International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Best Algorithm in Sentiment Analysis of Presidential Election in Indonesia on Twitter

April Lia Hananto[1,] Aprilia Putri Nardilasari[2], Ahmad Fauzi[3], Agustia Hananto [4], Bayu Priyatna[5], Aviv Yuniar Rahman[6]

Submitted: 12/02/2023 **Revised:** 13/04/2023 **Accepted:** 12/05/2023

Abstract: The election of presidential candidates for 2024 is included in the democratic process to elect the president and vice president for the period 2024-2029. In this case, there are already names of presidential candidates who have been nominated and many survey institutions have published survey results on several candidates who are eligible to become presidential candidates, based on this, not a few netizens have expressed their opinions that can be made regarding public sentiment. about the trend of presidential candidates which is currently being discussed on Twitter social media. In this study, public sentiment analysis was carried out on trends in presidential candidates by comparing three classification algorithms, namely support vector machine (SVM), K-Nearest Neighbor (K-NN) and Naïve Bayes (NB). Comparisons are made to find out which algorithm has better accuracy. This research is also expected to provide references and knowledge to the public about the trends of presidential candidates in the upcoming presidential election. The data taken are 9966 twitter data regarding presidential and presidential candidates as well as tweet data taken in the second week of 09-17 September 2022. The results of this test concluded that the SVM algorithm is superior to K-NN and Naïve Bayes which get an accuracy rate of 79.57%. The results of this study get the best and most effective algorithm in classifying positive and negative comments on the 2024 presidential candidate trend.

Keywords- *Capres, Pilpres, Sentiment Analysis, SVM, Naive Bayes, KNN*

1. Introduction

Indonesia is a country with a democratic system. The democratic system here refers to the election of the president and vice president. Elections in democracies are usually held periodically. In 2024, presidential and vice presidential elections will be held. The trend of the popularity of presidential candidates is currently being hotly discussed by the public through social media, especially social media Twitter. Unlike people in the past who conveyed their criticisms, suggestions, and opinions through print media, not a few people have writing skills or have the opportunity to publish their writings. However, currently, the advancement of communication technology has made changes, one of which is the habit of people expressing their opinions on social media networks. One of the social media that is known among internet users today is Twitter. Indonesia is the fourth most populous democracy in the world and the fifth most Twitter users in the world. The 2024 presidential election in Indonesia has become an interesting topic for Twitter users as well as an effective and efficient campaign media[1].

Twitter is an internet service or online social media service that facilitates its users to read and send messages of up to 280 characters called tweets. Previously, Twitter messages were limited to 140 characters and later increased to 280 characters on November 7, 2017. The services that Twitter provides to its users include creating statuses called Tweets that other Twitter users can read, and Twitter is a website that provides a collection of opinion data from people around the world. As a result of channelling opinions and comments, Tweets become a source of information that can be used to analyze public opinion against institutions and individuals[2],[3]. This is because Tweets contain sentiments that can be used as a measure of public opinion that can be used as a basis for evaluation in the future. To determine the sentiment of a tweet, we can divide it into two sentiments, namely positive and negative. Opinions from tweets can serve to assess the sentiments submitted, one of which is opinions about the trend of political figures who are ready to run for president of Indonesia in 2024[4].

Sentiment analysis is part of Natural Language Processing (NLP) which is useful in creating a system to recognize, identify, and obtain opinions in text form and is also a process that functions to identify forms of opinion or sentiment from content[5]. Dataset, in the form of text about positive and negative topics or events Information in text form is now widely distributed on the internet in the

