

# Twitter Sentiment Analysis in Indonesian Language using Naive Bayes Classification Method

Cakra Adipura Wicaksana  
Department of Informatics  
University of Sultan Ageng Tirtayasa  
Cilegon, Indonesia  
cakraadipura@untirta.ac.id

Mohammad Fatkhurrohman  
Department of Electrical Engineering  
Vocational Education  
University of Sultan Ageng Tirtayasa  
Serang, Indonesia  
fatkhur0404@untirta.ac.id

Hafiyyan Putra Pratama  
Department of Telecommunication System  
Indonesia University of Education  
Purwakarta, Indonesia  
hafiyyan@upi.edu

Rifaldi Tryawan  
Department of Electrical Engineering  
University of Sultan Ageng Tirtayasa  
Cilegon, Indonesia  
3332150004@untirta.ac.id

Alimuddin  
Department of Electrical Engineering  
University of Sultan Ageng Tirtayasa  
Cilegon, Indonesia  
alimuddin@untirta.ac.id

Rinanda Febriani  
Department of Informatics  
University of Muhammadiyah Bandung  
Bandung, Indonesia  
rinanda@umbandung.ac.id

**Abstract**— Social media is a very large source of data and has continuous and very short changes, the time span of which changes also has considerable relevance for various fields. The purpose of this study as a sentiment analysis aimed at twitter status discussing daily topics to determine the attitude of opinion from tweets posted on a trending topic online using the Naive Bayes Classifier method. With various series of steps carried out, namely data collection from Twitter, preprocessing, classification process with the Naives Bayes Classifier algorithm, and evaluation of the results determined by this method. The results at the time of testing, namely in the form of the highest accuracy value obtained in testing 40 tweet data by 82% with a time of 19.92 seconds, for the highest precision value in testing 10, 20, and 50 tweet data by 100% with a time of 14, 85 seconds, for the highest recall value in testing 40 tweet data of 81.82% with a time of 16.09 seconds, for the highest F1-Score value in testing 40 tweet data of 88.52% with a time of 16.38 seconds.

**Keywords**—twitter, bayes, sentiment, python, naïve bayes

## I. INTRODUCTION

Data is one thing that everyone will produce every day. It always increases as time goes on, named Big Data. Big data is a term that has a meaning of a bunch of data. The far-range of folks will produce data every day, for example it can be from the daily history of their application on smartphones, financial transaction, social media, online shopping, Text Messaging and so on. All their history of that activity will produce a bunch of data. It will be a shame if it can't be used for analytics purposes. Remembering, that data from Simon [1] said about eighteen million prospective twitter users.

The Purpose of The Data Analytics can be meaningful information that can be useful for another stakeholder's sake. For Example, what the topic that all the people talk about for a day, a week, a month, even a year or what people usually need to buy during the range of time. Thus, data is the fuel for this process.

Social media is a very large source of data and has continuous change and the time span of the change is very short and has considerable relevance for various fields. At one time there are millions of changes that occur in social media

so that the data that can be used in social media is very massive. There are so many things that become issues in social media from various fields, from economics, politics, religion to sports, and the data comes from various existing social networks and provides access to each system to get the data needed from some account information opened by each user (*permitted to be shared*) to timeline information regarding posts from each account.

The potential for the application of big data and analysis to the processed data is very large and valuable. So that the results of the analysis carried out from social media are very useful and are needed to improve the quality of decision making. Data or information plays a central role in every study and scientific discovery as well as the sophistication of social media today, greatly supporting the collaboration of professional researchers and the public to research and make new discoveries.

Most social media provide opportunities for users to interact with content, provide comments, vote, share content, or communicate between users. The process of sending messages through social media is faster than other media, thus making information interconnected, easily accessible and quickly spread widely. Social media users can connect with other users through the link facility at the address of an information obtained from the internet.

Sentiment Analysis refers to the use of Natural Language Processing to determine the attitudes, opinions and emotions of a speaker, writer, or other subject. The Naive Bayes classifier exhibits high accuracy and speed when applied to large databases. The Naive Bayes method is often used in solving problems in the field of machine learning because this method is known to have a high level of accuracy with simple calculations. Bayes classifier is a statistical classifier, where this classifier can predict the probability of class membership of a data that will fall into a certain class, according to the probability calculation. The Bayesian classifier is based on the Bayes theorem discovered by Thomas Bayes in the 18th century. In a comparative study of classification algorithms, a simple Bayesian or commonly known as Naive Bayes classifier has been found.

