Jurnal Teknologi Informasi dan Komunikasi

Vol: 15 No 01 2024 **E-ISSN: 2477-3255**
Accepted by the Editor: 06-04-2024 | Revision: 21-04-2024 | Published: 31-05-2024

Application of the Naïve Bayes Algorithm in Sentiment Analysis of Using the Shopee Application on the Play Store

Rina Afriani Sitorus[1], Ilka Zufria[2]
[1,2]Program Studi Ilmu Komputer Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara
[1,2]Jl. Lapangan Golf, Kp. Tengah, Kec. Pancur Batu, Kab. Deli Serdang
e-mail: [1]rina.afrianisitorus@uinsu.ac.id, [2]ilkazufria@uinsu.ac.id

Abstract

This research aims to find out the opinions of users of the Shopee application on the Play Store using the Naive Bayes Naive Algorithm and to find out the suitability of the correct application of the Naive Bayes algorithm in carrying out sentiment analysis with the classification of three sentiment classes. The dataset used in this study consisted of 2000 customer reviews obtained from the Play Store in 2024 collected by the scraping process using the Python library. The dataset has 1,198 examples of negative attitudes, 583 examples of good sentiment, and 219 examples of neutral sentiment. The results of this study are expected to be used as evaluation material for Shopee Apilkation to improve the performance of Shopee applications. Research findings show that the Bayes naive approach reaches accuracy determined by various aspects, such as the quantity of data collections and positive and negative data distribution. This study shows that the Bayes naive algorithm can function properly as a technique to evaluate user sentiment for applications in the Play Store. However, with the classification of three classes, another algorithm is needed to produce higher accuracy.

Keywords— *Sentiment Analysis, Naïve Bayes algorithm, Shopee, Play Store, User Reviews*

1. Introduction

Indonesia is one of the countries with the most smartphone users in the world. A population of 250 million people is a huge market for growing smartphone users. The Ministry of Research, Technology and Higher Education (Menristekdikti) stated that the number of smartphone users in Indonesia has now reached 25% of the total population. With so many smartphone users, Indonesia is also one of the countries that has very fast e-commerce growth. Most Indonesians use e-commerce to carry out online shopping transactions [1].

Many e-commerce applications are present in the Indonesian economy. It goes without saying that people select the best e-commerce application when they shop online. One of the most widely used apps among Indonesians is Shopee. To find out the advantages and disadvantages of the Shopee application and what features are not optimal according to user comments, sentiment analysis of previous user comments regarding the Shopee application is necessary, which cannot be judged solely based on views. From the results of this analysis,

users will know the percentage of positive, negative and neutral comments from previous users about the Shopee application. Apart from that, the results of this analysis will also be a benchmark for Shopee management to develop Shopee application features and pay attention to complaints from users.

The study "Implementation of the Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on the Google Play Store" was conducted previously Agustina, et al analyzed the sentiment of customer comments regarding the use of the Shopee application through 1000 comment data contained in the Play Store by the Naive Bayes algorithm and obtained sufficient accuracy results high [2].

Different from previous research, this research uses more recent comment data and also looks at improvements made by the Shopee application following the results of previous research. In order to confirm that the Naive Bayes technique is appropriate for sentiment analysis, 2000 data points from comments on the Play Store will be used in this study's sentiment analysis of user comments on the Shopee application as a whole. When compared with previous research, this research will provide better results regarding sentiment analysis of users of the Shopee application because in this research sentiment analysis was carried out by dividing user comments into three classes, namely positive class, negative class and neutral class sentiment.

2. Research Methodology

This research uses quantitative research methodology, which is a type of research that is mathematical in nature and has a structure with clear stages, and also aims to describe an event [3]. The following are the research stages.