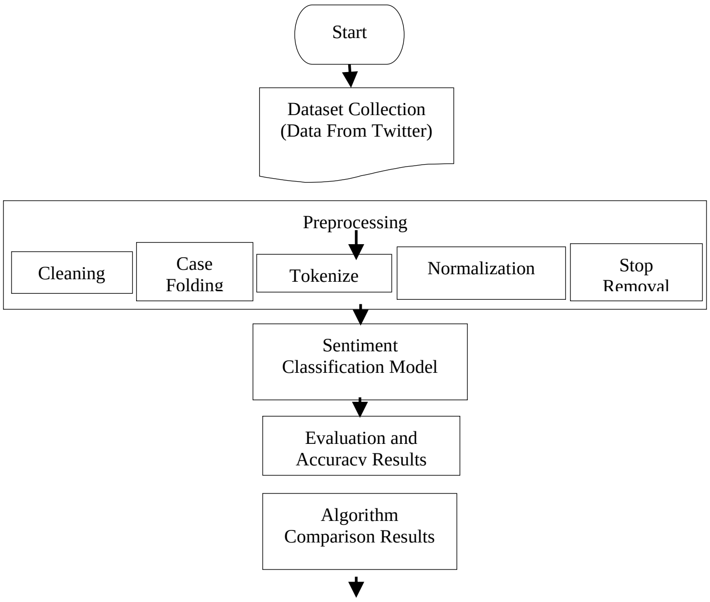
[1,2,4,5] Faculty of Computer Science, Information System Study Program, Universitas Buana Perjuangan Karawang,
[5]Faculty of Computer Science, Informatics Engineering Study Program, Universitas Buana Perjuangan Karawang
[6]Faculty of Engineering, Informatics Engineering Study Program, Universitas Widyagama Malang
[1]aprilia@ubpkarawang.ac.id,
[2]si19.aprilianardilasari@mhs.ubpkarawang.ac.id,
[3]agustia.hananto@ubpkarawang.ac.id, [4]afauzi@ubpkarawang.ac.id,
[5]bayu.priyatna@ubpkarawang.ac.id, [6]aviv@widyagama.ac.id

form of blogs, forums, social media, and sites containing reviews. Positive or negative sentiments from an opinion can be processed manually, but of course, the more the number of opinion sources, the more time and energy are needed to classify the polarity of the opinion. Therefore, it is proposed to apply machine learning methods to classify the polarity of the pinion of this large data source. Thus, you can take advantage of the text mining function[6].

There are several classification techniques, including SVM, KNN and Naïve Bayes. Classification algorithms have advantages and disadvantages for classifying text data. In this study, a comparison was made between the three classification algorithms Support Vector Machine, Naïve Bayes and K-Nearest Neighbor in terms of accuracy. And what kind of data classification results will be visualized as well as a more appropriate algorithm for classifying public opinion data on Twitter. In conducting this research, it is done by observation and literature study[7],[8].

2. Research Methodology

The research methodology is a stage of systematic research carried out from research preparation to final result collection[9], [10]. At this stage, it will be explained about the method used in this research, by going through several stages, namely collecting data, processing data, classifying text using 3 algorithm models and classification results. The following are the stages of the flow in this research.


Start
Dataset Collection (Data From Twitter)
Preprocessing
Cleaning
Case Folding
Tokenize
Normalization
Stop Removal
Sentiment Classification Model
Evaluation and Accuracy Results
Algorithm Comparison Results

2.1 Data Collection Method

Various methods of generating data or information to solve a problem. The methods used are:

a. Literature study

data collection using the library method; is carried out through the collection of literature, journals, books, papers, and online media sites as library sources [11],[12] that are relevant to the topic of writing, namely comparison of sentiment analysis using the SVM, Naïve Bayes, and KNN algorithms.

b. Twitter Data Collection

The data that was successfully obtained was sourced from data taken directly from Twitter. The data is the public's opinion of the Indonesian presidential candidate 2024 on Twitter social media using the search keywords #capres and #pilpres. The data obtained and the amount of data collected amounted to 9966 tweet data.