## II. PREVIOUS WORKS

There are some research or study that have been there before which talk about twitter analysis. For example, as elaborated in journals and papers in [2], [3], and [4]. Rani at. All [2] also used a naïve bayes as their method to analyze data from twitter, but she combine the method with SVM, so I think in my opinion, this such of combination will lead to more of time consumption. Simanjuntak at. All [3] also propose the similar way in conducting twitter analysis. He use of crawling method to gather data tweet from twitter, but he only used a well-package library from python programming language, namely Varder Sentiment Package. Ultimately, Narasamma at. All [4] have concise all methods or approaches by elaborated merits and demerits, respectively. In his paper, especially in aspects of merits, naïve bayes still have better information recall. In Another hand, naïve bayes aspect still has drawback at the time consumption and the time in processing.

## III. METHODOLOGY

This study carried out several stages, namely Twitter social media data collection, data processing, system implementation, text classification with the Naive Bayes Classifier method and evaluation of the system to classify sentiment analysis data which will be concluded with the resulting data. The stages that will be carried out in this research are depicted as bellow in figure 1 and figure 2.

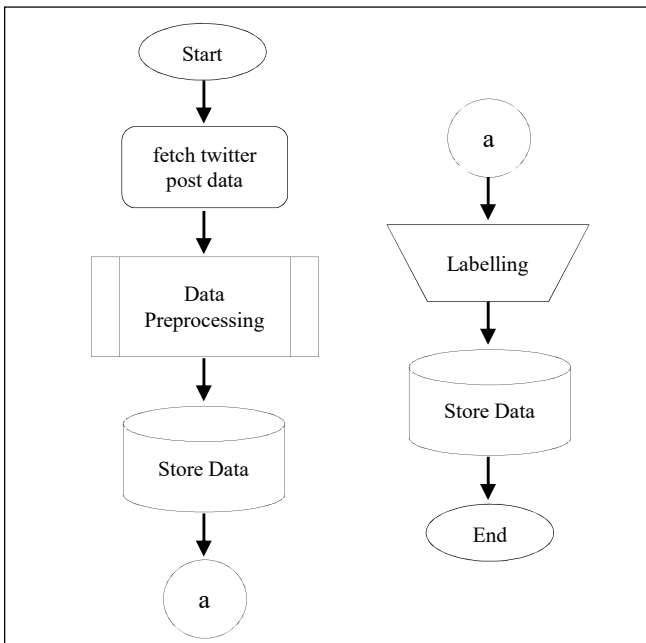


Fig. 1. Data Collection Process

There are many methodologies in collecting data from the twitter application, for example can be retrieved from API or crawl method. In this papers, crawl method will be used to gather tweet data. Here is the basic explanation of crwaling method [5][6][7]. A web crawler that is also called a web spider is a program that browses the web in a methodical manner to gather information.

TABLE I. COMPARISON OF WORDS IN CLASSIFICATION

Target	Example of Words (Indonesian Language)		
	Word 1	Word 1	Word 1
Positive	Saya	Suka	Kamu
Positive	Kamu	Sangat	Cantik
Positive	Dan	Juga	Pintar
Negative	Tapi	Saya	Jelek
Negative	dan	juga	misikin

Here is an example of naïve bayes aspect. Given a sentence in Indonesian Language “Saya Cantik dan Pintar”.

- Calculate target odds  
Positive odds =  $3/5 = 0.6$   
Negative odds =  $2/5 = 0.4$
- Calculate the probability of each word against each target

Here is a positive target:

“saya” odds | positive =  $1/3 = 0.33$   
 “Cantik” odds | positive =  $1/3 = 0.33$   
 “dan” odds | positive =  $1/3 = 0.33$   
 “Pintar” odds | positive =  $1/3 = 0.33$

Here is a negative target:

“saya” odds | negative =  $1/2 = 0.5$   
 “Cantik” odds | negative =  $0/2 = 0$   
 “dan” odds | negative =  $1/2 = 0.5$   
 “Pintar” opportunity | negative =  $0/2 = 0$

Positive P =  $0.6 + 0.33 + 0.33 + 0.33 + 0.33 = 1.92$   
 Negative P =  $0.4 + 0.5 + 0 + 0.5 + 0 = 1.4$   
 Result = Positive

