

# Multinomial Naïve Bayes for sentiment analysis of Indonesian's local government performance



Ahmad Azhari <sup>a,1,\*</sup>, Muhammad Saepul Hadi <sup>a,2</sup>

<sup>a</sup>Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

<sup>1</sup>[ahmad.azhari@tif.uad.ac.id](mailto:ahmad.azhari@tif.uad.ac.id); <sup>2</sup>[mashadie10@gmail.com](mailto:mashadie10@gmail.com)

\* corresponding author

## ARTICLE INFO

## ABSTRACT

### Keywords

Sentiment Analysis  
Multinomial Naïve Bayes  
Twitter  
Local Government  
Performance

Digitalization of government performance, in conveying information and getting criticism, suggestions, and complaints from the public, is currently being carried out using social media. The use of social media is a form of government responsibility and openness to society. The high number of Twitter users in Indonesia, which reaches 6.43 million, allows the government to get many responses from the public. This background provides an opportunity for the public to be able to measure government performance based on a number of criticisms, suggestions, and complaints that the government responds to. However, public sentiment towards government performance has not been used as an evaluation and benchmark for the government in determining policies. The purpose of this research is to build a social media twitter sentiment analysis system to measure public sentiment towards local government performance by implementing Multinomial Naïve Bayes. This research is divided into several stages including tweet grabbing, manual tweet filtering, tweet labeling, split tweets, preprocessing tweets, term frequency, classification, and evaluation. The tweet retrieval process was carried out on 1 June - 31 July 2020 with 2000 tweets used from the total tweets obtained after manual filtering was carried out. This study shows that the sentiment analysis carried out obtained an accuracy of 80%, a precision of 78%, and a recall of 82%.

This is an open access article under the [CC-BY-SA](#) license.



## 1. Introduction

The We Are Social survey conducted in 2019 of 6.43 million twitter users in Indonesia shows that Indonesia is in 10th place in the world with the highest number of Twitter users [1]. The high number of Twitter users in Indonesia certainly has an influence, where a lot of information in the form of complaints, criticisms, and public reports via the Twitter application is displayed to the central and local governments. Almost every public institution (central government, ministry, local government) has a social media account that serves as a tool for disseminating information to the public or getting responses from the public.

As an organization that runs a government system, the government has several tasks including disseminating information, communicating policies, work plans, and performance achievements to the wider community, finance, policies in all aspects, through conventional and modern media [2].

Regional heads also have social media accounts as a means of disseminating information with a greater number of followers and interactions than the official regional government accounts themselves. The official Twitter account of the West Java Provincial Government @jabarprov.go.id

has 19,987 followers while the regional head @ridwankamil has 4 million followers, the DKI Jakarta Provincial Government's account is 1.2 million followers while the regional head @aniesbaswedan has 3.7 million followers, with followers who have interaction with high communication intensity.

This study aims to analyze the social media accounts of regional heads with Multinomial Naive Bayes. Sentiment analysis on the twitter account of regional heads is needed, because the response from the community is a demand for the unfulfilled services provided by the government. This research also aims to support the government to create quality services for the community, with this sentiment analysis the government can measure public sentiment through the Twitter platform whether it is positive or negative, so that the government can classify public sentiment to evaluate policies or determine further policies.

## 2. Literature Review

### 2.1. Previous Research

Research conducted by Ginting, Lhaksmana, & Murdiansyah in 2018 with the title "Classification of Sentiment Towards Candidates for West Java Governor 2018 on Twitter Using Naive Bayes". Tweet public opinion obtained from Twitter is then processed with classification techniques, while for the classification of sentiment is divided into 3 indicators, namely positive, neutral and negative sentiments, with indicators based on public tweets related to candidates. After getting a classifier system, it is followed by a preference value to find out the trending or positive response from the community towards the prospective candidates for governor of West Java 2018 [3].

