JAISI
JOURNAL OF APPLIED INFORMATON SYSTEM AND INFORMATIC

Implementation of the Naive Bayes Classifier for Sentiment Analysis of Shopee E-Commerce Application Review Data on the Google Play Store

Adilia Tri Rizkya[a] ,Rianto [a], Acep Irham Gufroni[b,*]

[a]*Informatics Department, Universitas Siliwangi, Tasikmalaya, Indonesia*

[b]*Information System Department, Universitas Siliwangi, Tasikmalaya, Indonesia*

Corresponding author: acep@unsil.ac.id

Abstract*— E-commerce in Indonesia is growing very quickly every year. The Ministry of Communication and Information (KEMKOMINFO) stated that Indonesia is the 10th largest e-commerce growth country with score 78%. One of the effects from increasing number of internet users in Indonesia is the mushrooming of shopping activities through internet media. This causes internet users want everything that instant and easy. Knowing this, most business people use it to market their products, especially in the field of goods and services. As it grows, e-commerce becomes easier to use and download. One example of an e-commerce application that is in great demand is Shopee and can be downloaded via the Google Play Store. Google Play Store has a review feature which contains user comments about the downloaded apps. Sentiment analysis is carried out to extract information related to Shopee E-commerce. The Naïve Bayes Classifier algorithm is suitable for use in sentiment analysis because this algorithm is purposeful as a classification method into positive and negative categories. The data was used from November 2022 to January 2023. From a total of 4902 review data obtained, after going through preprocessing, translation and then classification, the total data is obtained that is 4849 review data. From the data obtained it is classified 2348 positive reviews, 1259 neutral reviews, and 1242 negative reviews. Based on the results of the naive Bayes classifier method and testing with the confusion matrix, an accuracy value of 79% has been obtainednprecision 77%, recall 86%, and f1-score 81% on positive sentiment with support 2127. For neutral sentiment with an accuracy value of 83%, precision 87%, and recall 85% with support 1209, while for negative sentiment is with an accuracy value of 78%, precision 64%, and recall 70% with support 1513. From this data it is obtained micro AVG values for precision 80%, recall 79%, f1-score 79%, and support 4849, then for weighted average for precision 79%, recall 79%, f1- score 79%, and support 4849.

***Keywords*— Sentiment Analysis, E-Commerce, Google Play, Naïve Bayes Classifier , Shopee**

Manuscript received 15 Jul. 2023; revised 29 Oct. 2023; accepted 15 Nov. 2022. Date of publication Nov. 2023.

I. INTRODUCTION

The rapid development of technology can make it easier for people in various aspects of life. The constantly growing technology produces huge amounts of data which can provide useful and useful information when it can be processed and used [1] The number of active Internet users today makes the amount of data that can be generated also huge. With big data technology can help in the processing of large, large, and complex data so that the data that has been processed can give useful information [2].

E-commerce in Indonesia is growing very quickly every year. The Ministry of Communication and Information (KEMKOMINFO) stated that Indonesia is the 10th largest e-commerce growth country with growth of 78% and is ranked 1st. This cannot be separated from the facts that the number of internet users in Indonesia continues to increase. Based on the results survey of the Indonesian Internet Service Providers Association (APJII), internet users in Indonesia it will reach 215.63 million people in the 2022-2022 period. That amount an increase of 2.67% compared to the previous period which was 210.03 million users, and with this achievement Indonesia is ranked 4th in this world [3][4].

Google has a service called Play Store which provides digital content such as games, applications, films, music and books in various categories. One of the features found in the Play Store is the rating and review feature where users of products from the Play Store can give their opinions on the products they have used [3]. One of the e-commerce

applications available on the Play Store is Shopee. Shopee is an application used to carry out online buying and selling processes that can be used via smartphone. Shopee is a popular marketplace application, where in 2020 Shopee became the most clicked e-commerce application in Indonesia [5].

Sentiment analysis is a process carried out to provide information contained in unstructured datasets. This process is a computational process that is carried out by understanding, extracting and processing data in textual form automatically so as to obtain information contained in a person's opinion or behavior [3]. Sentiment analysis is useful for finding out whether users respond well or not to a product by extracting text from a review for knowing user emotions [6][7][8].

