Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm

Auliya Rahman Isnain*[1], Nurman Satya Marga[2], Debby Alita[3]
[1,2,3]Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Lampung, Indonesia
e-mail: ***[1]auliyarahman@teknokrat.ac.id**, [2]debbyalita@teknokrat.ac.id,
[3]nurman@teknokrat.ac.id

Abstrak
Pemerintah Indonesia memberlakukan peraturan New Normal dalam menjaga stabilisasi ekonomi dan juga menahan penyebaran virus dimasa pademi covid 19. Hal ini menjadi perbincangan hangat di media sosial twitter banyak masyarakat yang beranggapan positif maupun negatif.

Penelitian yang dilakukan adalah representasi dari text mining dan text processing menggunakan pembelajaran mesin dengan menggunakan metode klasifikasi Naive Bayes Classifier, tujuan analisis adalah untuk mengetahui apakah sentimen masyarakat terhadap kebijakan New Normal positif atau negatif, dan juga sebagai dasar pengukuran performa ekstrasi fitur TF-Idf dan N-gram pada machine learning menggunakan metode Naive Bayes.

Hasil dari penelitian ini menghasilkan tingkat akurasi metode Naive Bayes dengan seleksi fitur TF-IDF mendapat total akurasi sebesar 81% dengan nilai Precission 78%, Recall 91% dan f1-Score 84%. Hasil tertinggi didapatkan dari penggunaan parameter algoritma Naive Bayes dan Trigram yaitu sebesar 84% yaitu Precission 84%, Recall 86%, dan f1-Score 85%. Algoritma Naive Bayes dengan penggunaan ekstrasi fitur N-Gram jenis trigram menunjukan performa yang cukup baik dalam proses pengklasifikasian data tweet masyarakat.

Kata kunci— *Sentimen analisis, corona, new normal, Naive Bayes, Tf-IDF, dan N-Gram.*

Abstract
The Indonesian government has enforced the New Normal rule in maintaining economic stabilization and also restraining the spread of the virus during the Covid 19 pandemic. This has become a hot topic of conversation on social media Twitter, many people think positive and negative.

The research conducted is a representation of text mining and text processing using machine learning using the Naive Bayes Classifier classification method, the objective of the analysis is to determine whether public sentiment towards the New Normal policy is positive or negative, and also as a basis for measuring the performance of the TF-IDF feature extraction and N-gram in machine learning uses the Naive Bayes method.

The results of this study resulted in the accuracy rate of the Naive Bayes method with the TF-IDF feature selection. The total accuracy was 81% with a Precision value of 78%, Recall 91%, and f1-Score 84%. The highest results were obtained from the use of the Naive Bayes and Trigram algorithm parameters, namely 84%, namely 84% Precision, 86% Recall, and 85% f1-Score. The Naive Bayes algorithm with the use of the trigram type N-Gram feature extraction shows a fairly good performance in the process of classifying public tweet data.

Keywords— *Sentiment analysis, coronavirus, new normal, Naive Bayes, Tf-IDF, and N-Gram.*

1. INTRODUCTION

In this era, social media has become an important part of everyday life for the wider community [1]. Social media that are still widely used by people today, one of which is Twitter [2][3]. Twitter is an online social networking and microblogging service that allows users to send and read text-based messages of up to 140 characters, known as Tweets [4]. Twitter is an online social networking and microblogging service that allows users to send and read text-based messages of up to 140 characters, known as Tweet [5][6].

Sentiment analysis or Opinion Minning is a field of textual data management that conducts studies based on opinion, sentiment, evaluation [7], behavior and emotions of a person that can be used as evaluation material. There have been many studies on sentiment analysis by applying several methods such as Naive Bayes, Support Vector Machine (SVM) [8], Multinomial Naive Bayes [9]. Some residents considered the government policy positively in imposing the New Normal during the Covid-19 pandemic, from monitoring on social media, Twitter, it was known that the public gave various opinions ranging from positive to negative opinions on the policy. Negative opinions can be caused because there is a lack or imperfection of information on government policies in dealing with Covid-19 or because of subjectivity factors. These public opinions can then be processed and analyzed using methods known as sentiment analysis.