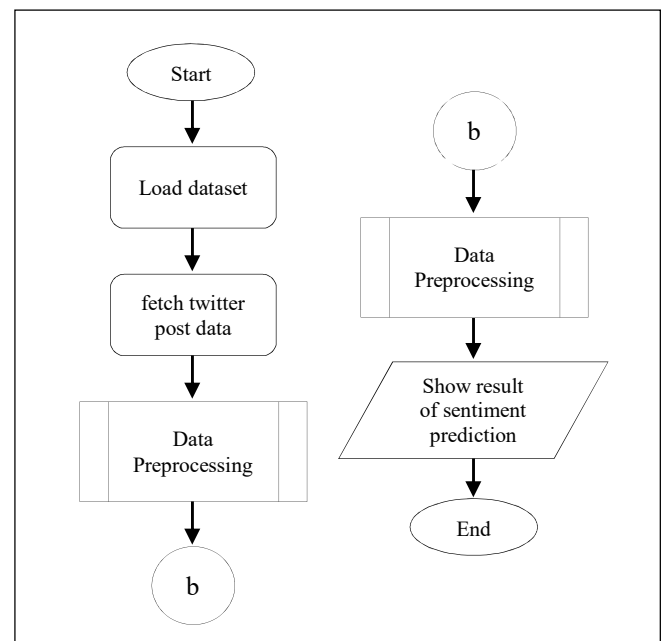


Fig. 2. Data Testing Process

The stages that will be carried out in system testing can be described as follows.

#### A. Dataset Retrieval

The dataset retrieval process is the initial process that must be done. In this process we take random twitter posts and manually label each one with the label "POSITIVE" or "NEGATIVE". And save it in one file. This data will be used as a reference during the test process to detect sentiment. Data retrieval is done by taking post tweets using the twitter API.

#### B. Text Preprocessing

In the preprocessing process, several things were done to process the twitter post data so that it could be used in the next process. Preprocessing is done, among others, is to remove unnecessary characters to become clean text. Clean text in question is text that only contains alphabets and numbers without punctuation, mathematical characters, emoticons etc. Then the clean text is converted into lowercase format or non-capital letters. This is done to equalize and avoid mistakes during the next process.

#### C. Sentiment Detection

The sentiment detection process consists of two processes, namely the dataset retrieval process and data processing. In this data retrieval process, the stored dataset is a dataset that has passed normalization, meaning that the filtered dataset only displays text that is processed in the Navies Bayes Classifier processing. For the dataset testing process (training data and testing data) this process compares tweets into words to distinguish the classification results. In this case, the data has gone through labeling before being classified in the program.

#### D. Classification with Naive Bayes Classifier

In the Naive Bayes Classifier method, the text classification process is carried out based on data that has previously been taken on the large number of tweets that have been downloaded from Twitter, in accordance with the purpose of this search, namely to get the most votes from the analytical sentiment generated by the classifier, then generate a total positive or negative percentage. , which is a score from tweet data in the form of posts or topics that are currently trending.

#### E. System Testing

The testing process is the main process, in this process to take posts from twitter which are processed, and the estimated sentiment is determined using the Naive Bayes method based on the data obtained from the previous process. This process determines the success rate of running or not the system programmed in the analysis shown next. The dataset has been classified from full sentence tweets and word classification by displaying the results of calculating positive or negative opportunities based on the calculation of words in Twitter posts. After going through the calculation results, the data is displayed by calculating the value of positive and negative words and their classification.

## IV. RESULT AND ANALYSIS

This training data used in the current test is in the form of tweet data and slang words data which is then stored into a dataset for classification data as training data and test data. The stored training data is in the form of sentence parameters that are posted to Twitter and then used as training data parameters, in addition to tweet sentence parameters and slang words parameters can be detected properly. The next process is sentiment analysis using the opportunity formula, then each sentence is assessed verbally as the data being tested, then for non-standard words it is matched first into the standard words that have been made so that the results displayed can adjust the existing sentences and can be understood by the programming language.