The research entitled "Classification On Categories Of Public Responses On Television Programs Using Naive Bayes Method" classifies public responses to programs on television. The results of this research show that of the 326 data that have been used as datasets with a percentage of 80% of the training data and 20% of the test data, it produces an accuracy value of 82%. [9]. Apart from that, research was carried out to analyze public opinion sentiment towards JNE expedition services using the Naive Bayes algorithm which produced an accuracy value of 85%, precision 78% and recall 67%. [10]. Research on sentiment analysis of anti-LGBT campaign cases in Indonesia comparing the Naïve Bayes, Decision Tree and Random Forest algorithms. Produces an accuracy value for Naïve Bayes of 86.43% accuracy, where the accuracy is higher than other algorithms, Decision Tree and Random Forest, which is 82.91% [11]. Research was carried out by comparing the Naïve Bayes and Decision Tree algorithms for analysis. Public sentiment towards COVID-19 vaccination in Indonesia shows that public opinion tends to be negative and the best algorithm in this research is the Naïve Bayes algorithm with an accuracy of 100.00%, while the Decision Tree algorithm produces an accuracy of 50.39% [12].

This research will be carried out using the Naïve Bayes algorithm with Google Colaboratory tools to carry out sentiment analysis on one of the e-commerce applications, namely Shopee. It is hoped that this research can help provide information about the sentiment contained in reviews given by customers about the application.

II. Materials and Method

The methodology in a research has guidelines in the form of a research flow or steps so that the expected results are in accordance with the initial objectives. The research methodology has a structured and appropriate flow design. The plot design in this research can be seen in Figure. 1 following

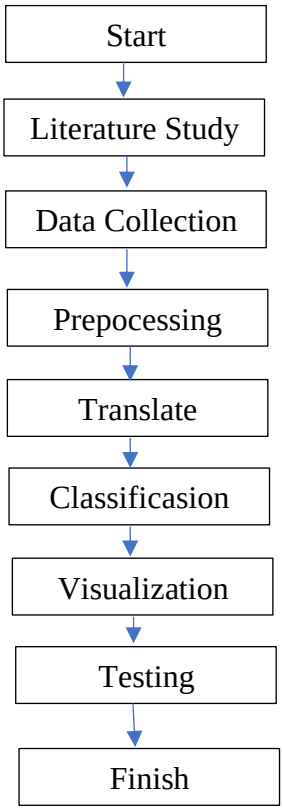


Fig. 1 Research Stages

A. *Literature Study*

The literature study stage is carried out to collect information and references from journals or scientific works related to concepts and theories regarding Sentiment Analysis, Naive Bayes Classifier, Preprocessing and others related to research.

B. *Data Collection*

This research uses a dataset of 4902 dataset obtained from e-commerce applications Shopee on the Google Play Store website. Dataset retrieval was carried out by Web scrapping from the Google Play Store using the Python programming language

C. *Preprocessing*

At this stage, data selection and cleaning of the review data that has been taken will be carried out. The following are the stages of the preprocessing process:

1) Cleaning
   At the cleaning stage, a process will be carried out to remove punctuation marks and unnecessary characters such as periods, commas, question marks, exclamation marks, emojis, as well as removing irrelevant characters.
2) Tokenization
   At this tokenization stage, a sentence from the review will be separated into chunks of words before being analyzed further.
3) Case Folding
   In the case folding stage, a collection of review sentences will be changed to all lowercase letters.

4) Filtering or Stopwords Removal
The filtering stage will carry out the process of eliminating words that have no meaning or stop words in order to focus on words that are more meaningful. In this way, the classification process will be faster and more efficient because the number of words processed will be fewer.

5) Stemming
At this stage, we will change the words with affixes into base words.

D. Translate

After the preprocessing stage is carried out, stage next is translating the review data in Indonesian into an English review. The translation stage is carried out because the next stage will use the textblob library which uses English.

E. Classification

After the translation stage is carried out, the review data classification stage continues. At this stage the sentiment labeling process is carried out and will produce a polarity score. The method used at this stage is the Naive Bayes Classifier.

Algorithms that use the concept of chance or what is usually called the probability used in classification for sentiment analysis it is called as a Naive Bayes Classifier. Naive Bayes Classifiers are also included in the algorithm easy to use and simple and can predicting an event based on the results from classification well [13]. Following is a formula for the calculation equation of value probability of Naive Bayes Classifier method:

$$P(X|Y) = \frac{P(Y|X)\ x\ P(X)}{P(Y)} \quad (1)$$

X = Temporary estimate of data from a specific class
Y = Data with unknown class
P(X|Y) = Estimated probability of X with conditions Y (posterior probability)
P(X) = Estimated probability of X (prior probability)
P(Y|X) = Estimated probability of Y with X
P(Y) = Probability of Y

Information :
Posterior probability: the possibility that class X exists
Prior probability: the possibility of the initial sample of class Y

F. Visualization

After each stage and process is carried out,Next is the visualization stage. In this research, the visualization stage was carried out using the Matplotlib and Wordcloud libraries. The output from the wordcloud visualization displays words that frequently appear in each sentiment.