2.2 Preprocessing

There are several methods used in processing sentiment analysis of text data using rapid miner tools such as [13]:

a. Cleansing is the process of cleaning text data by removing irrelevant and inconsistent data. Cleaning functions to remove unimportant characters such as hashtags (#), numbers, usernames (@), URLs, punctuation marks, and emoticons [14].

b. Case folding is the process of changing the word form of a character so that it becomes uniform (lowercase).

c. Tokenize is the step to divide the text into token parts. The initial stage of preprocessing text is tokenization, which is a tool to break text into smaller parts (sentences, words, and bigrams) [15].

d. StopWord filter is the process of removing words in Indonesian. Remove stopwords to remove or as an eraser every word that can be ignored. The StopWord filter is used by the Indonesian stopwords dictionary [16].

2.3 Classification Model

The first algorithm is Naïve Bayes. Naïve Bayes is a simple machine learning algorithm according to the Bayes theorem invented by Thomas Bayes in the 18th century. In

Thomas Bayes' theory a conditional probability with the following equation:

$$P(F|X) = \frac{p(X|F)P(F)}{p(x)}$$

P(F|X) = Probability that the hypothesis is true for observed sample X data P(X|F) = Probability of sample X data if it is assumed that the hypothesis is true. P(F) = Probability of the hypothesis F P(X) = Probability of observed sample data [17].

The next algorithm is SVM. SVM is a machine learning algorithm that serves to analyze sentiment. The purpose of the SVM algorithm is to apply a hyperplane task to the data to create a regional shape for each class. The hyperplane is in charge of separating the existing classes.

The third classification method is KNN. KNN is an algorithm method in charge of classifying text and data by classifying objects based on the closest distance to the object. The KNN algorithm aims to classify objects based on attributes and examples of training data.

All the algorithms described above will be compared to assess the performance of each sentiment data classification algorithm using the K-fold cross-validation evaluation technique. K-fold cross-validation serves to find out the best value of a system by repeating and randomizing input attributes until the system is tested [18],[19].

2.4 Evaluation and Final Results

This evaluation serves to measure the accuracy value using K-fold cross-validation. The evaluation here is to analyze the results of the classification. The evaluation process uses a confusion matrix. The validation used is by dividing the training data and test data [13]. This evaluation process serves to see the performance of the classification model that has been processed and determine its accuracy. All test data that has been collected is then divided into four categories, namely:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

TP : True Positive
TN : True Negative
FP : False Positive
FN : False Negative

Accuracy is obtained by calculating the total value of True divided by the sum of all data. In this study, the precision and recall of the classification model were also calculated [16].

2.5 Comparison of Algorithm Methods

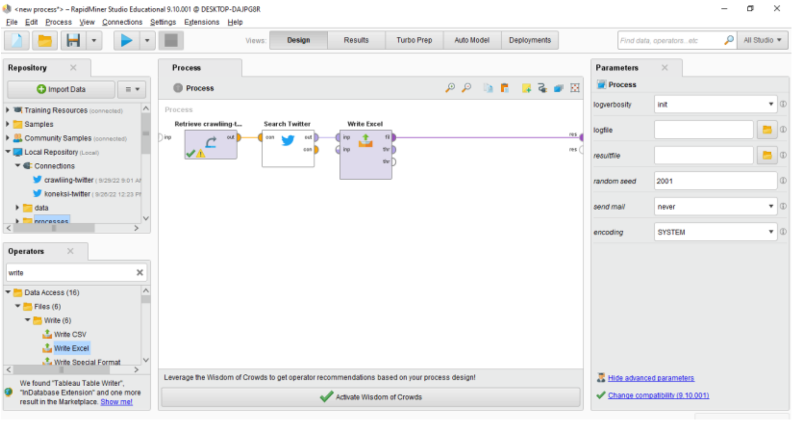
After the evaluation process, the next step is to compare the level of accuracy between the SVM, Naïve Bayes, and KNN algorithms to see a better algorithm in the accuracy value obtained.

3. Result and Discussion

This result and discussion will explain the stages in each algorithm that will be compared in analyzing sentiment using the Rapid Miner application. The steps of the process of results and discussion in this study, among others, are as follows:

3.1 Data Collection

The data collection obtained in this study is through the Twitter Crawler data using RapidMiner. The data obtained were 9966 tweets with the keywords 'candidate' and 'election'. After that, delete duplicate data by doing data cleaning. The cleaning data obtained are 3472. Then the data that has been successfully collected can be saved in the form of .csv or .xlsx. The following is an example of crawling data from twitter which can be seen in table 1.