#### A. Twitter Dataset

This dataset is used for sentence parameters that will be run in the program, with this sentence parameter that will judge how accurate the data is, but the Naive Bayes method has strong independence so there are several sentences that according to human language are not appropriate, because in this program method the parameters The assessment is based on the calculation of the opportunities that arise and requires a separate sentence or dictionary adjustment to produce the appropriate data. To retrieve tweets and sentiment data in the database, pseudocode is used. The more numbers in the dataset, the better the assessment in the classification, because this data is compared to determine the assessment of a sentence. The data taken is 10,000 tweets that will be stored as a classification parameter. Tweet data is taken randomly either from ordinary users or online media on Twitter. This data is the parameter in the search for tweets, with this data as a parameter, a lot of data is needed so that the data classification is better at identifying.

```
initialize database_connection

get database from database_connection
get tweet, sentiment from database

get positive_tweet, negative_tweet from tweet
get positive_words from positive_tweet
get negative_words from negative_tweet
```

Fig. 3. Pseudocode database

#### B. Slang words

In this study, the slang words dataset process was used to identify non-standard words. This problem can make the program unable to identify words in tweet sentiment analysis. To be able to identify words in the program, you must create a slang words dataset. In the program used for this research, pseudocode is used to run the program. To enter slang words data into this program, it is very necessary to identify non-standard words. Slang words have become a language that is often used to communicate and is commonly used in everyday life. This language that is not included in any major dictionary is indeed difficult to understand and requires to enter the meaning of every word that comes out of the other person while communicating. Slang words are a collection of words that are often used in writing tweets on Twitter or on other social media, this is useful for facilitating understanding in program languages

to recognize undetected words or words that are not based on standard words.

```
get slang_words from database
```

Fig. 4. Pseudocode Slang Words

### C. Sentiment Classification

This stage is as data preprocessing, the steps taken to process the initial data to become ideal data for the next process. To classify sentiments, pseudocode is used. In this study, the selection of positive sentiment tweets and negative sentiments was used. Positive tweets must go through the negative tweet selection stage and vice versa, this is because there are words that contain positive meanings but are used in negative tweet sentences, in this case it is necessary to take sample data in each classification to calculate the number of opportunities that most appear on sentiment. In the Naive Bayes method, the more word opportunities that appear, the smaller the assessment because it uses probability data.

```
get positive_tweet, negative_tweet from tweet
get positive_words from positive_tweet
get negative_words from negative_tweet
```

Fig. 5. Sentiment Classification Pseudocode

### D. Displaying Text Processing words

This process is used to correct words that are not used in general or words that are not understood such as abbreviations, capital letters, punctuation marks, symbols, etc. The displayed letters, which are lowercase, adjust in data collection. The program for word processing uses pseudocode in the program flow. The data stored in the form of a collection of tweets that have been given a positive and negative classification label. The testing process uses samples from the tweet and slang words dataset, the data stored in the database as a calculation parameter in this method, for some of the data taken in the form of a probability calculation which will be broken down again in the tweet, the value obtained in the classification results is displayed in the program.

```
get tweet, sentiment = data_uji
get tweet_normal = normalisasi(data_uji)
get tweet_standar = standarisasi(tweet_normal)
get tweet_words = tokenisasi(tweet_standar)
```

Fig. 6. Pseudocode processing text

### E. Testing Program

This stage is the stage to run the classification program and display the dataset in the form of training data and test data which will be processed into a classification program using the python programming language, this classification process uses the Naive Bayes Classifier method by comparing each sentence in the data contained using a dataset that has been After that, enter the Twitter API token bearer for online data retrieval, after that enter the query or input the word you want to search, then the results in this program are sentences that are inputted and classified into positive and negative sentiments. The number of words is calculated based on the sentences contained in the search

for tweet data, the calculation that is assessed is based on the word opportunity value, if the assessment is greater than the positive opportunity result, the result is positive and vice versa. The results that can be obtained are that there are 100 tweets every time the program runs, this is due to the policy of the Twitter developer which limits every request for tweet data retrieval. From the tweet data, it can be identified by preprocessing the data. The assessment parameters are based on the dataset and slang words that have been made, from these parameters the resulting value can use the opportunity formula, in this case each word contained in the sentence is assessed based on the existing parameters for further analysis of the classification results.

### F. Result

Each word as a liaison sentence can be assessed if the content of the comment indicates positive/negative, in this case the amount produced is different depending on the content of the sentence. The abbreviations and non-standard languages have been adapted to the programming language so that they can be judged if there is writing using a language outside the official dictionary (colloquial language). The results of the study obtained several tests of tweet samples in the form of assessment of accuracy, precision, recall, F1-score, and response time.