G. Testing

When using a method, you certainly have an idea of the method's performance in the data classification process. The method used in this research is the confusion matrix method. Where the method used to calculate accuracy is by comparing the actual classification results with the classification results from the method [13]. For actual data classification, this research carried out labeling manually to determine the polarity.

Benchmark for the calculation results of the confusion method matrix namely Precision, Recall, F1-Score, Macro avg, Weighted avg and Accuracy.

1) Precision

Precision is a visualization of the percentage accuracy of the estimation results by the method used.

$$Precision\ = \frac{TP}{FP + TP}\ x\ 100\% \quad (2)$$

2) Recall

Recall yaitu visualisasi kesesuaian metode dalam mencari ulang sebuah informasi. Recall is a visualization of the suitability of a method for re-searching information.

$$Recall\ = \frac{TP}{TP + FN}\ x\ 100\% \quad (3)$$

3) F1-Score

F1-Score is a comparison between the average precision and recall values of test results.

$$f1 - score = \frac{TP}{TP + \frac{1}{2} + (FP + FN)}\ \ atau \quad (4)$$

$$f1 - score = 2\ * \frac{precision * recall}{precision + recall} \quad (5)$$

4) Macro avg

Macro avg is the unweighted average of all F1-Scores per class.

$$macro\ avg\ = \frac{jumlah\ nilai\ f1 - score}{jumlah\ kelas} \quad (6)$$

5) Weighted avg

Weighted avg is the average of all F1-Scores per class taking into account support each class.

$$weighted\ avg\ = \text{€}\ f1score\ per\ kelas\ * support\ proporsion \quad (7)$$

6) Accuracy

Accuracy is a visualization of the model's accuracy in grouping correctly.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}\ x\ 100\% \quad (8)$$

III. Result and Discussion

A. Literature Study

The stages of literature study carried out produced several theories that can be used as references in research.

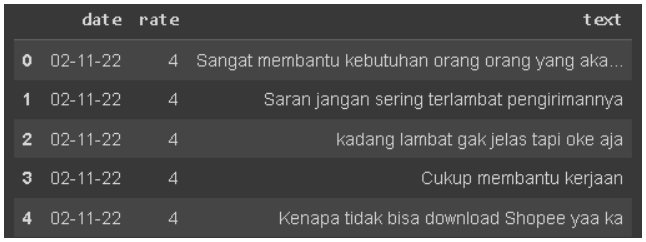
B. Data Collection

Data collection was 4902 review data from November 2 2022 to January 31 2023. After the review data was collected, the data collected obtained is converted into tabular data so that it is easy to process at the next stage. The data frame contains three attributes, the attributes are:

1) rating: rating contains the number of ratings given by the reviewer.

2) at : at contains the time when the user made the review.
3) content: content contains content that comes from reviews that have been made

The results of scraping research reviews are presented in Figure 2 below:



	date	rate	text
0	02-11-22	4	Sangat membantu kebutuhan orang orang yang aka...
1	02-11-22	4	Saran jangan sering terlambat pengirimannya
2	02-11-22	4	kadang lambat gak jelas tapi oke aja
3	02-11-22	4	Cukup membantu kerjaan
4	02-11-22	4	Kenapa tidak bisa download Shopee yaa ka

Fig. 2 Scraping Results

C. Preprocessing

In the preprocessing stage, several stages will be carried out, namely cleaning, case folding, tokenization, filtering or stopwords removal, and stemming. These stages are carried out for cleaning and deleting data from punctuation as well as unnecessary symbols such as periods, commas, question marks, exclamation marks, removing emojis, and removing irrelevant symbols. In the preprocessing stage, you can also change all the letters in the review data to lowercase, cut a sentence into word fragments, remove stop words, and change words with affixes into basic words.