<new process*> – RapidMiner Studio Educational 9.10.001 @ DESKTOP-DAJPG8R
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
crawling-twitter (9/29/22 9:01 AM
koneksi-twitter (9/26/22 12:23 PM
data
processes
Operators
write
Data Access (16)
Files (6)
Write (6)
Write CSV
Write Excel
Write Special Format
Process
Process
Retrieve crawling-t...
Search Twitter
Write Excel
inp
out
con
fil
thr
res
Parameters
Process
Hide advanced parameters
Change compatibility (9.10.001)
Activate Wisdom of Crowds

We found "Tableau Table Writer", "InDatabase Extension" and one more result in the Marketplace. Show me!

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

logverbosity: init
logfile:
resultfile:
random seed: 2001
send mail: never
encoding: SYSTEM

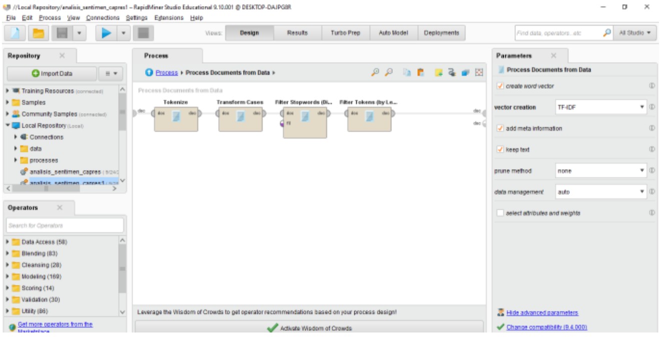
Fig 1. Preparation for Crawling Data

Table 1. Example of Crawling Result Data

No	Tweet
1	RT @aLy_Bima: Ini baru servei original tanpa margin error. Fiks ya bang anies capres terkuat versi survei sungguhan. https://t.co/uqnkWOMB…
2	mantap sekali banyak yang dukung pak prabowo menjadi capres tahun 2024 semangat pak #PrabowoPalingLayak https://t.co/WnPF4Q5VXe
3	Memang untuk sementara kita masih menunggu keputusan kpu siapa siapa nama capres-cawapres untuk tahun 2024 mendatang nanti klo sudah ada keputusan kpu yg tetap baik capres-cawapres maupun partai2 pengusungnya baru kita mulai perang urat syaraf,debat, adu argumen dll,wait and see https://t.co/yvgcpkikpo

3.2 Data Preprocessing

The next stage is preprocessing. The steps carried out in the preprocessing process are as follows:


//Local Repository/analisis_sentimen_capres1 – RapidMiner Studio Educational 9.10.001 @ DESKTOP-DAJPG8R
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
analisis_sentimen_capres
analisis_sentimen_capres1
Operators
Search for Operators
Data Access (58)
Blending (83)
Cleansing (28)
Modeling (169)
Scoring (14)
Validation (30)
Utility (86)
Get more operators from the Marketplace
Process
Process ▸ Process Documents from Data ▸
Process Documents from Data
Tokenize
Transform Cases
Filter Stopwords (Di...
Filter Tokens (by Le...
doc
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Hide advanced parameters
Change compatibility (9.4.000)

Process Documents from Data

☑ create word vector

vector creation: TF-IDF

☑ add meta information

☑ keep text

prune method: none

data management: auto

□ select attributes and weights

Fig 2. Preprocessing Data

Table 2. Example of Cleaning Result Data

Text	*Cleaning*
RT @RizkiR4madani: PDIP kemungkinan besar akan menang dlm kontestasi pilpres2024 apabila mencalonkan @ganjarpranowo sebagai calon presiden…	PDIP kemungkinan besar akan menang dlm kontestasi pilpres2024 apabila mencalonkan sebagai calon presiden…
@alimuhsuharli Semoga tidak salah pilih Capres 2024, bpk @aniesbaswedan lah yg dapat merubah saat ini. Save. #AniesBaswedan	Semoga tidak salah pilih Capres 2024 bpk lah yg dapat merubah saat ini Save AniesBaswedan
RT @cnxy3_: Puan didorong jadi capres PDIP	Puan didorong jadi capres PDIP