TABLE II. NAIVE BAYES TEST RESULTS

Tweet Sample	Measurement				
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Response Time (s)
10	80	100	71,43	83,33	19,92
20	65	100	63,16	77,42	14,85
30	73	85	78,26	81,82	16,09
40	82	96	81,82	88,52	16,38
50	80	100	71,43	83,33	12,92

The results obtained are based on word searches that are processed to find the value of a tweet based on slang words. Accuracy is the percentage of the total sentiment that is correctly recognized. The accuracy calculation is done by dividing the correct amount of sentiment data by the total data and test data. Precision is a comparison of the amount of relevant data found to the amount of data found. The calculation of precision is done by dividing the number of correct data with positive values divided by the number of correct data with positive values and false data with positive values. The value of the false positive data is taken from the number of values other than the true positive column that corresponds to each class. Recall is a comparison of the amount of relevant material found to the amount of relevant material. The recall calculation is done by dividing the correct data with a positive value by the sum of the correct data which has a positive value and the incorrect data with a negative value. The value of the false data that is negative is taken from the number of values other than the true positive row that corresponds to each class. F1-score is a single parameter measuring retrieval success that combines recall and precision. The F1-score value is obtained from the calculation of the multiplication of precision and recall divided by the result of the addition

of precision and recall and then multiplied by two. To get the classification results, you must go through the word splitting process from each tweet taken, the word from the tweet is calculated its value so that it becomes a positive / negative classification value, with the parameter values can be processed based on positive and negative opportunities that make that value as the final result of classification. If the value is greater on a positive probability then the classification result will be positive, this is due to the calculation in the Naives Bayes method which emphasizes the comparison of the dataset values collected, if the summed value is greater then the result is what makes the final value or the classification results in the analysis used. The results of this study can be seen in the form of the results of the test graphs that have been carried out as follows:

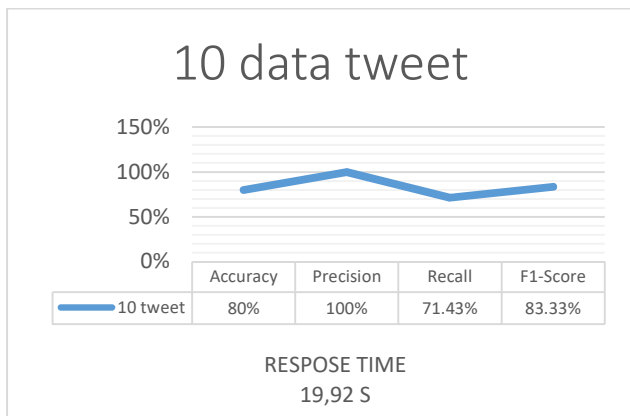


Fig. 7. Ten Tweet

The results on Figure 7 show that the response time decreases with the addition of data. This is because the queries that are searched are few so that the program can access the data more quickly. On the results of the classification test for 10 tweet data, the accuracy value is 80%, the precision value is 100%, the recall value is 71.43%, the F1-score is 83.33% and the load response time data is 19.92 s. The results on the graph show the performance value of the program.

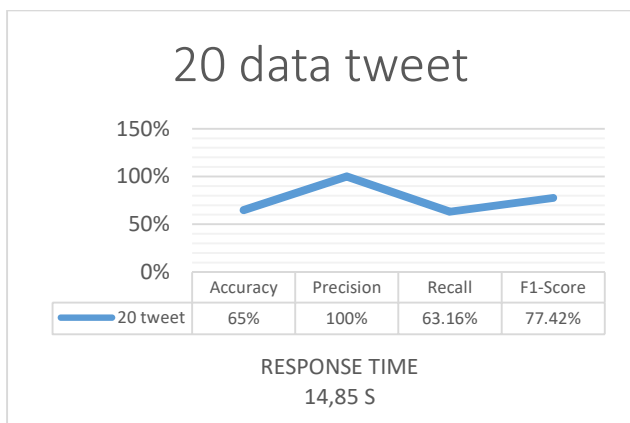


Fig. 8. Twenty Tweet

The results on Figure 8 show that the response time decreases with the addition of data. This is because the queries that are searched are few so that the program can

access the data more quickly. In the classification test results of 20 tweet data, the accuracy value is 65%, the precision value is 100%, the recall value is 63.16%, the F1-score is 77.42% and the load response time data is 14.85 s. The accuracy value in testing 20 tweet data is quite low, this is because the results of the test data or data samples used in the form of tweets are inaccurate in the assessment of the program to classify the contents of the content. In data collection the sample is taken randomly and collected on the sample dataset to be tested. The recall value experienced a decrease in the performance of the test dataset because the data collected in the sample was not very good, this had an effect if the data taken had a lot of inappropriate assessments.