In the research of [9] the Machine Learning approach in the research carried out was using the Naive Bayes Classifier (NBC) method, with the collaboration of the Chi-Square feature selection. This research was conducted with a total of 200 data opinions about the English-language opinion of mobile phone users divided into 100 positive opinions and 100 negative opinions. As training data used as many as 100 data, namely data divided respectively 50 positive opinions and 50 opinions. The remaining 100 opinions that are equally divided between positive and negative opinions are used as test data. The results of the NBC classification on the negative test data obtained an accuracy of 72% while the positive test data obtained 96%. Overall, the test data, classification with NBC get an accuracy of 83% and a harmonic average of 90.713%. In this study, NBC made a misclassification of four test data, unable to classify 13 test data from a total of 100 test data. While [4], in their research, analyzed by preprocessing normalization and stemming using 2 methods, namely the Naive Bayes Classifier (NBC) and Support Vector Machine (SVM). In this study, Boolean Searching was used to obtain the required data, from the results of the test data obtained and the method test, the accuracy results were obtained when the data was stemming, there was an average increase of 0.85% for the Naive Bayes method and 0, 85% for the SVM method. The accuracy produced by the SVM method is not always superior to the Naive Bayes method, and vice versa. For the highest method in each experiment is the SVM method in experiment 1 getting an accuracy of 88.7006% for the preprocessing technique C (TF-IDF = Yes, Lowercase = Yes, minTermFreq = 1, Normalize all data, Stopwords = No, Tokenizer = N-Gram, and Emoticons = Yes.). experiment 2 got an accuracy of 89.2655% for the preProcessing technique D (TF-IDF = Yes, Lowercase = Yes, minTermFreq = 1, Normalize all data, Stopwords = No, Tokenizer = Unigram, and Emoticon = Yes).

Based on the description of the existing problems, sentiment analysis using the machine learning method, namely the Naive Bayes Classifier, was carried out in this study to know the public's response to government policy in handling the Covid19 case by dividing 2 sentiment classifications, namely Positive and Negative. As far as we know, there is still no research that examines public opinion regarding the new normal. Furthermore, in this study the data we took from the public's tweets on Twitter from 6 July 2020 to 10 September 2020. This analysis process was carried out using a text mining approach, and machine learning in its implementation to produce conclusions about public opinion about the policy and also testing the level of accuracy, precision, recall, and f1-score on the Naive Bayes Classifier method with the extraction of the N-Gram and TF-IDF features in classifying tweet data about government policies regarding the Covid case19.

2. METHODS

2.1 Research Stages

Research is starting from the existence of an important, interesting problem and the need for a solution. To create effective & efficient research, a structured frame of mind is needed and conveyed through pictures with structured stages related to the actions to be taken, which can be seen in Figure 1.
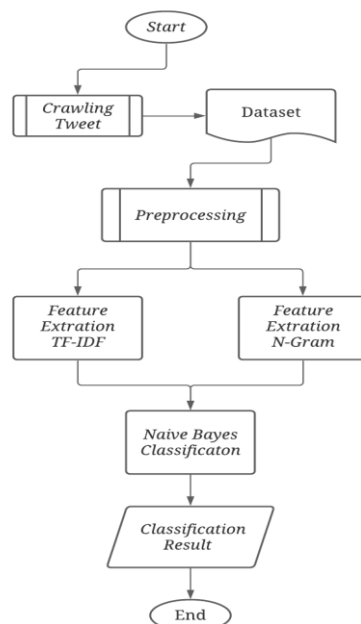

Figure 1. Research Stages

2. 2. Data Crawling

Data crawling is the initial stage of this research, data crawling functions in collecting the dataset. The tweet data crawling was carried out on tweets written by Twitter users on July 6, 2020 - September 10, 2020, using the Twitter API (Application Programming Interface).