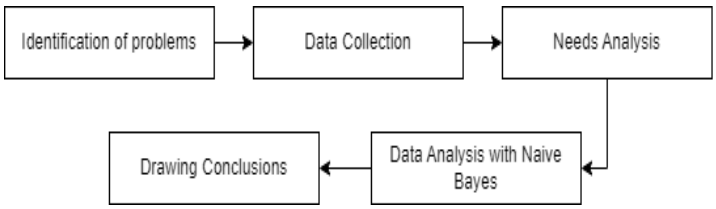

Figure 1. Reseacrh Stages

Information :

1. Determine the issue. Many people who use the Shopee application post their thoughts about how well it works in the Play Store comments section. The large number of opinions makes it difficult to conclude whether the Shopee application is the best e-commerce application or not. For this reason, sentiment analysis is needed to classify user opinions regarding the Shopee application which will produce a percentage of positive, negative and neutral sentiment.
2. Data Collection. To collect data in this research, field research and a literature review were carried out. Field research was carried out using the observation method or observing the comments column of the Shopee application on the Play Store. Then 2000 comment data were collected using scraping techniques using the Python programming language. In this research, a literature review was also carried out by summarizing the contents of accredited journals, reading library books and theses and observing previous research.
3. Needs Analysis. In the system that will be built, data will be collected using the scraping technique using google-play-scraper. In this system, a sentiment analysis process will be carried out to classify the opinions of Shopee application users using the Naïve Bayes algorithm.

https://doi.org/10.31849/digitalzone.v15i1.19828

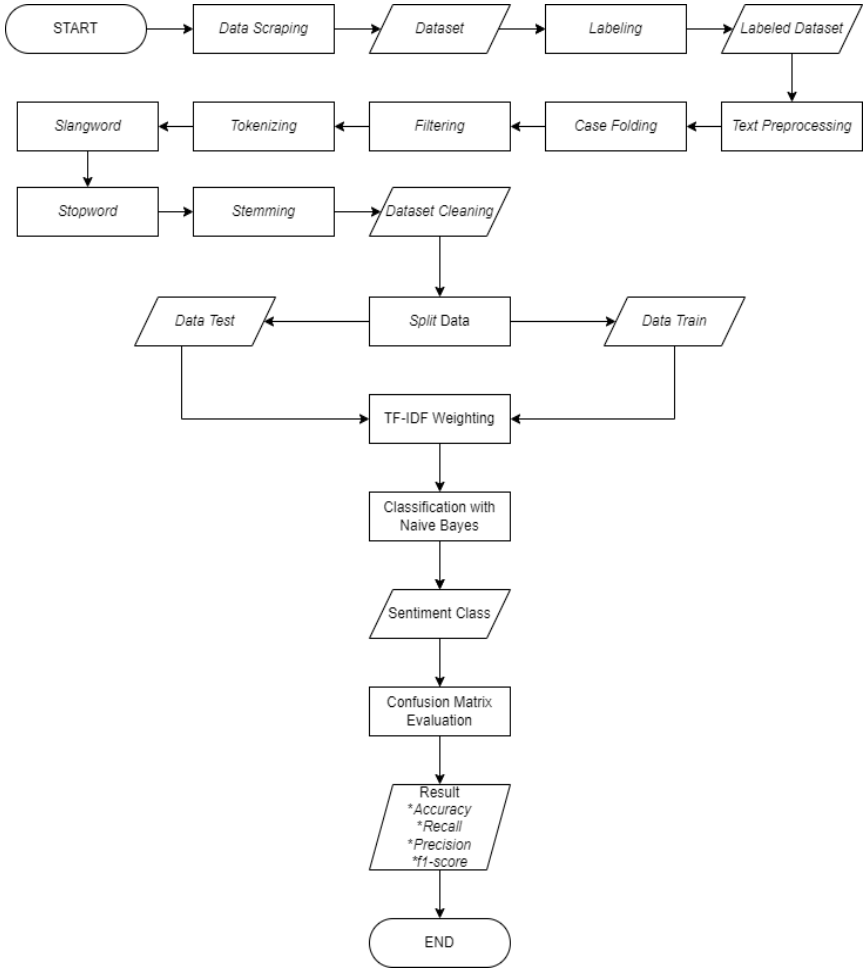

Figure 2. System Flowchart

4. Data Analysis. This study will use the Naïve Bayes Classifier method for sentiment analysis and TF-IDF for word weighting. Before carrying out sentiment classification using the Naïve Bayes Classifier algorithm, the data will go through a data cleaning process or what is called data preprocessing, then word weighting will be carried out using TF-IDF and then classified using the Naïve Bayes Classifier algorithm. Then, from the classification results, calculations will be carried out using a confusion matrix. The results of the confusion matrix calculation are in the form of percentage values for accuracy, recall, f1-score and precision.
5. Drawing conclusions. Making inferences from the analysis is the next step after the data has been examined and categorized using the Naïve Bayes Classifier. The analysis in this research will produce percentage values of accuracy, recall, f-1 score from positive and negative sentiment.