Another research conducted by Berliana, Shaufiah, & Sa'adah in 2018 was entitled "Classification of Tweet Postings on Government Policy Using the Naive Bayesian Classification". In this study, tweets that state public opinion regarding tax amnesty are classified by the Naive Bayes Classifier method, the tweet data is then carried out preprocessing. In this research, the system is able to classify data into three classes, namely positive, negative, and neutral with extraction using unigram and term frequency with an accuracy of 53.45% with learning data of 80% from 578 Twitter data [4].

Previous research conducted by Lorosae, Prakoso, Saifudin, & Kusri in 2018 entitled "Sentiment Analysis Based on Public Opinion on Twitter Using Naive Bayes". The research was conducted in several stages, including the identification of input data, pre-processing, classification of the Naive Bayes method, and output. Input data is in the form of training data and test data in the form of public opinion about services from the TIKI shipping company Twitter account. As for the results of analysis and testing using 100 test data, the Naive Bayes Classifier method with pre-processing produces an average accuracy of 84%, and also results in a classification of positive, negative sentiments [5].

The research that was carried out by Nurulbaiti & Subekti in 2018 was entitled "Sentiment Analysis on Tweet Data for Social Security Administering Bodies (BPJS) Using the R Program". In this study, the aim of this research is to find out the topics that are often discussed by the Indonesian people related to the BPJS with sentiment analysis. The test data used is 1000 tweets with limited only tweets that use Indonesian. The R program is used to assist the process of data retrieval, data processing, and information visualization, test data is then processed using text mining and sentiment scoring. The results obtained from this research are public opinion tweets on the social media Twitter network regarding BPJS, it is found that the dominant public response is negative [6].

Previous research conducted by Soepriadi & Permata in 2015 entitled "Sentiment Analysis to Assess Community Satisfaction with Local Government Performance Using the Naive Bayes Classifier (Case Study: Mayor of Bandung 2013 - 2018)." In this study, the researchers classified tweets from the Bandung education office's twitter account with the account name @disdik\_bandung, the transportation department with the account name @dishub\_kotabdg, the environmental service with the account name @dbmpkotabdg. All data from the tweet account will be classified into six classes including; positive education, positive transportation, positive environment, negative education, negative transportation, and negative environment [7].

### 2.2. Sentiment Analysis

- 1) Sentiment Analysis is the process of understanding, extracting, and processing textual data automatically to obtain sentiment information contained in an opinion sentence. Sentiment analysis is carried out to see an opinion or tendency of opinions on a problem or object by

someone, whether they tend to have a positive or negative view or opinion [8]. Based on the data source level, sentiment analysis is divided into 2 major groups, namely: Coarse-grained Sentiment Analysis and Fined-grained Sentiment Analysis. In Coarse-grained Sentiment Analysis, the sentiment analysis carried out is at the document level. Broadly speaking, the main focus of this type of sentiment analysis is to regard the entire document as a positive sentiment or negative sentiment. Fined-grained Sentiment Analysis is a sentiment analysis at the sentence level. The main focus of the fined-grained Sentiment Analysis is to determine the sentiment in each sentence [9].

