157 with 90:10 data splitting using AdaBoost Naïve Bayes. Here are the results of this

study for four data splitting ratios, namely 50:50, 70:30, 80:20, and 90:10 with feature extraction, presented in the following table and graph.

Table 6. Comparison of Test Results of Methods based on Accuracy with Feature Extraction

Splitting Data	*Naïve Bayes*	*Naïve Bayes + K-Fold*	*Naïve Bayes + Adaboost*	*SVM*	*SVM + K-Fold*	*SVM + Adaboost*	*Decision Tree*	*Decision Tree + K-Fold*	*Decision Tree + Adaboost*
50 : 50	0,6943	0,6943	0,7029	0,7143	0,7257	0,3571	0,7257	0,7086	0,74
70 : 30	0,7286	0,7122	0,7286	0,7810	0,7490	0,7426	0,7238	0,7061	0,7190
80 : 20	0,7429	0,7143	0,7286	0,7714	0,7643	0,7643	0,7	0,7214	0,7143
90 : 10	0,8143	0,7175	0,7714	0,8286	0,7587	0,6714	0,7714	0,7048	0,7714

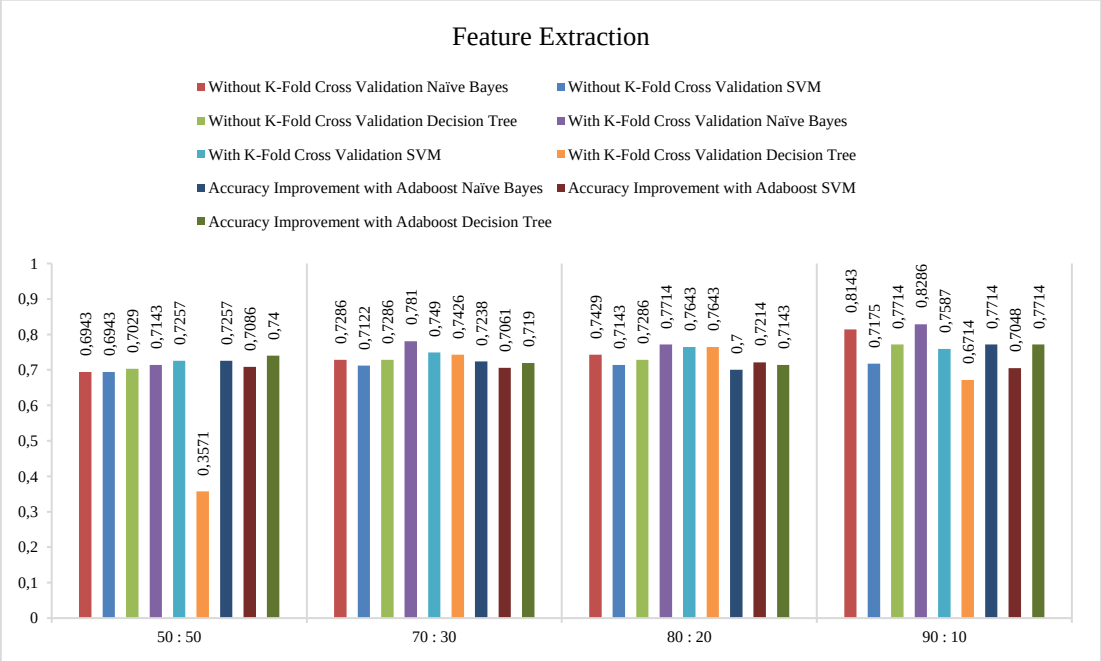

Figure 4. Graph of Comparison Results of Method Testing based on Accuracy with Feature Extraction

From the table and graph above, it can be seen that the highest accuracy obtained without using feature extraction is 0.8286 with a 90:10 data splitting ratio using the SVM method. Here are the results of this study on four data splitting ratios, namely 50:50, 70:30, 80:20 and 90:10 using feature extraction, presented in the following table and comparison graph. The results of sentiment analysis are in the form of negative, neutral, and positive opinion categories. For more details, please refer to the table below.