2. 3 Data Preprocessing

The dataset collected using the tweepy library in the Python programming language is not easy enough to understand because it requires word processing (Text Processing) so that the data produces more concise words that contain sentiments by selecting and removing unnecessary words. required to produce good input data in the labeling, training, and testing process. Data pre-processing includes several processes, namely : 1. Escapping HTML characters, 2. Case Folding, 3. Stemming, 4. Remove of punctuation, 5. Tokenization, 6. Stopwords Removal[10][11].

1. Escapping HTML characters
   At the beginning of data collection (crawling), there is a tweet that contains a link, in this case, the process of Escaping HTML characters is very necessary because this process aims to remove URL links and also HTML characters found in the tweet.
2. Case Folding
   A process of uniformity of words in a tweet, in this case, the letters used in the word are lowercase. Text data containing uppercase words will be converted into lowercase words.
3. Stemming
   The process of selecting words that have conjunctions, affixes, pronouns, and verbs into root words, by removing the prefix or suffix.
4. Remove of punctuation

This process is used to delete special characters such as (#), @, (RT). In addition, punctuation marks (,.?!, Etc.), numeric numbers (0 ... 9), and other characters (+, $, ~, etc.) are removed.

5. Tokenization
The process for cutting words in a sentence into token form, where each word in one sentence is separated by a space.

6. Stopwords Removal
The process of deleting words that do not affect the classification process for example: and, to, or, from, who, etc.

2. 4 Feature Selection

In this study, the features used in processing the dataset are TF-IDF and N-Gram. N-Gram is a combination of adjectives that often appear for showing a sentiment[12]. An example of using N-Gram can be seen in Table 1. The weighting process using the TF-IDF feature selection can be seen in Table 2.

Table 1 Example of N-Gram

N-Gram	Results of N-Gram Extraction
Unigram	kasus, meninggal, covid, tinggi
Bigram	kasus meninggal, meninggal covid, covid tinggi
Trigram	kasus meninggal covid, meninggal covid tinggi

Table 2 TF-IDF weighting

Term(t)	TF						IDF
	D1	D2	D3	D4	D5	D6	
Kasus	1	0	0	0	0	0	0,778
Meninggal	1	0	0	0	0	0	0,778
Covid	1	1	0	0	0	0	0,477
Tinggi	1	0	0	0	0	0	0,778
Newnormal	0	1	1	0	0	0	0,477
Indonesia	0	1	1	0	0	0	0,477
Atasi	0	1	0	0	0	0	0,778
Apa	0	0	1	0	0	0	0,778
Sudah	0	0	1	0	0	0	0,778
Siap	0	0	1	0	0	0	0,778
Kebijakan	0	0	0	1	0	0	0,778
Baik	0	0	0	1	0	0	0,778
Ekonomi	0	0	0	1	0	0	0,778
Pemimpin	0	0	0	0	1	0	0,778
Dungu	0	0	0	0	1	0	0,778
Bencana	0	0	0	0	1	0	0,778
Optimis	0	0	0	0	0	1	0,778
Industri	0	0	0	0	0	1	0,778
kembang	0	0	0	0	0	1	0,778

2. 5 Classification

Classification of data in this study uses a machine learning method, namely the Naive Bayes Classifier (NBC) using a Feature Selection. Feature Selection is a process of selecting features in a more efficient classification process by identifying relevant features which are then processed based on the classifier model that has been generated from the training dataset process. The feature selection method used in this research is TF-IDF. In its calculation, the Naive Bayes algorithm has the following rules. Example of a training data set:

In this sentiment analysis, words are divided into three classes (categories), namely:

1. A1= Positive Sentiment
2. A2= Negative Sentiment

In each class, the probability value of each vocabulary is obtained through a set of tweets through equations $P(Vj) = \frac{|docs\ j|}{|contoh|}$, as shown in Table 3 below:

Table 3 P value (Vj) in each class

Data	Emergence	Sentiment	P(Vj)
A1	newnormal (2) baik (1) ekonomi (1) berkembang (1) optimis(1)	Positif	$\frac{1}{2} = 0{,}5$
A2	kasus (1) meninggal (2) covid (2) tinggi (1) bencana(1) newnormal(1)	Negatif	$\frac{1}{2} = 0{,}5$