- 2) Twitter is a service for friends, family, and coworkers to communicate and stay connected through the exchange of quick, frequent messages. People post Tweets, which may contain photos, videos, links, and text. These messages are posted to your profile, sent to your followers, and are searchable on Twitter search [10]. Twitter itself is a platform with 280 characters of text in one time sending a status or what is often called a tweet.
- 3) Twitter API On its official website, Twitter explains that the Twitter Application Programming Interface or what is commonly called the Twitter API is a way for computer programs to "talk" to each other so that they can request and present information. It does this by allowing a software application to call what is known as an endpoint, the address associated with a certain type of information provided [10].
- 4) Text Mining can be broadly defined as a knowledge intensive process where users interact with a collection of documents over time using a series of analytical tools. In a manner similar to data mining, text mining seeks to extract useful information from data sources through identification and exploration of interesting patterns [11].
- 5) Term Frequency and Inverse Document Frequency (TF-IDF) are numerical statistics that show the importance of words in a document. TF-IDF is used as a factor to calculate the weight of information retrieval [12]. In this research, what is done is Term Frequency, this is done to get the number of each word that appears in one tweet, the term frequency will count every word that appears in one tweet which will later be used as one of the processes in the Naïve Bayes algorithm.
- 6) Naïve Bayes Algorithm is one of the algorithms contained in the classification technique. Naïve Bayes is a classification using probability and statistical methods proposed by the British scientist Thomas Bayes, which predicts future opportunities based on previous experience, so it is known as the Bayes Theorem. The theorem is combined with Naïve where it is assumed that the conditions between the attributes are mutually independent [13].
- 7) Multinomial Naïve Bayes, One of the Naïve Bayes models that is often used in text classification is the Multinomial Naïve Bayes, this model utilizes the probability theorem, namely the Bayes Theorem and data mining functionality, namely the Naïve Bayes classification. Multinomial Naïve Bayes is one of the Bayes methods used by calculating the frequency of each word occurrence in a document and the probability. The advantages of Naïve Bayes Multinomials are a high level of accuracy, easy to implement, low computation time and minimum error rate [14]. In this research, the method used is Naïve Bayes with the Multinomial Naïve Bayes model, in the Multinomial Naïve Bayes method the probability of a document  $d$  is in class  $c$ , this condition can be expressed by equation 1:

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (1)$$

In the above equation  $P(t_k|c)$  is the conditional probability of  $t_k$  that is in the document that belongs to class  $c$ .  $P(c)$  is the prior probability of the documents contained in class  $c$ .  $t_1, t_2, t_3 \dots, t_n$  are tokens in document  $d$  which are part of the vocabulary used for classification and the number of tokens from document  $d$ . To estimate the prior probability  $P(c)$  expressed by equation 2:

$$P(c) = \frac{N_c}{N} \quad (2)$$

In equation 2,  $N_c$  is the number of training data documents in class  $c$ ,  $N$  is the total number of training data documents from all classes. Furthermore, to estimate the conditional probability  $P(t_k|c)$  expressed by equation 3:

$$P(t_k|c) = \frac{\text{count}(t_n|c) + 1}{\text{count}(c) + |V|} \quad (3)$$

In equation 3,  $\text{count}(t_n|c)$  is the number of words found throughout the training data with category  $c$ .  $\text{count}(c)$  the number of words in the training data with categories  $c$ ,  $|V|$  is the total number of words in the training data.

- 8) Confusion Matrix is a matrix that contains the actual and predictive classification carried out by the classification system. The performance of a system is evaluated using a matrix [15]. The Confusion Matrix is used to test the level of accuracy by using three calculated parameters, including accuracy, recall, and precision. The confusion matrix table is shown in the Table. 1. Confusion Matrix

**Table 1.** Confusion Matrix

		Prediction Result Class	
		TRUE	FALSE
Actual Class	TRUE	TP (True Positive)	FP (False Positive)
	FALSE	FN (False Negative)	TN (True Negative)

In table 1 of the Confusion Matrix, True Positive (TP) is the amount of data with the actual class being a positive class and the prediction class being a positive class. False Positive (FP) is the amount of data with the actual class being negative class with the prediction class being the positive class. True Negative (TN) is the amount of data where the actual class is a negative class and the prediction class is a negative class. False Negative (FN) is the amount of data where the actual class is a positive class and the prediction class is a negative class. Testing with a Confusion matrix based on three parameters including:

- Accuracy is the level of recognition which states that much of the data in the test data is correctly classified by the classification model. To determine the accuracy value of equation 4:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = * 100\% \quad (4)$$

- Precision is the level of accuracy between the information requested and the answers given by the system. To determine the value of precision as in equation 5:

$$\text{Precision} = \frac{TP}{TP + FP} = * 100\% \quad (5)$$

- Recall is the success rate of the system in recovering information. To determine the recall formula as in equation 6:

$$\text{Recall} = \frac{TP}{TP + FN} = * 100\% \quad (6)$$

### 3. Method

#### 3.1. Object of Research

The object of this research is to build a social media Twitter sentiment analysis system on the performance of local government, in this case the governor of West Java for the period 2018-2023 with the twitter account name @Ridwankamil while the data taken are all people who mention the word ridwankamil, while the documents studied are tweets using Indonesian. This system is expected to help the government as a yardstick in determining policies or evaluating government performance in every process of service to the community.