2.1 Sentiment Analysis

Sentiment analysis is a method for classifying material into sentences and identifying sentiment. These feelings fall into three categories: neutral, positive, and negative [4]. In sentiment analysis, data mining is carried out to analyze, process and extract data in the form of text in an object, such as a service, product, topic or certain event [5]. It may be inferred that sentiment analysis is a procedure for identifying a person's sentiments that are conveyed in text form and can be classified into positive and negative sentiments. Sentiment analysis is one of the things that is needed in filtering opinions on the internet [6].

2.2 Text Mining

Text Mining is a text analysis method carried out by a computer automatically to obtain quality information from the text structure contained in a document. When utilizing techniques like group analysis, association, and classification to examine the relationships between documents, text mining extracts words that can be considered samples from the texts [7].

2.3 Term Frequency-Inverse Documen Frequency (TF-IDF)

The TF-IDF word weighting approach, which assigns a value to each word in the data to determine the term's frequency of occurrence, is highly effective at determining the importance of a word in a document. This value is then multiplied by the inverse of the document frequency. TF is considered to have importance according to the total occurrence of words in the text and IDF is a token weighting that is used to observe the occurrence of tokens in the text or document [8].

2.4 Naïve Bayes Algorithm

Using an existing dataset's frequencies and combinations of values, the Naïve Bayes classification algorithm adds them up to determine probability. This approach makes use of the Bayes theorem, which makes the assumption that none of the dataset's qualities are dependent on the class variable's value [9]. Using opportunities from the data, the Naïve Bayes method is a data mining/machine learning technique that groups and classifies data into many classes [10]. Using a variety of sentiment analysis techniques, Naïve Bayes classification aims to disseminate user feedback findings [11].

2.5 Confussion Matrix

Confusion Matrix is a technique that can be used to evaluate analysis results. Confusion Matrix is frequently used to assess a classification method's effectiveness and derive conclusions from the conducted study.

3. Result and Discussion

3.1 Data Analysis

The method used to collect data is the process of scraping comment data on Playstore using google-play-scraper which is a method for retrieving data from Playstore using the Python programming language[12]. 2000 data points were gathered during the scraping operation, representing customer reviews of the Shopee app on the Playstore in 2024.

Table 1 Scraping Data

Content	*Score*
Lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua aneh!!jd males blenje.pdhl.sering belanja dsni!!!	2
Semenjak kenal sama Shopee, belum pernah dikecewain,,,, semakin besar minat saya untuk selalu belanja di Shopee	5
Tolong performa shopee di perbaiki, buka gambar, scroll yg lain sangat lambat sekali,	3

3.2 Sentiment Labeling

Sentiment labeling is done manually, where the labeling process utilizes human reasoning in determining it. In this research, positive, negative and neutral labels were given manually based on the meanings of words that researchers studied in the book Bahasa Language Scholars by Erwan Juhara, et al. In this book, linguists interpret the meaning of a word based on the feeling contained in the word. For example, a sentence is labeled "positive" if the content praises a product that contains a pleasant taste, "negative" if the content contains criticism and

https://doi.org/10.31849/digitalzone.v15i1.19828

tends to be dismissive, and "neutral" if it is judged to have no taste. The following is the distribution of sentiment labels from the entire dataset used.