Puan Next President #CapresKuatPDIP https://t.co/tVABS5ad1O	Puan Next President

Table 3. Example of Case Folding Result Data

Text	*Case Folding*
PDIP kemungkinan besar akan menang dlm kontestasi pilpres2024 apabila mencalonkan sebagai calon presiden…	pdip kemungkinan besar akan menang dlm kontestasi pilpres2024 apabila mencalonkan sebagai calon presiden…
Semoga tidak salah pilih Capres 2024 bpk lah yg dapat merubah saat ini Save AniesBaswedan	semoga tidak salah pilih capres 2024 bpk lah yg dapat merubah saat ini save aniesbaswedan
Puan didorong jadi capres PDIP Puan Next President	puan didorong jadi capres pdip puan next president

Table 4. Example of Tokenize Result Data

Text	*Tokenize*
pdip kemungkinan besar akan menang dlm kontestasi pilpres2024 apabila mencalonkan sebagai calon presiden…	['pdip', 'kemungkinan', 'besar', 'akan', 'menang', 'dlm', 'kontestasi', 'pilpres2024', 'apabila', 'mencalonkan', 'sebagai', 'calon', 'presiden']
semoga tidak salah pilih capres 2024 bpk lah yg dapat merubah saat ini save aniesbaswedan	['semoga:', 'tidak', 'salah', 'pilih', 'capres', '2024', 'bpk', 'lah', 'yg', 'dapat', 'merubah', 'saat', 'ini', 'save', 'aniesbaswedan']
puan didorong jadi capres pdip puan next president	['puan', 'didorong', 'jadi', 'capres', 'pdip', 'puan', 'next', 'president']

Table 5. Example of Remove Stopwords Result Data

Text	*Remove Stopwords*
['pdip', 'kemungkinan', 'besar', 'akan', 'menang', 'dlm', 'kontestasi', 'pilpres2024', 'apabila',	['pdip','menang', 'kontestasi', 'pilpres2024', 'mencalonkan', 'calon', 'presiden']

'mencalonkan', 'sebagai', 'calon', 'presiden']	
['semoga:', 'tidak', 'salah', 'pilih', 'capres', '2024', 'bpk', 'lah', 'yg', 'dapat', 'merubah', 'saat', 'ini', 'save', 'aniesbaswedan']	['semoga:', 'tidak', 'salah', 'pilih', 'capres', '2024', 'merubah', 'save', 'aniesbaswedan']
['puan', 'didorong', 'jadi', 'capres', 'pdip', 'puan', 'next', 'president']	['puan', 'didorong', 'capres', 'pdip', 'puan', 'president']

3.3 Classification Model

The classification model in this study uses three methods: SVM, Naïve Bayes and KNN. In this process, data mining techniques are used using the RapidMiner application with a model design that can be seen in Figure 4. Before processing, the data will go through the SMOTE Upsampling operator which is useful so that the class is balanced. The data file was obtained using the Read Excel operator.