#### 3.2. Data Collection

The data collection method used is literature study, where the researcher conducts a review of previous research both journals, books and other scientific publications. Then the tweet grabbing process is to get data from Twitter based on predetermined keywords. The time used for grabbing tweets is from 1 June 2020 to 31 July 2020 and the data used is data taken from Twitter which mentions @ridwankamil and tweet reply from the @ridwankamil account.

#### 3.3. Research Method

The stages of research carried out in this study are shown in Fig. 1. Flowchart of Research Stages.

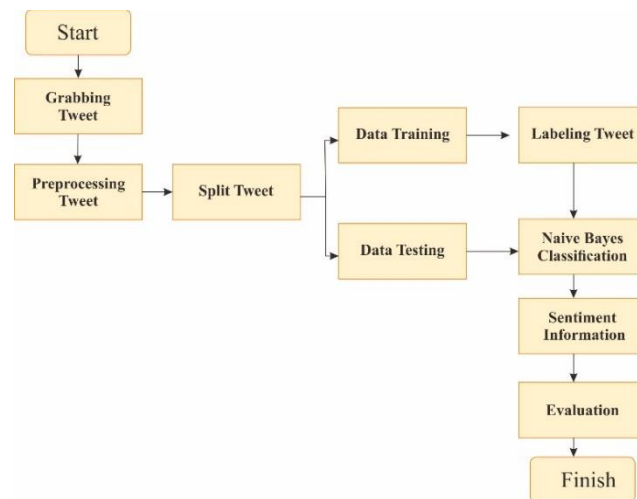


Fig. 1. Flowchart of Research Stages

- **Grabbing Tweet**, this stage, it is carried out to get tweets based on predetermined keywords, namely @ridwankamil, the grabbing process takes advantage of the API provided by Twitter. The results from grabbing are then automatically saved into a file with csv format. Due to the many multiple tweets and the large number of spam tweets, the author performs manual filtering of the grabbing CSV files to make the system more accurate in determining sentiment.
- **Preprocessing Tweet**, after getting the results of the grabbing tweet, the next process is preprocessing the tweet with the tokenizing stage used to cut the sentence input (string) based on each constituent word, removing hashtags at and URL, removing punctuation, case folding is used to change all tweets to lowercase and only accept A - Z received, Filtering is used to retrieve important words in each tweet, using a predetermined stop word or stop list, Stemming is used to change the word into the original word.
- **Split Tweet**, In the split data process, the data is divided into two parts, namely training data and test data, with a percentage of 70% training data and 30% test data. The training data are then labeled for the classification process.

- **Labeling**, after splitting data, the training data is then labeled, positive label is given a value of 2 and negative label is given a value of 1, the melting is done with a balance between positive and negative to get the maximum accuracy value.
- **Multinomial Naïve Bayes Classification**, Basically, Naïve Bayes Multinomial Algorithm uses two stages in the process, including training and testing. At the training stage, training is carried out on each tweet whose category is already known, while in the testing, the classification of the tweet that is not known is the category. In this process, before the classification process with the Multinomial Naïve Bayes, first looking for the probability of the frequency of occurrence of words in the training data and giving categories or labels to the training data, this process will later become part of the Multinomial Naïve Bayes classification process. Where the training data that has known the frequency of occurrence of words and their categories are carried out with a probability test data with unknown frequency of occurrence of words and their categories.

### 3.4. System Implementation

System implementation is the implementation stage of the design results into the system to be created, in this case the Python programming language, Flask Framework, HTML and SQL are used. The results of this implementation are in the form of a Twitter social media sentiment analysis system on local government performance using the naïve Bayes method and in the form of a website.