Table 2. Data Labeling

Content	Label
Lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua aneh!!jd males blenje.pdhl.sering belanja dsni!!!	Negatif
Semenjak kenal sama Shopee, belum pernah dikecewain,,,, semakin besar minat saya untuk selalu belanja di Shopee	Positif
Tolong performa shopee di perbaiki, buka gambar, scroll yg lain sangat lambat sekali,	Netral

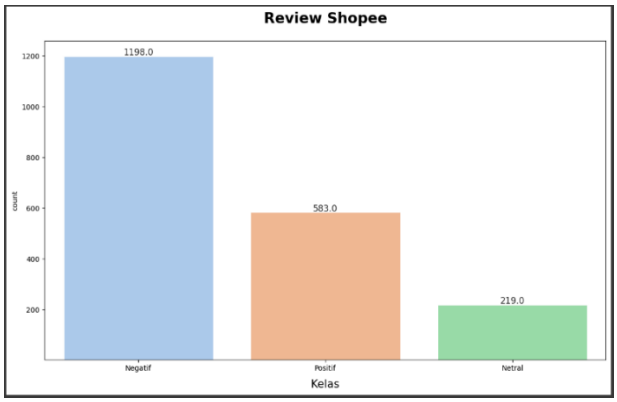

Figure 3. Distribution of Sentiment Data

3.3 Data Preprocessing

Data preprocessing is the initial stage in the text classification process to prepare text data before processing. In the preprocessing stage, the data will be converted into a better form and will produce information that is ready to be used in the next process [13]. The goal of this stage is to improve data quality [14]. At this stage the data will go through several data preprocessing processes in the form of case folding, filtering, tokenizing, slangwords, stopwords, and stemming.

a. Case Folding

The technique of changing text in review data to lowercase is known as "case folding" [15]. The following are the results of case folding.

Table 3. Case Folding

Content	Result of Case Folding
Lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua aneh!!jd males blenje.pdhl.sering belanja dsni!!!	lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua aneh!!jd males blenje.pdhl.sering belanja dsni!!!

b. Filtering

Filtering is the process of eliminating data that is irrelevant or not needed for the analysis being carried out. Such as special signs, emojis, numbers and punctuation [16]. The following are the results of filtering.

Table 4. Filtering

Content	Result of Filtering
lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua aneh!!jd males blenje.pdhl.sering belanja dsni!!!	lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua anehjd males blenjepdhlsering belanja dsni

c. Tokenizing

Tokenizing is the process of breaking text or text data into smaller units [17] such as words, phrases, or other tokens. Tokenization is an important step in text analysis. For example, whitespace characters, such as enter, tabulate, space are considered as word separators. The following are the results of tokenizing.

Table 5. Tokenizing

Content	Result of Tokenizing
lg ga ada free ong gila mahal bgt deh dan masa harus pakai kurir instant semua anehjd males blenjepdhlsering belanja dsni	['lg', 'ga', 'ada', 'free', 'ong', 'gila', 'mahal', 'bgt', 'deh', 'dan', 'masa', 'harus', 'pakai', 'kurir', 'instant', 'semua', 'anehjd', 'males', 'blenjepdhlsering', 'belanja', 'dsni']

d. Slang Word

The process of converting slang words (slang or informal words) is a step in pre-processing text data which aims to convert slang words into standard forms that are more commonly used in more formal or official language. This is important in text analysis and natural language processing to ensure consistency and accuracy in data processing.

Table 6. Slang Word

Content	Slangword
['lg', 'ga', 'ada', 'free', 'ong', 'gila', 'mahal', 'bgt', 'deh', 'dan', 'masa', 'harus', 'pakai', 'kurir', 'instant', 'semua', 'anehjd', 'males', 'blenjepdhlsering', 'belanja', 'dsni']	['lagi', 'enggak', 'ada', 'free', 'ong', 'gila', 'mahal', 'banget', 'deh', 'dan', 'masa', 'harus', 'pakai', 'kurir', 'instant', 'semua', 'anehjd', 'malas', 'blenjepdhlsering', 'belanja', 'disini']

e. Stopword Removal

The stopword process in text pre-processing is a step to remove words that are common and do not provide important information in text analysis. These words are called stopwords and usually do not make a significant contribution to understanding the meaning or patterns in the text.