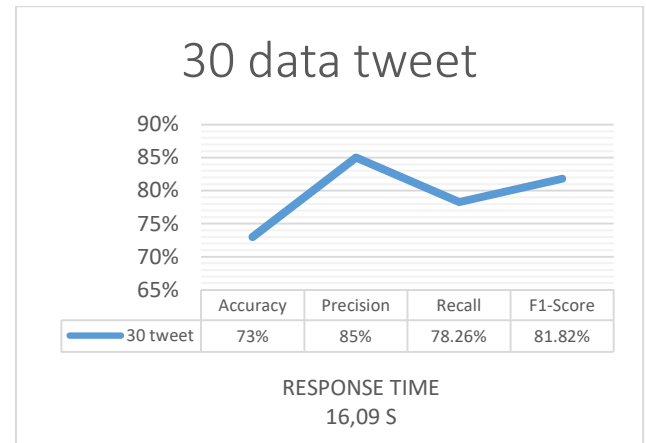


Fig. 9. Thrity Tweet

The results on Figure 9 show that the response time decreases with the addition of data. This is because the queries that are searched are few so that the program can access the data more quickly. The results of the classification test for 30 tweet data obtained an accuracy value of 73%, a precision value of 85%, a recall value of 78.26%, an F1-score of 81.82% and a load response time of 16.09 s. The accuracy value in testing 30 tweet data is quite good, this is because the results of the test data or data samples used in the form of tweets are quite accurate in the assessment of the program for classifying content. In data collection the sample is taken randomly and collected on the sample dataset to be tested.

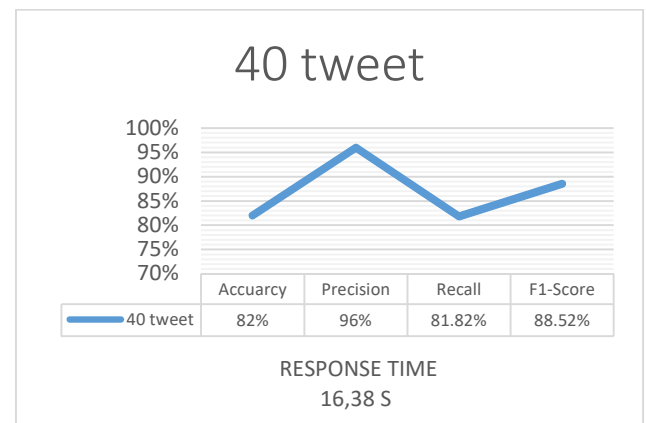


Fig. 10. Forty Tweet

The results on Figure 10 show that the response time decreases with the addition of data. This is because the queries that are searched are few so that the program can access the data more quickly. In the classification test results of 40 tweet data, the accuracy value is 82%, the precision value is 96%, the recall value is 81.82%, the F1-score is 88.52% and the load data response time is 19.92 s. The accuracy value in the test of 40 tweet data is quite good, this is because the results of the test data or data samples used are in the form of a more accurate collection of tweet datasets in the assessment of the 40 tweet data testing program to classify the content. In data collection the sample is taken randomly and collected on the dataset to be tested. The recall value has increased the performance of the test dataset because the data collected on the sample is good enough, this can have an effect if the data taken has a lot of appropriate judgments in predicting sentiment.