### 3.5. Evaluation

The evaluation process is needed to determine whether the system built can be used as it should, in this study testing was carried out to determine the accuracy value provided by Multinomial Naïve Bayes in analyzing every public sentiment in both the positive and negative categories. At this testing stage, it was carried out using the Confusion Matrix with three calculated parameters, including accuracy, recall, and precision.

## 4. Results and Discussion

In this research, the data used is tweet data which mentions the keyword of military personnel, the data is obtained by grabbing from Twitter by utilizing the Twitter API. The data from the grabbing tweets are then stored in a CSV file format and manually filtered to eliminate duplicate tweets and tweet spam. From the filtering results collected a dataset of 2000 tweets taken from June 1 to July 31, 2020 with the attributes index, screen, name, created at, text. Samples of grabbing tweets can be seen in the Table 2. Sample Grabbing Tweet Results.

**Table 2.** Sample Grabbing Tweet Results

Screen name	Created at	Text
KSecoria	2020-07-01 12:03:36	""@ridwankamil Untung saja gubernur Jabar sekarang kang emil, kalo dilanda pandemi begini bisa dipercaya kinerjanya. \xe2\x80\xa6 https://t.co/96jfCTWj8M""
realita1987	2020-07-01 11:58:11	'@tianfillio @gojekindonesia @GrabID @ridwankamil Sabar ya bang \xf0\x9f\x98\x81\xf0\x9f\x98\x81'
tpserasi	2020-06-18 14:36:03	'@ridwankamil @nilacobek semoga bermanfaat pak gubernur'

### 4.1. Tweet Grabbing Stages

At this stage, it is carried out to get tweets based on predetermined keywords, namely @ridwankamil, the grabbing process takes advantage of the API provided by Twitter. The process of grabbing tweets is carried out using the Tweepy library, the results of the grabbing are then automatically saved into a file with csv format. The number of multiple tweets and spam tweets is then done manually filtering the csv file to eliminate duplicate tweets and spam so that the system can be more accurate in determining sentiment.

### 4.2. Preprocessing Tweet

After getting the results of grabbing the tweet then uploading it to the system for the preprocessing process of the tweet with the Tokenizing stage for cutting the tweet based on each word of its



composition, then removing hashtags and urls in the tweet, then Case Folding to change every tweet to lowercase and only accept A-Z and other than then it is not accepted, next is Filtering where in this process the important words are taken in each tweet based on the specified Stopword and Stoplist, and the last one is Stemming which is used to change the tweet into the original word. The examples of the results of preprocessing tweets can be seen in Table 3. Examples of Tweet Preprocessing.

**Table 3.** Example of Tweet Preprocessing

Tweet	Preprocessing Tweet
@ridwankamil Bikin aturan tegas dong pak gubernur..!! Jgn cuma himbauan aja..percuma tidak ada manfaatnya.karena kita\xe2\x80\xa6 himbauan percuma tidak manfaat https://t.co/9Yy51q4Z5A	bikin atur tegas dong gubernur cuma himbauan percuma tidak manfaat

#### 4.3. Multinomial Naïve Bayes Clasification

In this research, the process of labeling positive and negative on the training data was still manual, where each tweet was labeled textually. The process of applying the Naïve Bayes Multinomial method in this research is to take a sample of tweets from Twitter which uses the Twitter API for grabbing tweets with the keywords ridwankamil, while 3 tweets are taken randomly from the results of grabbing tweets then used as training data, and 1 tweet. for test data. For training data samples and test data taken can be seen in the Table. 4. Sample Training Data and Test Data.

**Table 4.** Sample Training Data and Test Data

Document	Tweet	Tweet Preprocessing	Label
<b>Training Data</b>	1 @ridwankamil Alhamdulillah,...\nDan semoga tetap disiplin	alhamdulillah semoga disiplin	Positive
	2 @ridwankamil @nilacobek mantap, semoga bermanfaat pak gubernur	mantap moga manfaat gubernur	Positive
	3 anjing emang lu @ridwankamil kerjanya tidak jelas	anjing kerja tidak jelas	Negative
<b>Testing Data</b>	4 Mantap pak @ridwankamil kerjanya semoga disiplin	mantap kerja semoga disiplin	?