1) Cleaning

At this cleaning stage, several characters will be removed, periods, commas, exclamation marks, question marks, removing emojis, and removing irrelevant symbols. Table 1 below is the result of the cleaning stage:

TABLE I
DATA CLEANING RESULTS

Before	after
Sangat membantu kebutuhan orang2 yang akan belanja karena sangat murah, dan mudah	Sangat membantu kebutuhan orang yang akan belanja karena sangat murah dan mudah

2) Tokenization

In the tokenization stage, the NLTK library in This research is used to tokenize reviews. This tokenization stage is used to obtain break down a sentence from the review into the pieces, the pieces in the form of word fragments. Table 2 below is results before and after going through the process tokenization.

TABLE II
DATA TOKENIZATION RESULTS

sebelum	sesudah
Sangat membantu kebutuhan orang orang yang akan belanja karena sangat murah dan mudah	[‘Sangat’ , ’membantu’, ‘kebutuhan’, ‘orang’, ‘orang’, ‘yang’, ‘akan’, belanja karena ‘sangat’, ‘murah’, ‘dan’, ‘mudah’]

3) Folding case

In the case folding stage this will change The letters contained in the review data are lowercase or lowercase so that they are easier to read by computers

4) Filtering or Remove Stopwords

In the filtering stage, we will use the NLTK library in Indonesian to make it easier to remove sentences. In this research, researchers also added several words that often appear in review data such as "yang", "tidak", "untuk" and others in order to reduce the occurrence of noise and make the data cleaner.

5) Stemming

At this stemming stage we will use stemmer factory library to make the process easier. The research objective of adding a stemming stage is to be able to make the base word of each word in the review data and remove affixes that appear at the beginning, insertion or suffix of the review words. Table 3 below is a comparison before and after the data passes through the case stage folding, filtering, and stemming.

TABLE III
RESULT OF CASE FOLDING, FILTERING, AND STEMMING

Before	after
Sangat membantu kebutuhan orang orang yang akan belanja karena sangat murah dan mudah	bantu butuh belanja murah mudah

D. Translate

After the review data has gone through the preprocessing stage, the next step is to carry out the translation process of the review data using the translator library. Table 4 shows the results before and after going through the translation stage.

TABLE IV
RESULT TRANSLATE

Before	after
aplikasi shopeenya bagus	the shopee app is good
cocok bantu shopee	it's good to help the shopee,
mantap gratis ongkir	it's free shipping

E. Sentiment Analysis or Classification

After the translation stage is carried out on the clean review data, the next step is the classification stage. In this research, the classification stage was carried out using the Naive Bayes Classifier method.

Naive Bayes algorithm classification process The classifier here is to determine a sentences as a set of positive, neutral, or negative based on the larger value of the probability calculation from the Bayes formula. If the probability of the sentence being classified as positive is greater than for negative classification, then the sentence is included in the positive classification. If the sentence's probability of positive classification is the same as negative classification, then it is included in the neutral classification. While the opportunity for positive classification is smaller than negative classification, the sentence is included in

negative classification [14]. Table 5 below is the sentiment results from the Naive Bayes Classifier method.

TABLE V
SENTIMENT RESULTS WITH NAIVE BAYES CLASSIFIER

ulasan	klasifikasi	Klasifikasi Naive Bayes
the bug is really bad, i swear, the discount is like a long one, it's really stingi	Negatif	Negatif
i really like the application because i see good stuff, it's funny, it's funny, it's really good	Positif	positif

The composition of the classification results using Naive Bayes, obtained the following data with the results shown in Table 6.

TABLE VI
CLASSIFICATION RESULTS USING NAIVE BAYES

Sentimen	Jumlah Ulasan
Positif	2291
Netral	1650
Negatif	925

After crawling, preprocessing, automatic labeling with TextBlob, and classification with Naïve Bayes, the next step is to measure accuracy, precision, recall and F1_Score using the Sklearn library in the Python programming language with testing data and training data. The test results using Naïve Bayes are shown in Figure 3.

	precision	recall	f1-score	support
Negatif	0.78	0.64	0.70	1513
Netral	0.83	0.87	0.85	1209
Positif	0.77	0.86	0.81	2127
accuracy			0.79	4849
macro avg	0.80	0.79	0.79	4849
weighted avg	0.79	0.79	0.79	4849

Fig. 3 Test Result using Naive Bayes

After all stages have been carried out, next is the visualization stage. The output from the visualization stage in this research is in the form of a wordcloud. which will display images in the form of words that often appear in each sentiment. Figure 4 and Figure 5 are wordcloud displays of positive and negative sentients:

Fig. 4 Wordcloud Positive Sentiment Naive Bayes Classifier

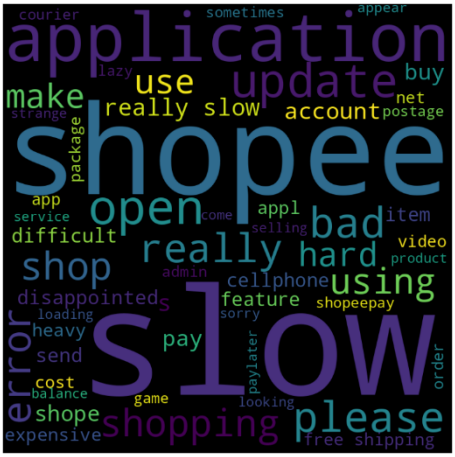


Fig. 5 Wordcloud Negative Sentiment Naive Bayes Classifier

F. Method Testing

At the metric performance testing stage, testing is carried out using the confusion matrix method, where researchers will compare the classification results carried out manually with the classification results that have been produced by the model using the Naive Bayes Classifier method. Table 7 is the results of classification carried out manually which will be compared with the Naive Bayes method. The Sklearn library is used to assist the process of displaying a classification report from this comparison. In building this machine learning model, researchers use a measure or what is usually called a metric to find out how well the model that has been created performs. Precision, recall, f1-score, macro avg, and weighted avg are used for evaluation metrics in this classification model. This classification model is used because it can show how the model that has been created can make decisions in the real world

TABLE VII
CLASSIFICATION RESULTS

ulasan	klasifikasi
the bug is really bad, i swear, the discount is like a long one, it's really stingi	Negatif
i really like the application because i see good stuff, it's funny, it's funny, it's really good	Positif

Testing the accuracy results in this research This is done by comparing classifications actually with the resulting classification results by models. Testing is carried out in this way displays the classification report, namely calculating accuracy, precision, recall, f1-score, macro avg, and weighted avg. Table 8 is the result of the multiclass confusion matrix from the naive Bayes classifier method which is the result of opinion data which has been converted into numerical data and table 9 is the result of the classification report of the naive Bayes classifier method

TABLE VIII
MULTICLASS CONFUSION MATRIX NAIVE BAYES CCLASSIFIER RESULTS

Klasifikasi	TP *(True Positif)*	FP *(False Positif)*	FN *(False Negatif)*
Positif	1819	529	308
Netral	1047	212	162
Negatif	970	272	543

TABLE IX
CLASSIFICATION REPORT RESULTS OF THE NAIVE BAYES CLASSIFIER METHOD

	Precision	Recall	FI-Score	Support
Positif	0,77	0,86	0,81	2127
Netral	0,83	0,87	0,85	1209
Negatif	0,78	0,64	0,70	1513
Accuracy	0,79	0,79	0,79	4849
Macro avg	0,80	0,79	0,79	4849
Weighted avg	0,79	0,79	0,79	4849

Based on the results of the classification report of the naive Bayes classifier method, if converted into percent form, an accuracy value of 79% has been obtained, the precision, recall and f1-score values for positive sentiment are respectively 77%, 86% and 81% with support of 2127. while neutral sentiment is with values of 83%, 87%, and 85% with support of 1209, and for negative sentiment it is with values of 78%, 64%, and 70% with support of 1513. From the data above, the micro average for precision is obtained. 80%, recall 79%, f1-score 79%, and support 4849, then for weighted average for precision 79%, recall 79%, f1-score 79%, and support 4849.

IV. CONCLUSION

This research is a sentiment analysis regarding Shopee E-commerce user reviews obtained from the Goggle Play Store using the naive Bayes classifier method. Scraping data from Google Play Store is stored and then classified into three polarities, namely positive, neutral and negative. From a total of 4902 review data obtained, after carrying out the preprocessing, translation, and then entering the Naive Bayes classifier method, the total data obtained was 4849 review data with the criteria of 2348 positive reviews, 1259 neutral reviews and 1242 negative reviews. Results the performance accuracy of the naive Bayes classifier method is 79%.

REFERENCES

[1] S. Wahyu Handani, D. Intan Surya Saputra, Hasirun, R. Mega Arino, and G. Fiza Asyrofi Ramadhan, "Sentiment analysis for go-jek on google play store," J. Phys. Conf. Ser., vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012032.