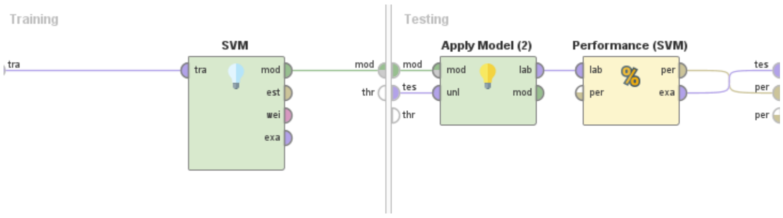


Fig 3. SVM Cross Validation Process

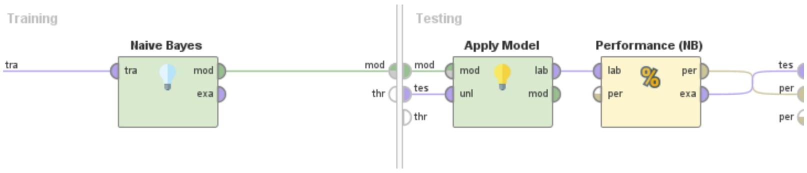


Fig 4. NB Cross Validation Process

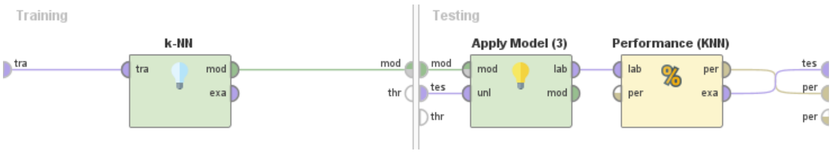


Fig 5. KNN Cross Validation Process

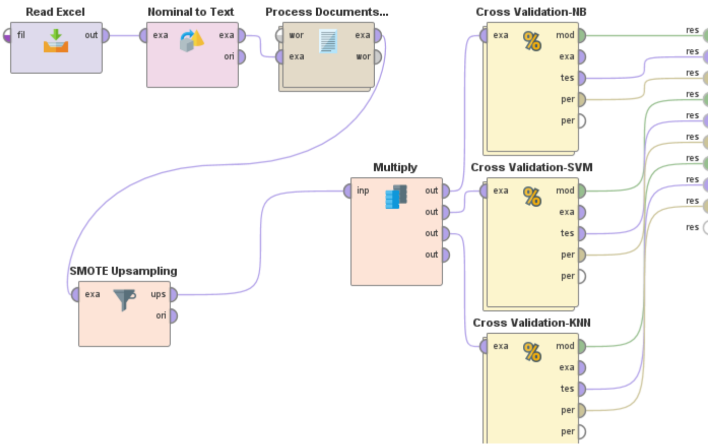
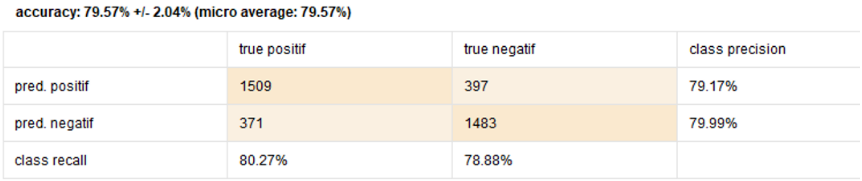


Fig 6. Data Mining Classification Model Design Comparison of SVM, NB, and KNN

3.4 Evaluation and Final Results

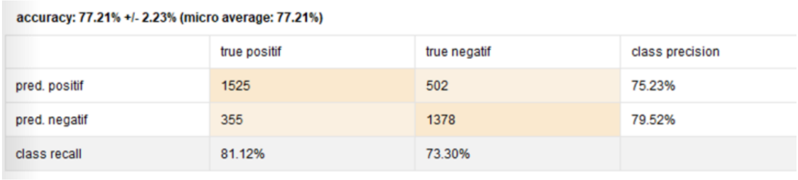
At the evaluation stage and the accuracy, results are obtained to determine the value of the classification comparison that was successfully built in the previous classification model process [20]. The evaluation method used is the 10 K-fold validation method. In determining the accuracy results, an evaluation measurement is needed called a confusion matrix. The confusion matrix is a table that contains most of the rows of testing data that are predicted to be true and false by the classification model used in determining the performance of a classification model that can produce recall, precision and accuracy [14].



accuracy: 79.57% +/- 2.04% (micro average: 79.57%)

	true positif	true negatif	class precision
pred. positif	1509	397	79.17%
pred. negatif	371	1483	79.99%
class recall	80.27%	78.88%	