After obtaining the preprocessing training data and labeling it, the next step is to predict the label on the test data by calculating the prior probability of the positive and negative classes.

$$P(Positif) = \frac{2}{3} = 0,6 \quad P(Negatif) = \frac{1}{3} = 0,3$$

After getting the prior probability value for each class, the next step is to calculate the conditional probability.

$$P(t_k|c) = \frac{\text{count}(t_n|c) + 1}{\text{count}(c) + |V|}$$

$$p(\text{mantap} | \text{Positif}) = \frac{1 + 1}{7 + 15} = \frac{2}{22} = 0,09 \quad p(\text{mantap} | \text{Negatif}) = \frac{0 + 1}{4 + 14} = \frac{1}{18} = 0,05$$

$$p(\text{kerja} | \text{Positif}) = \frac{0 + 1}{7 + 15} = \frac{1}{22} = 0,04 \quad p(\text{kerja} | \text{Negatif}) = \frac{1 + 1}{4 + 14} = \frac{2}{18} = 0,11$$

$$p(\text{semoga} | \text{Positif}) = \frac{1 + 1}{7 + 15} = \frac{1}{22} = 0,09 \quad p(\text{semoga} | \text{Negatif}) = \frac{0 + 1}{4 + 14} = \frac{1}{18} = 0,05$$

$$p(\text{disiplin} | \text{Positif}) = \frac{1 + 1}{7 + 15} = \frac{2}{22} = 0,09 \quad p(\text{disiplin} | \text{Negatif}) = \frac{0 + 1}{4 + 14} = \frac{1}{18} = 0,05$$

After getting the conditional probability value, the next process is a calculation to determine whether the test data tweet in document 4 is included in the positive or negative label. The calculations are as follows:

$$p(\text{Positif} | \text{Dokument 4}) = 0,6 \times 0,09 \times 0,04 \times 0,09 \times 0,09 = 0,0000017496$$

$$p(\text{Negatif} | \text{Dokument 4}) = 0,3 \times 0,05 \times 0,11 \times 0,05 \times 0,05 = 0,000004125$$

After getting the probability value from the test data against the training data, then the probability results are compared with the positive and negative labels. In document 4 with a positive label, a probability value is obtained of **0,0000017496**, while in document 4 with a negative label, a probability value is obtained of **0,000004125**, thus the test data in document 4 labeled positive is greater than document 4 labeled negative, so document 4 is included into the **"Positive"** label.

#### 4.4. Interface System

Implementation of a system created using the Micro Web Frame Work Flask with the Python programming language for uploading tweet data, training data dictionaries, and sentiment analysis results. The implementation is shown in Fig. 2. Results Sentiment Analysis Page.

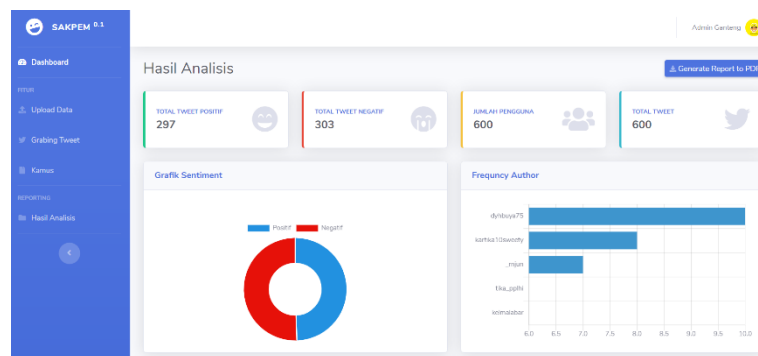


Fig. 2. Results Sentiment Analysis Page.