Table 7. Distribution of Public Opinion Categories

Category	Count
Negative	118
Neutral	15
Positive	217

Based on the table above, the public opinion shows a higher positive value of 217 in the research data. The word 'vaccine' has the highest frequency of appearance, followed by 'Pfizer', 'Moderna', 'AstraZeneca', 'COVID', 'Indonesia', 'dose', and 'vaccination'. The results indicate that the positive category is more dominant than neutral and negative categories.

4. CONCLUSION

Based on the implementation, testing, and evaluation conducted in the previous chapters, it can be concluded that the government's COVID-19 vaccination program has received a positive response. This can be seen from the sentiment analysis of COVID-19 vaccination data with a test data of 350 for the three algorithms, namely Naïve Bayes, SVM, and Decision Tree. From the three algorithms used for classification with 700 vaccine data, it can be concluded that the SVM algorithm outperforms the other four Splitting Data. This indicates that SVM is the best algorithm to use for classification on vaccine data.

In the experiment with the Naïve Bayes algorithm with accuracy improvement using Feature Selection Adaboost, there was an increase in accuracy for each splitting data. This indicates that Adaboost is an effective Feature Selection technique for the Naïve Bayes algorithm. Using Feature Extraction on the Naïve Bayes algorithm, two splitting data outperformed, namely Naïve Bayes pure and Naïve Bayes + Adaboost. This indicates that Feature Extraction can improve the performance of the Naïve Bayes algorithm on vaccine datasets. For three splitting data with Feature Extraction of the SVM algorithm, it can be concluded that pure SVM outperforms in accuracy. This indicates that Feature Extraction does not always improve the performance of the SVM algorithm on vaccine datasets. For two splitting data with Feature Extraction using the pure Decision Tree algorithm and Decision Tree + Adaboost, the accuracy outperforms. This indicates that the Decision Tree can be used for classification on vaccine datasets effectively.

Overall, the testing and evaluation results show that the SVM algorithm is the best algorithm to use for classification on vaccine datasets with the highest accuracy. However, Naïve Bayes and Decision Tree can also be used as alternatives if SVM cannot be used. Additionally, the use of Feature Selection and Feature Extraction techniques can improve the performance of all three algorithms on vaccine datasets.

REFERENCES

[1] Aljameel SS, Alabbad DA, Alzahrani NA, Alqarni SM, Alamoudi FA, Babili LM, et al. A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. Int J Environ Res Public Health 2021;18:218.

[2] Rachman FF, Pramana S. Analisis sentimen pro dan kontra masyarakat Indonesia tentang vaksin COVID-19 pada media sosial Twitter. Indonesian of Health Information Management Journal (INOHIM) 2020;8:100–9.

[3] Ratino RR, Hafidz NHH, Anggraeni SAA, Gata WGG. Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes. JUPITER (Jurnal Penelitian Ilmu Dan Teknik Komputer) 2020;12:1–11.

[4] Indah RNG, Novita R, Kharisma OB, Vebrianto R, Sanjaya S, Andriani T, et al. DBSCAN algorithm: twitter text clustering of trend topic pilkada pekanbaru. J Phys Conf Ser, vol. 1363, IOP Publishing; 2019, p. 012001.

[5] Kartino A, Anam MK. Analisis Akun Twitter Berpengaruh terkait Covid-19 menggunakan Social Network Analysis. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi) 2021;5:697–704.

[6] Que VKS, Iriani A, Purnomo HD. Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. Jurnal Nasional Teknik Elektro Dan Teknologi Informasi 2020;9.

[7] Kurniawan R, Apriliani A. Analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter berbasis web scraper. Jurnal INSTEK (Informatika Sains Dan Teknologi) 2020;5:67–75.

[8] Mahardhika YS, Zuliarso E. Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier 2018.

[9] Vitandy SWU, Supianto AA, Bachtiar FA. Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naïve Bayes Classifier. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN 2019;2548:964X.