For each word xi in class Vj, a calculation based on the equation $P\,(xi \mid vj) = nk + 1n + \mid kosakata \mid$ as an example to display the calculation, one word will be taken in each class, namely the calculation of the word "case" with a total vocabulary of 11 words, as in table 3.11 below: For each word xi in class Vj, a calculation based on the equation $P\,(xi \mid vj) = nk + 1n + \mid kosakata \mid$ as an example to display the calculation, one word will be taken in each class, namely the calculation of the word "case" with a total vocabulary of 11 words, as in Table 4 below:

Table 4 P values (cases) in each sentiment class

V_i	Sentiment Positive		Sentiment Negative	
	Nk	N	Nk	N
	0	6	1	7
P(xi | Vj)	$\dfrac{1}{17}$		$\dfrac{2}{18}$	

The same is applied to each word xi so that the P value (xi) is obtained for each class Vj and a probabilistic model is obtained as in Table 5 :

Table 5 Probabilistic Model

V	P(X$_i$|V$_j$)	
	Positive P(V$_j$) = 0,5	Negative P(V$_j$) = 0,5
Kasus	$\dfrac{1}{17}$	$\dfrac{2}{18}$
Meninggal	$\dfrac{1}{17}$	$\dfrac{3}{18}$
Covid	$\dfrac{1}{17}$	$\dfrac{3}{18}$
Tinggi	$\dfrac{1}{17}$	$\dfrac{2}{18}$
Bencana	$\dfrac{1}{17}$	$\dfrac{2}{18}$
Newnormal	$\dfrac{3}{17}$	$\dfrac{3}{18}$
Baik	$\dfrac{2}{17}$	$\dfrac{1}{18}$
Ekonomi	$\dfrac{2}{17}$	$\dfrac{1}{18}$
Optimis	$\dfrac{2}{17}$	$\dfrac{1}{18}$

The process of calculating predictions on the Naive Bayes algorithm can be seen in the example of predictive data in Table 6 :

Table 6 Examples of predictive data

Tweets that have been in Preprocessing	Test Data
Korban covid semakin tinggi pasca berlaku kebijakan	Korban(1) covid(1) semakin (1) tinggi(1) pasca(1) berlaku(1) kebijakan(1)

Stage 1, Calculating the Vmap value for each class

$$Vmap = V_j\{\text{Positif, Negatif}\}P(Vj)\Pi\,P(X_i|V_j) \tag{3.2}$$

$Vmap = Vj\{Positive,$Negative$\}$ $P(Vj)$ P("korban" | Vj) P("covid" | Vj) P("semakin" | Vj) P("tambah" | Vj) P("pasca" | Vj) P("berlaku" | Vj) P("kebijakan" | Vj)

Value Vmap Positive :

$Vmap$ ("Positive") = P("Positif") P("korban"|"Positif") P("covid"|"Positif") P("semakin"|"Positif") P("tinggi"|"Positif") P("pasca"|"Positif") P("berlaku"|"Positif") P("kebijakan"|"Positif")

$$= \frac{1}{2}x\frac{1}{17}x\frac{1}{17}x\frac{1}{17}x\frac{1}{17}x\frac{1}{17}x\frac{1}{17}x\frac{1}{17}$$
$$= 1{,}2 \times 10^{-9}$$

$Vmap$ ("Negative") = P("Negative") P("korban"|"Negative") P("covid"|"Negative") P("semakin"|"Negative") P("tinggi"|"Negative") P("pasca"|"Negative") P("berlaku"|"Negative") P("kebijakan"|"Negative")

$$=\frac{1}{2}x\frac{1}{18}x\frac{4}{18}x\frac{1}{18}x\frac{3}{18}x\frac{1}{18}x\frac{1}{18}x\frac{1}{18}$$
$$= 9{,}8 \times 10^{-9}$$

From the calculations done above, it is found that the Vmap value for the negative sentiment class has the highest value, this shows that the tweets are included in negative sentiment. If the value of Vmap on a positive sentiment and a negative sentiment is the same, then the Tweet will be considered a negative sentiment class by assuming negative sentiment, the government will review the policy.