#### 4.5. Evaluation

At this testing stage, it was carried out using the Sklearn Confusion Matrix Library with three calculated parameters including Accuracy, Recall and Precision. Based on 2000 data with 1400 training data and 600 test data with a balance of 1000 data labeled positive and 1000 data labeled negative. In the testing process, the values are obtained as in Table 5. Testing the Confusion Matrix:

Table 5. Testing the Confusion Matrix

Categories	Total
True Positive	247
False Positive	70
True Negative	230
False Negative	53

The following is a test with a Confusion Matrix with three calculated parameters including Accuracy, Recall, and Precision on the positive label:

$$Accuracy = \frac{247 + 230}{247 + 230 + 70 + 53} = \frac{477}{600} = 0,80$$

$$Recall = \frac{247}{247 + 53} = \frac{247}{300} = 0,82$$

$$Precision = \frac{247}{247 + 70} = \frac{247}{317} = 0,78$$



The following is a test with a Confusion Matrix with three calculated parameters, including Accuracy, Recall, and Precision on the negative label:

$$Accuracy = \frac{247 + 230}{247 + 230 + 70 + 53} = \frac{477}{600} = 0,80$$

$$Recall = \frac{230}{230 + 70} = \frac{230}{300} = 0,77$$

$$Precision = \frac{230}{230 + 53} = \frac{230}{283} = 0,81$$

## 5. Conclusion

Based on the results of the sentiment analysis by applying the Multinomial Naïve Bayes which was carried out on 2000 tweets, it resulted in an accuracy value of 80%, 78% positive precision, 82% positive recall, 80% positive f1-score. While the accuracy value for negative tweets is 80%, negative precision is 81%, Recall is 77% and f1-score is 79%. In completing this research, further researchers are advised to do automatic linking of positive and negative tweets, filter out some tweets and spam from the results of grabbing tweets to be done automatically. The more training data the more accurate the results of the sentiment analysis accuracy classification are.

## References

- [1] A. Aswari, "How digital technology driven millennial consumer behaviour in Indonesia," *Journal of Distribution Science*, vol. 17, no. 8, pp. 25-34, 2019.
- [2] M. N. Aziz, A. Firmanto, A. M. Fajrin and R. V. Hari Ginardi, "Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction," *2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, pp. 125-130, 2018.
- [3] P. P. Surya and B. Subbulakshmi, "Sentimental Analysis using Naive Bayes Classifier," *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, pp. 1-5, 2019.
- [4] A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pp. 257-261, 2016.
- [5] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," *Big Data and Cognitive Computing*, vol. 4, no. 1, p. 1, 2020.
- [6] I. Kurniawati and H. F. Pardede, "Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis," *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 1-5, 2018.
- [7] M. M. Saritas, and A. Yasar, "Performance analysis of ANN and Naive Bayes classification algorithm for data classification," *International journal of intelligent systems and applications in engineering*, vol. 7, no. 2, pp. 88-91, 2019.
- [8] K. Sarkar and M. Bhowmick, "Sentiment polarity detection in bengali tweets using multinomial Naïve Bayes and support vector machines," *2017 IEEE Calcutta Conference (CALCON)*, 2017, pp. 31-36, 2017.
- [9] D. Wang and R. Alfred, "A Review on Sentiment Analysis Model for Chinese Weibo Text," *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pp. 456-463, 2020.
- [10] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin and L. Marlinda, "Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter," *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1-6, 2018.

- [11] E. Y. Sari, A. D. Wierfi and A. Setyanto, "Sentiment Analysis of Customer Satisfaction on Transportation Network Company Using Naive Bayes Classifier," *2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, pp. 1-6, 2019.
- [12] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler: Twitter," *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pp. 1-5, 2019.
- [13] S. Das and A. K. Kolya, "Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm," *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pp. 239-244, 2017.
- [14] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 416-419, 2016.
- [15] S. P. Kristanto, J. A. Prasetyo and E. Pramana, "Naive Bayes Classifier on Twitter Sentiment Analysis BPJS of HEALTH," *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)*, pp. 24-28, 2019.