[10] Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence–enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. J Med Internet Res 2021;23:e26627.

[11] Gunawan B, Sastypratiwi H, Pratama EE. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. JEPIN (Jurnal Edukasi Dan Penelitian Informatika) 2018;4:113–8.

[12] Sari FV, Wibowo A. Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer 2019;10:681–6.

[13] Permadi VA. Analisis sentimen menggunakan algoritma Naïve Bayes terhadap review restoran di Singapura. Jurnal Buana Informatika 2020;11:141–51.

[14] Syarifuddinn M. Analisis sentimen opini publik terhadap efek PSBB pada twitter dengan algoritma decision tree, knn, dan naïve bayes. INTI Nusa Mandiri 2020;15:87–94.

[15] Puspita R, Widodo A. Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. J Inform Univ Pamulang 2021;5:646.

[16] Romadloni NT, Santoso I, Budilaksono S. Perbandingan Metode Naïve Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. Ikraith-Informatika 2019;3:1–9.

[17] Chinnasamy P, Suresh V, Ramprathap K, Jebamani BJA, Rao KS, Kranthi MS. COVID-19 vaccine sentiment analysis using public opinions on Twitter. Mater Today Proc 2022;64:448–51.

[18] Prasad KS, Reddy NCS, Puneeth BN. A framework for diagnosing kidney disease in diabetes patients using classification algorithms. SN Comput Sci 2020;1:101.

[19] Sari IC, Ruldeviyani Y. Sentiment analysis of the covid-19 virus infection in indonesian public transportation on twitter data: A case study of commuter line passengers. 2020 International Workshop on Big Data and Information Security (IWBIS), IEEE; 2020, p. 23–8.

[20] Lutfi AA, Permanasari AE, Fauziati S. Sentiment analysis in the sales review of Indonesian marketplace by utilizing Support Vector Machine. Journal of Information Systems Engineering and Business Intelligence 2018;4:57–64.

[21] Rahmaddeni R, Anam MK, Irawan Y, Susanti S, Jamaris M. Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination. ILKOM Jurnal Ilmiah 2022;14:32–8.

[22] Khakharia A, Shah V, Gupta P. Sentiment analysis of COVID-19 vaccine tweets using machine learning. Available at SSRN 3869531 2021.
[23] Rathi M, Malik A, Varshney D, Sharma R, Mendiratta S. Sentiment analysis of tweets using machine learning approach. 2018 Eleventh international conference on contemporary computing (IC3), IEEE; 2018, p. 1–3.
[24] Untoro MC, Praseptiawan M, Widianingsih M, Ashari IF, Afriansyah A. Evaluation of Decision Tree, k-NN, Naive Bayes and SVM with MWMOTE on UCI Dataset. J Phys Conf Ser, vol. 1477, IOP Publishing; 2020, p. 032005.
[25] Samsir S, Ambiyar A, Verawardina U, Edi F, Watrianthos R. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. Jurnal Media Informatika Budidarma 2021;5:157–63.
[26] Luqyana WA, Cholissodin I, Perdana RS. Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN 2018;2:4704–13.
[27] Manuaba IBNW, Dantes GR, Indrawan G. Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine Dengan Penambahan Algoritma Levenshtein Distance. Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan) 2022;5:9–17.
[28] Luqyana WA, Cholissodin I, Perdana RS. Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer 2018;2:4704–13.

BIBLIOGRAPHY OF AUTHORS

Rahmaddeni was born in Padang and currently a Lecturer from Department of Informatics at STMIK Amik Riau. He actively teaches in the fields of Machine Learning and Data Mining. Received Bachelor's Degree in Informatic Engineering Department in STMIK Amik Riau and Master's Degree in Informatic Engineering in Universitas Putra Indonesia "YPTK" Padang. The focus of he research is Machine Learning and Data Mining.

Firman Akbar, student at the STMIK Amik Riau School of Informatics & Computer Management, departement of informatics engineering for the 2021 academic year who is interested in data science. The focus of researcher is data mining, deep learning, and machine learning.