3. RESULTS AND DISCUSSION

This stage is carried out by calculating and measuring the system performance against the level of accuracy, recall, precision, and f1-Score on the results of the Naive Bayes classification method from Twitter data regarding government policies regarding the implementation of the New Normal. The results of the test can be seen whether the Naive Bayes method has a good performance in predicting tweet sentiment about government policy on the implementation of the New Normal, from the test results also seen positive responses or more negative responses about the tweet so that it can be used as a reference by decision-makers.

In testing tweet sentiment data about government policies regarding the implementation of the New Normal, the results of negative sentiment and positive sentiment were manually obtained. From the results of the data labeling, training data, and data testing were carried out, namely 80% training data and 20% random data testing. The results of the calculation of negative class sentiment data are 795 data, and the positive class sentiment is 1028 data, while the comparison in percentage form can be seen in Figure 2.
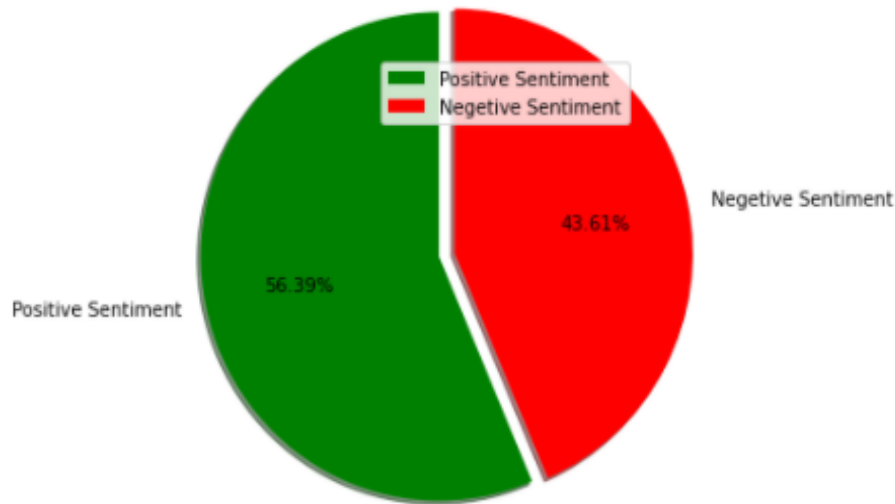
Figure 2 The percentage of negative and positive class comparisons

From Figure 2 we can see a comparison between the public's tweets that commented positively on government policies in the implementation of the New Normal order system with those of the public who commented negatively on government policies regarding the New Normal, namely 56.39% positive tweets compared to 43.61% negative tweets with the number Negative class is 795 data, and the positive class sentiment is 1028 data.

3.1 TF-IDF and N-Gram Testing Against Naive Bayes

Testing is a process in implementing a program with the aim of finding an error or testing the functional data [10]. Machine learning testing uses the Naive Bayes method with test parameters using the TF-IDF and N-Gram feature selection, and the results are as shown in Table 7.

Table 7 Classification Results using TF-IDF & N-Gram

Feature Selection	Accuracy
TF-IDF	0.819
N-Gram (Unigram)	0.827
N-Gram (Bigram)	0.835
N-Gram (Trigram)	0.841