Table 7. Stopword Removal

Content	Stopword
['lagi', 'enggak', 'ada', 'free', 'ong', 'gila', 'mahal', 'banget', 'deh', 'dan', 'masa', 'harus', 'pakai', 'kurir', 'instant', 'semua', 'anehjd', 'malas', 'blenjepdhlsering', 'belanja', 'disini']	['free', 'ong', 'gila', 'mahal', 'pakai', 'kurir', 'instant', 'anehjd', 'malas', 'blenjepdhlsering', 'belanja']

f. Stemming

The stemming stage is the stage of discovering the root word for each filtered word. At this step, words are cut in the form of prefixes, insertions and word endings to make fundamental words. The purpose of the stemming process is to limit the morphological variances of words in a text such that words that have the same root may be identified as the same entity.

Table 8. Stemming

Content	Stemming
['free', 'ong', 'gila', 'mahal', 'pakai', 'kurir', 'instant', 'anehjd', 'malas', 'blenjepdhlsering', 'belanja']	['free', 'ong', 'gila', 'mahal', 'pakai', 'kurir', 'instant', 'anehjd', 'malas', 'blenjepdhlsering', 'belanja']

3.4 Word Weighting

Following sentiment cleaning and labeling comes TF-IDF (Term Frequency-Inverse Document Frequency) weighting. This technique multiplies the word frequency (term frequency) by the inverse of the document frequency (inverse document frequency) to determine the weighting of words (terms) in document data.

An illustration of how to calculate the TF-IDF value is shown below. The training data's TF values are listed below.

Table 9. TF Value

Term	TF					DF
	D1	D2	D3	D4	D5	
Bantu	1	0	0	0	0	1
Keranjang	1	0	1	0	0	2
Belanja	1	1	0	0	0	2
Enak	0	1	0	0	0	1
Cepat	0	0	1	0	0	1
Bayar	0	0	1	0	0	1
Pay	0	0	1	0	0	1
Bisa	0	0	1	0	0	1
Banyak	0	0	0	1	1	2
Iklan	0	0	0	3	1	2
Aplikasi	0	0	0	1	1	2
Jual	0	0	0	1	0	1
Busuk	0	0	0	0	1	1
Ganggu	0	0	0	0	1	1

After getting the word frequency value (term frequency), the next step is to determine the IDF (Inverse Document Frequency) value for each word. The following is the formula for calculating the IDF value of these words.

$$IDF = \ln\left(\frac{D+1}{df+1}\right)+1 \quad (1)$$

Table 10. IDF Value

Term	DF	IDF
Bantu	1	2.098
Keranjang	2	1.693
Belanja	2	1.693
Enak	1	2.098

Term	DF	IDF
Cepat	1	2.098
Bayar	1	2.098
Pay	1	2.098
Bisa	1	2.098
Banyak	2	1.693
Iklan	2	1.693
Aplikasi	2	1.693
Jual	1	2.098
Busuk	1	2.098
Ganggu	1	2.098

The TF and IDF values must then be normalized in order to obtain the same interval for every data set. The following equation was applied to normalize the data.

$$W = TF\ x\ IDF \quad (2)$$

Table 11. TF-IDF Value

Term	TF-IDF				
	D1	**D2**	**D3**	**D4**	**D5**
Bantu	2.098	0	0	0	0
Keranjang	1.693	0	1.693	0	0
Belanja	1.693	1.693	0	0	0
Enak	0	2.098	0	0	0
Cepat	0	0	2.098	0	0
Bayar	0	0	2.098	0	0
Pay	0	0	2.098	0	0
Bisa	0	0	2.098	0	0
Banyak	0	0	0	1.693	1.693
Iklan	0	0	0	5.079	1.693
Aplikasi	0	0	0	1.693	1.693
Jual	0	0	0	2.098	0

https://doi.org/10.31849/digitalzone.v15i1.19828

Term	TF-IDF				
	D1	**D2**	**D3**	**D4**	**D5**
Busuk	0	0	0	0	2.098
Ganggu	0	0	0	0	2.098

The TF-IDF values must then be normalized in order to get the same interval for every set of data. The following equation was applied to standardize the data.