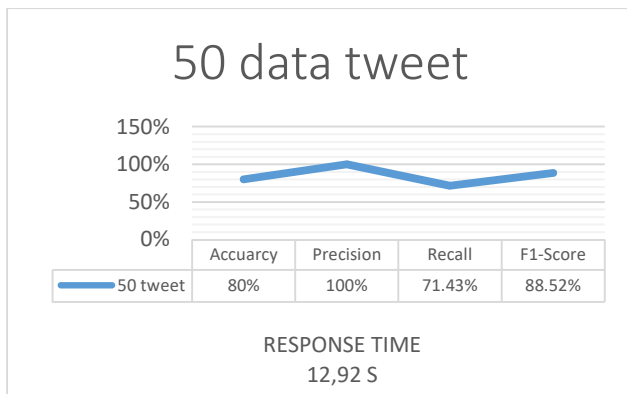


Fig. 11. Fifty Tweet

The results on Figure 11 show that the response time decreases with the addition of data. This is because the queries that are searched are few so that the program can access the data more quickly. The accuracy value in testing 50 tweet data is quite good, this is because the results of the test data or data samples used in the form of tweets are quite accurate in the assessment of the program to classify the contents of the content. In data collection the sample is taken randomly and collected on the sample dataset to be tested. The recall value experienced a decrease in the performance of the test dataset because the data collected in the sample was not very good, this had an effect if the data taken had a lot of inappropriate assessments. In an application, recall is the ratio of true positive predictions compared to the total number of true positive data. On the results of the classification test of 50 tweet data, the accuracy value is 80%, the precision value is 100%, the recall value is 71.43%, the F1-score is 88.52% and the load response time data is 12.92 s.

### G. Discussion

After conducting the testing, we can see that the main problem from this research is time consumption. It needs to be an important and main concern, remembering data always changes and increases as time goes on. Based on data from the chart, it indicates that more than ten second on average. It seems to be a long process or heavy. Some alternative methods of computing, for example parallel

programming or computing need to be considered as one of the solutions. If we are leaning toward parallel computing, all the procedure and model are needed converting into a matrix and linear algebra.

### V. CONCLUSION

In the sentiment analysis program using the Naïve Bayes method, it is used to determine the classification of twitter sentiment, using the opportunity parameter that most often appears in each word. The opportunity value is calculated based on the word value of the tweets obtained, then classified with positive or negative results. The number of values obtained is based on the parameters of the data repository and slang words. In the test results obtained the highest accuracy value, namely in testing 40 tweet data with a value of 82% and for the lowest value obtained in testing 20 tweet data at 65%. The highest precision when testing 10 and 20 tweet data with the results obtained 100% and the lowest value obtained in testing 30 tweet data the results obtained were 85%. The highest recall value obtained in testing 40 tweet data is 81.82% and for the lowest test it is obtained in testing 20 tweet data at 63.16%. The highest f1-score test results were obtained when testing 40 and 50 tweet data with the same value of 88.52% and for the lowest test it was obtained at 20 tweet data of 77.42%.

### REFERENCES

- [1] S. Kemp, "Digital 2022: Indonesia," 2022. <https://datareportal.com/reports/digital-2022-indonesia>.
- [2] S. Rani, N. S. Gill, and P. Gulia, "Survey of Tools and Techniques for Sentiment Analysis of Social Networking Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 4, 2021, doi: 10.14569/IJACSA.2021.0120430.
- [3] T. N. Simanjuntak and S. Pramana, "Sentiment Analysis on Overseas Tweets on the Impact of COVID-19 in Indonesia," *Indones. J. Stat. Its Appl.*, vol. 5, no. 2, 2021, doi: 10.29244/ijasa.v5i2p304-313.
- [4] V. L. Narasamma and M. Sreedevi, "Twitter based Data Analysis in Natural Language Processing using a Novel Catboost Recurrent Neural Framework," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 5, 2021, doi: 10.14569/IJACSA.2021.0120555.
- [5] F. M. Javed Mehedi Shamrat, Z. Tasnim, A. K. M. Sazzadur Rahman, N. I. Nobel, and S. A. Hossain, "An effective implementation of web crawling technology to retrieve data from the world wide web (Www)," *Int. J. Sci. Technol. Res.*, vol. 9, no. 1, 2020.
- [6] I. Dongo, Y. Cadinale, A. Aguilera, F. Martinez, Y. Quintero, and S. Barrios, "Web Scraping versus Twitter API," 2020, doi: 10.1145/3428757.3429104.
- [7] L. C. Dewi, Meiliana, and A. Chandra, "Social media web scraping using social media developers API and regex," in *Procedia Computer Science*, 2019, vol. 157, doi: 10.1016/j.procs.2019.08.237.
- [8] S. M. Alzahrani, "Big Data analytics tools: Twitter API and spark," 2021, doi: 10.1109/WIDSTAIF52235.2021.9430205.