From the test data using TF-IDF feature extraction and N-Gram highest accuracy results obtained on the N-Gram (Trigram) amounted to 84.1%, while the lowest was in the use of the results of feature selection TF-IDF with total accuracy at 81, 9%, Based on the results of the accuracy-test using TF-IDF and N-Gram that has been done, it shows an increase in classification performance when using N-Gram while TF-IDF is deemed not sufficient to show good performance in applications using the Naive Bayes method. The improved performance of the Naive Bayes classification on the selected feature can be seen in Figure 3.
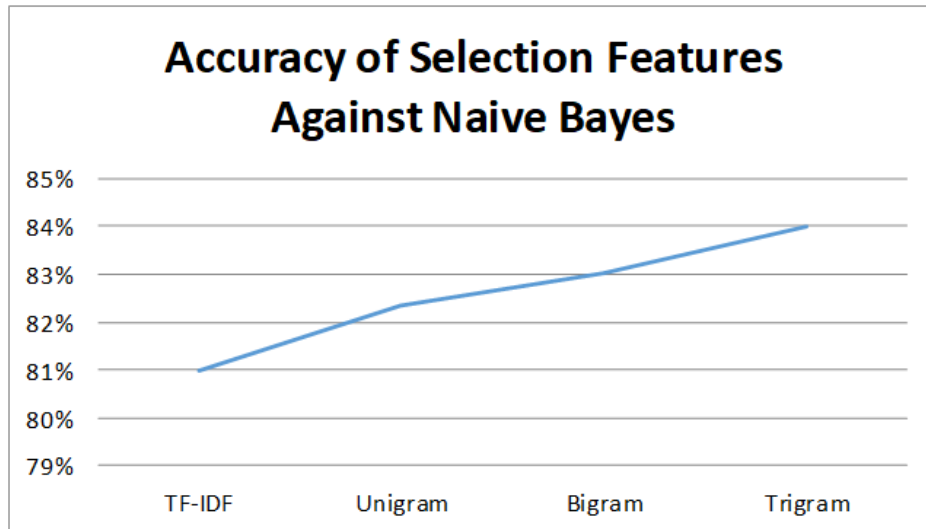
Accuracy of Selection Features Against Naive Bayes

Figure 3 Classification Accuracy Graph

3.2 Confusion Matrix Test Results

        Performance of the Naive Bayes Algorithm in analyzing the sentiment of the dataset about the public's view of tweeting about government policy in implementing new normal. In inaccuracy testing, the N-Gram feature selection with the Trigram parameter is the highest accuracy value, in testing this accuracy is obtained through confusion matrix testing concerning TP (positive results detected correctly), TF (Negative results detected correctly), FP (Positive results detected false), and FN (a negative result reads incorrectly). The confusion matrix in the TF-IDF feature selection test can be seen in Figure 4, while testing using the N-Gram feature selection with Trigram parameters in the Naive Bayes algorithm can be seen in Figure 5.
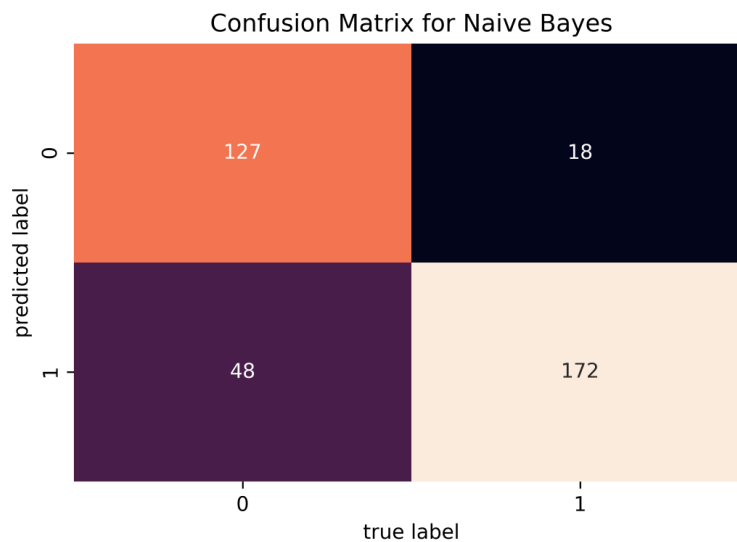
Figure 4 TF-IDF Confusion Matrix Test Results

The test results in Figure 4 can be seen that confusion matrix testing on TF-IDF feature extraction using the Naive Bayes classification gets the results TP = 172, TN = 127, FP = 48, and FN = 18.