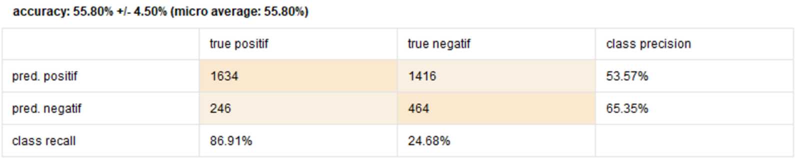
Fig 7. SVM Algorithm Accuracy Results



accuracy: 77.21% +/- 2.23% (micro average: 77.21%)

	true positif	true negatif	class precision
pred. positif	1525	502	75.23%
pred. negatif	355	1378	79.52%
class recall	81.12%	73.30%	

Fig 8. NB Algorithm Accuracy Results



accuracy: 55.80% +/- 4.50% (micro average: 55.80%)

	true positif	true negatif	class precision
pred. positif	1634	1416	53.57%
pred. negatif	246	464	65.35%
class recall	86.91%	24.68%	

Fig 9. SVM Algorithm Accuracy Results

3.5 Comparison of Algorithm Methods

The classification model runs well from the results of the comparison of experiments that have been carried out. The accuracy level of each algorithm is classified based on its AUC value as follows: 0.50-0.60=Failed. 0.60-0.70=Poor. 0.70-0.80=Enough. 0.80-0.90=Good. 0.90-1.00=Very good. [20]. Accuracy results can be seen in Figure 12 below.

Table 6. Comparison of Algorithm Classification

Algorhythm	Accuracy	AUC	Precision	Recall	Classification
SVM	79.57%	0.883	79,17%	80,27%	Good
NB	77.21%	0.706	75.23%	81,12%	Enough
KNN	55.80%	0.763	53.57%	86,91%	Enough

Based on the results above, it can be seen that the SVM algorithm has the best performance with an accuracy rate of 79.57% when compared to the Naïve Bayes algorithm 77.21%, and KNN 55.80%.

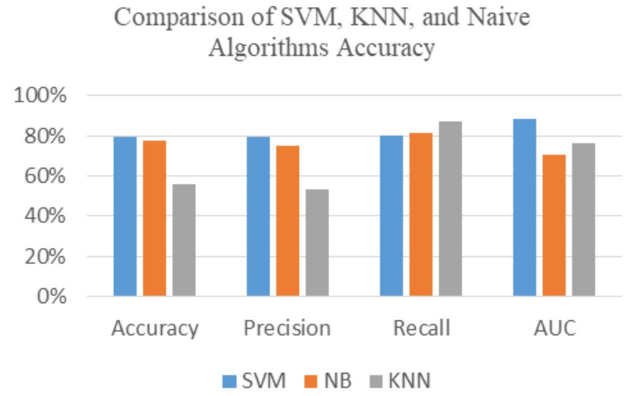


Fig 10. Graph of Comparison of Classification Algorithms

The comparison results are depicted by graphical data which can be seen in Figure 10. Based on the comparison of each model, the average results of the classification model are obtained, and the SVM Algorithm has the highest average of the NB and KNN algorithm models. Except for the recall results, KNN has a higher value.

4. Conclusion

Based on the research results, it can be concluded that the Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Naïve Bayes methods can be implemented for sentiment classification on the popularity of the 2024 presidential candidate trend. The results of this analysis are categorized into positive and negative sentiment labels. .

The results of this processing and testing show that the accuracy rate of the SVM algorithm is 79.57% and the Naïve Bayes algorithm has an accuracy rate of 77.21% and the KNN algorithm is 55.80%. The test shows that the performance of the SVM algorithm is superior to the NB and KNN algorithms for testing the 2024 presidential candidate trend case. The results above can be used as reference material by using different data sources for further research.

References

[1] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, “A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis,” *Expert Syst. Appl.*, vol. 110, pp. 298–310, Nov. 2018, doi: 10.1016/j.eswa.2018.06.022.