$$TF_{norm}(t,d) = \frac{TF(t,d)}{\sqrt{\sum_i (TF(t,d))^2}} \quad (3)$$

Table 12. Data Normalization Result

D1	D2	D3	D4	D5
0.2197	0	0	0	0
0.1773	0	0.1773	0	0
0. 1773	0.1773	0	0	0
0	0.2197	0	0	0
0	0	0.2197	0	0
0	0	0.2197	0	0
0	0	0.2197	0	0
0	0	0.2197	0	0
0	0	0	0.1773	0.1773
0	0	0	0.53182	0.1773
0	0	0	0.1773	0.1773
0	0	0	0.2197	0
0	0	0	0	0.2197
0	0	0	0	0.2197

3.5 Naïve Bayes Classification

Using the Naïve Bayes Algorithm, the initial dataset for the sentiment analysis study on Shopee in the Google Play Store has 2000 data points. The training set in this study comprises 1400 data points, whereas the test set has 600 data points. A 7:3 training-to-test data ratio is what is meant by this. The TF-IDF weighting technique will be used to produce a weight for each word in each class during the training phase. The following method is used to choose one test data point: five previously gathered sample data points are used as training data:

Table 13. Test Data Sample

Test Sentiment

['belanja', 'aman', 'bisa']

Sentiment is determined by estimating the document's likelihood in the test data while accounting for the word likelihood in the training data. The Naïve Bayes algorithm is used to automatically classify sentiment. This method utilizes the MultinomialNB function to compare the magnitude of each word in the test data with the terms found in the training data. Consequently, every training document will have a calculated total of the probabilities assigned to positive and negative terms, with each probability being given a numerical value. Density. Afterwards, The test document's weights are assessed. If the weight assigned to positive likelihood is greater than the weight assigned to negative probability, the mood is categorized as positive. Conversely, when the weight given to negative probability is greater than the weight given to positive probability, the emotion is classified as negative. Preliminary probabilities, conditional probabilities, and posterior probabilities are computed before class categorization begins. The Naïve Bayes method can be employed to classify test data, and the subsequent steps can be utilized to divide the procedure:

a. Calculation of Prior Probability Values

$$P\,(Kelas\,|Sentimen\,) = \frac{Jumlah\ Kelas\ X}{Jumlah\ Sentimen} \quad (4)$$

By using this equation, the probability of each class in sentiment will be obtained.

1. P (Positive | Sentiment) $= \frac{2}{5} = 0.4$
2. P (Netral | Sentiment) $= \frac{1}{5} = 0.2$
3. P (Negative | Sentimen) $= \frac{2}{5} = 0.4$

b. Calculation of Conditional Probability Values

$$P\,(Term\,|Kelas\,) = \frac{Total\ Bobot\ TF-IDF\ Term\ pada\ Kelas + 1}{Bobot\ TF-IDF\ Kelas + Total\ Bobot\ TF-IDF} \quad (5)$$

By using this equation, the probability of each word in each sentiment class will be obtained

1. Positive
   P (belanja | Positive) $= \frac{0.3546+1}{1.8501+3.927} = \frac{1.354}{5.777} = 0.234$
   P (aman | Positive) $= \frac{0+1}{1.8501+3.927} = \frac{1}{5.777} = 0.173$
   P (bisa | Positive) $= \frac{0+1}{1.8501+3.927} = \frac{1}{5.777} = 0.173$
2. Netral
   P (belanja | Netral) $= \frac{0.3546+1}{1.0561+3.927} = \frac{1.354}{4.983} = 0.271$
   P (aman | Netral) $= \frac{0+1}{1.0561+3.927} = \frac{1}{4.983} = 0.0002$
   P (bisa | Netral) $= \frac{0.2197+1}{1.0561+3.927} = \frac{1.219}{4.983} = 0.\ 244$
3. Negative
   P (belanja | Negative) $= \frac{0+1}{2.077+3.927} = \frac{1}{6.004} = 0.167$

https://doi.org/10.31849/digitalzone.v15i1.19828

$$P \text{ (aman | Negative)} = \frac{0+1}{2.077+3.927} = \frac{1}{6.004} = 0.167$$

$$P \text{ (bisa | Negative)} = \frac{0+1}{2.077+3.927} = \frac{1}{6.004} = 0.167$$

c. Calculation of Posterior Probability Values

$$P\ (Sentimen \mid Kelas) = P_{term\ 1}\ x\ ..\ x\ P_{term\ n}\ x\ P\ (Kelas \mid Sentimen) \quad (6)$$

The posterior probability value for every class will be computed using the equation above.