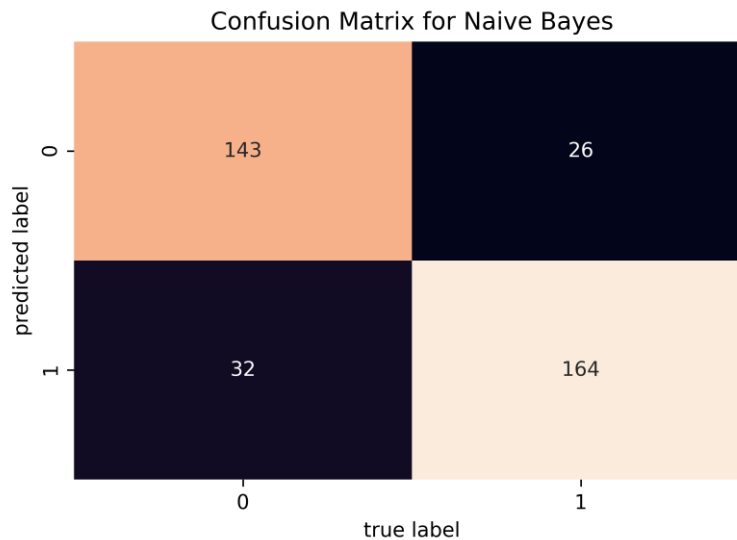Confusion Matrix for Naive Bayes

Figure 5 The Results of Testing the Confusion Matrix N-Gram (Trigram)

The results of the Confusion Matrix test using the Trigram parameter N-Gram feature extraction using the Naive Bayes classification obtained the results of TP = 164, TN = 143, FP = 32, and FN = 26 after the system obtained the data, the last test was precision, recall, and F1-Score in Table 8.

Table 8 Confusion Matrix Test Results

Testing	TF IDF	N-Gram (Trigram)
Precission	0.78	0.84
Recall	0.91	0.86
F1-score	0.84	0.85
Accuracy	0.81	0.84

In Table 8, it is found that the report information from the TF-IDF feature selection test results in 78% Precission, 91% Recall, 84% f1-score, and 81% accuracy. While the report from the selection of N-Gram features with Trigram parameters on the Naive Bayes classification with a Precission value of 84%, Recall 86%, f1-Score 85%, and 84% accuracy.

4. CONCLUSIONS

The test results using the Confusion Matrix obtained the lowest accuracy value with the Naive Bayes classification using the TF-IDF feature extraction, which is 81%, with a Precission value of 78%, Recall 91% and f1-Score of 84%. While the highest accuracy value was obtained using the parameters of the Naive Bayes algorithm with the N-Gram Trigram type, which was 84%, there was also a Precission value of 84%, Recall 86%, and f1-Score 85%. This shows that the Naive Bayes Algorithm using TF-Idf and N-Gram feature extraction can be used well in the process of classifying public tweet data against government policies on the implementation of the New Normal system. The increase in accuracy obtained after using N-Gram is because many Indonesian phases have 2 to 3 vocabulary words.

REFERENCES

[1]   I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekanata, "Hate Speech Detection in the Indonesian Language : A Dataset and Preliminary Study," in *9th Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS 2017)*, 2017, no. October.

[2]   A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, p. 169, 2020.

[3]   M. A. Nurrohmat and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 3, p. 209, 2019.

[4]   N. Saputra, T. B. Adji, and A. E. Permanasari, "Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM," *J. Din. Inform.*, vol. 5, no. November, p. 12, 2015.

[5]   D. Alita, Y. Fernando, and H. Sulistiani, "Sentiment Ana," *Tekno Kompak*, vol. 14, no. 2, pp. 86–91, 2020.

[6]   R. R. Suryono and I. Budi, "P2P Lending Sentiment Analysis in Indonesian Online News," vol. 172, no. Siconian 2019, 2019.

[7]   N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 8, no. 1, p. 91, 2014.

[8]   J. Patihullah and E. Winarko, "Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 1, p. 43, 2019.

[9]   L. Juen, I. P. E. N. Kencana, and T. B. Oka, "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square," *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014.

[10] Y. Woldemariam, "Sentiment analysis in a cross-media analysis framework," *2016 IEEE Int. Conf. Big Data Anal.*, pp. 1–5, 2016.

[11] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, pp. 1–5, 2019.

[12] R. Hidayatillah, M. Mirwan, M. Hakam, and A. Nugroho, "Levels of Political Participation Based on Naive Bayes Classifier," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 1, p. 73, 2019.