[2] A. Mittal and S. Patidar, “Sentiment analysis on twitter data: A survey,” *ACM Int. Conf. Proceeding Ser.*, pp. 91–95, 2019, doi: 10.1145/3348445.3348466.

[3] N. Naw, “Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers,” *Int. J. Sci. Res. Publ.*, vol. 8, no. 10, 2018, doi: 10.29322/ijsrp.8.10.2018.p8252.

[4] F. Firmansyah *et al.*, “Comparing Sentiment Analysis of Indonesian Presidential Election 2019 with Support Vector Machine and K-Nearest Neighbor Algorithm,” in *2020 6th International Conference on Computing Engineering and Design (ICCED)*, Oct. 2020, pp. 1–6. doi: 10.1109/ICCED51276.2020.9415767.

[5] A. Salma and W. Silfianti, “Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors,” *Int. Res. J. Adv. Eng. Sci.*, vol. 6, no. 4, pp. 158–162, 2021.

[6] A. Lia Hananto *et al.*, “Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm,” *J. Phys. Conf. Ser.*, vol. 1908, no. 1, 2021, doi:

10.1088/1742-6596/1908/1/012024.

[7] C. Steven and W. Wella, “The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter,” *IJNMT (International J. New Media Technol.*, vol. 7, no. 2, pp. 102–110, 2020, doi: 10.31937/ijnmt.v7i2.1732.

[8] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.

[9] E. Widawati, I. Iskandar, and C. Budiono, “Kajian Potensi Pengolahan Sampah (Studi Kasus : Kampung Banjarsari),” *J. Metris*, vol. 15, pp. 119–126, 2014.

[10] A. Lia Hananto, B. Priyatna, A. Fauzi, A. Yuniar Rahman, Y. Pangestika, and Tukino, “Analysis of the Best Employee Selection Decision Support System Using Analytical Hierarchy Process (AHP),” *J. Phys. Conf. Ser.*, vol. 1908, no. 1, 2021, doi: 10.1088/1742-6596/1908/1/012023.

[11] Y. S. Mahardika and E. Zuliarso, “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes,” *Pros. SINTAK 2018*, no. 2015, pp. 409–413, 2018.

[12] A. L. Hananto, B. Priyatna, and A. Y. Rahman, “Penerapan Algoritma Djikstra Pada Sistem Monitoring Petugas Lapangan Pemkab Bekasi Berbasis Android,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 4, no. 3, p. 95, 2019, doi: 10.31328/jointecs.v4i3.1078.

[13] F. Fathonah and A. Herliana, “Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes,” *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.

[14] A. Gormantara, “Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Menggunakan Metode Support Vector Machine,” *Konf. Nas. Ilmu Komput. 2020*, no. July, pp. 1–5, 2020.

[15] H. Nurrun Muchammad Shiddieqy, S. Paulus Insap, and W. Wing Wahyu, “Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. March, pp. 57–64, 2016.

[16] S. Chohan, A. Nugroho, A. M. B. Aji, and W. Gata, “Analisis Sentimen Pengguna Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique,” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 139–144, 2020, doi: 10.31294/p.v22i2.8251.

[17] A. Alwi, I. Iskandar, and D. Setyanto, “The Philosophy of Naive Bayes and its Comparison with the Tree Augmented Naive Bayes (TAN) in Making Predictions (Case Study Using Course Student Data),” *Saudi J. Eng. Technol.*, vol. 7, no. 7, pp. 377–385, 2022, doi: 10.36348/sjet.2022.v07i07.005.

[18] F. Sodik and I. Kharisudin, “Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter,” *Prisma*, vol. 4, pp. 628–634, 2021.

[19] G. Nugroho, D. T. Murdiansyah, and K. M. Lhaksmana, “Analisis Sentimen Pemilihan Presiden Amerika 2020 di Twitter Menggunakan Naïve Bayes dan Support Vector Machine,” vol. 8, no. 5, pp. 10106–10115, 2021.

[20] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.