[2] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen spada Ulasan Produk Menggunakan Metode Naive Bayes," J. Edukasi dan Penelit. Inform., vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.

[3] Indonesiabaik. Pengguna Internet di Indonesia Makin Tinggi. https://indonesiabaik.id/infografis/pengguna-internet-di-indonesia-makin-tinggi. 2023. Diakses pada tanggal 10 Januari 2023

[4] Indonesiainside. Riset Google: Nilai Ekonomi Internet RI Tumbuh Tercepat di ASEAN. 2019. https://indonesiainside.id/ekonomi/2019/10/13/riset-google-nilai-ekonomi-internet-ri-tumbuh-tercepat-di-asean Diakses pada tanggal 10 Januari 2023

[5] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," CESS (Journal Comput. Eng. Syst. Sci., vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.

[6] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," J. Phys. Conf. Ser., vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012043.

[7] Rahmatulloh, RN Shofa, Darmawan. Sentiment Analysis of Ojek Online User Satisfaction Based on the Naïve Bayes and Net Brand Reputation Method. 9th International Conference on Information and Communication Technology (ICoICT). 2021.

[8] R Akbar, RN Shofa, MI Paripurna. The implementation of Naïve Bayes algorithm for classifying tweets containing hate speech with political motive. The 4th the International Conference on Sustainable Engineering and Creative Computing (ICSECC). 2019.

[9] A. Rahman, E. Utami, and S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," J. Komtika (Komputasi dan Inform., vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.

[10] D. Dwimarcahyani, T. Badriyah and T. Karlita, "Classification On Category Of Public Responses On Television Program Using Naive Bayes Method," *2019 International Electronics Symposium (IES)*, Surabaya, Indonesia, 2019, pp. 225-231, doi: 10.1109/ELECSYM.2019.8901576.

[11] F. S. Jumeilah, "Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes," *JSINBIS (Jurnal Sistem Informasi Bisnis)*, vol. 8, no. 1, pp. 92-98, Apr. 2018. https://doi.org/10.21456/vol8iss1pp92-98

[12] V.A Fitri, R. Andreswari, M.A. Hasibuan Sentiment Analysis of Social Media Twitter with Case of AntiLGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm DOI:10.1016/j.procs.2019.11.181

[13] A. Harun and D. P. Ananda, "Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decission Tree Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decission Tree," Indones. J. Mach. Learn. Comput. Sci., vol. 1, no. April, pp. 58–63, 2021

[14] P. P. E. Indarbensyah, And N. Rochmawati, "Penerapan *N-Gram* menggunakan Algoritma *Random Forest* dan *Naive Bayes Classifier* pada Analisis Sentimen Kebijakan PPKM 2021," *J. of Informatics and Computer Science*, Vol. 2, No. 4, 2021

[15] D. A. Muthia, "Analisis Sentimen Pada Review Buku Menggunakan Algoritma Naive Bayes," *J. Paradigma*, Vol. 16, No. 1, 2014.

[16] Adhi dan Eri. Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier. 2018.

[17] Chakra A. S., , Gupta u., & Kumar, P. A. Analysing Stock Market. Movement Using Twitter Sentiment Analysis And Time Series Forecasting. 2018.

[18] Fauziah Afshoh. Analisa Sentimen menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial

[19] Gunawan, Fauzi, & Adikara. Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile). 2017.

[20] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari. "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifie. 2016.

[21] Suryadi, Andri; Harahap, Erwin. Sistem Rekomendasi Penerimaan Mahasiswa Baru Menggunakan Naive Bayes Classifier di Institut Pendidikan Indonesia. 2018.

[22] D Pramita, R Saptono, R Anggrainingsih. Paramita Dwi, Saptono. ACADEMIC ARTICLES CLASSIFICATION USING NAIVE BAYES CLASSIFIER (NBC) METHOD. ITSMART. Jurnal Ilmiah Teknologi dan Informasi. 2018.

[23] AP Wijaya, HA Santoso. Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government. Journal of Applied Intelligent System, Vol.1, No. 1, 2016

[24] NC Siregar, RRA Siregar, MYD Sudirman. Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ). Jurnal Teknologia Aliansi Perguruan Tinggi (APERTI) BUMN. 2020

[25] SK Wardani, YA Sari, Indriati. Analisis Sentimen menggunakan Metode Naive Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding. Jurnal Pengembangan Teknologi dan Ilmu Komputer Universitas Brawijaya. 2021.