P (Sentiment | Positive) = 0.234 x 0.173 x 0.173 x 0.4
=0.0028013544

P (Sentiment | Netral) = 0.271 x 0.0002 x 0.244 x 0.2
=0.00000264496

P (Sentiment | Negative) = 0.167 x 0.167 x 0.167 x 0.4
= 0.0018629852

Upon completing the calculations example, maximum value obtained from every classes and maximum number from the data of the test for Positive sentiment were selected, which amounted to 0.0028013544. The sentiment categorization outcome for the words "belanja, aman, senang" is Positive. Displayed below is a visual depiction of the data pertaining to each sentiment category.

Figure 4. Wordcloud Positive

Figure 5. Wordcloud Negative

Figure 6. Wordcloud Netral

3.6 Test Result

The purpose of this technique is to evaluate the built system's sentiment analysis performance for the Shopee app available on the Play Store. The research utilizes Google Collab and Python as the programming language. The outcomes of sentiment analysis will be displayed as a confusion matrix, which aids in evaluating the research's accuracy level.

a. Evaluation of Result

After testing sentiment by the Naïve Bayes algorithm, classification of sentiment results will be obtained in the form of sentiment labels. The labels resulting from this classification will be compared with the actual labels to determine the accuracy, precision, recall and F1 value of the model for the dataset used.

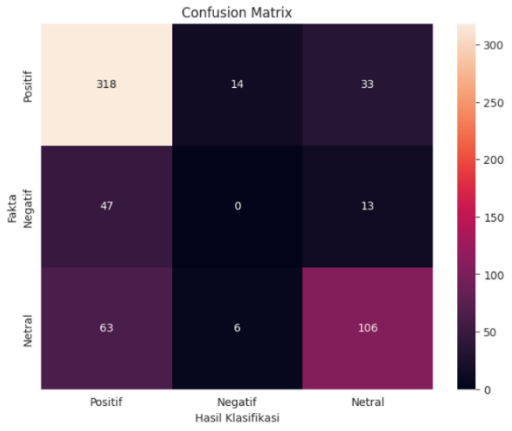

Figure 7. Confusion Matrix

In Figure 7, the accuracy, precision, recall and f1-score values can be calculated using the following equation.

$$Accuracy = \frac{318+0+106}{318+14+33+47+0+13+63+6+106} \text{ x } 100\% = 71\%$$

$$Precision = \frac{318}{318+47+63} \text{ x } 100\% = 74\%$$

$$Recall = \frac{318}{318+14+33} \text{ x } 100\% = 87\%$$

$$F1\text{-}score = \frac{2 \text{ x } 74 \text{ x } 87}{74 + 87} \text{ x } 100\% = 80\%$$

Of the 600 data used as test data in this research, it can be seen from the confusion matrix above that 365 data are in the form of positive comments, 60 data are negative

comments, and 175 data are neutral data. Based on the calculations carried out above, evidently the number of test data is 168 data and the accuracy value is 71%, precision 74%, recall 87%, f1-score 80%.

Through the analysis carried out in this research, it can be seen that users' views regarding the Shopee application from data obtained from the end of 2023 to 2024 are mostly negative comments. Shopee can use this to improve application performance by paying attention to user complaints.

In previous research, sentiment analysis for using the Shopee application with data collected through comments in the Play Store was carried out. Using a thousand comments, Positive and negative data were the two classes into which the researchers divided the data.In order to generate better study, the amount of data was then increased to 2000, the most recent comment data, and it was also classified using three sentiment classes: positive, negative, and neutral.

4. Conclusion

utilizing test data for the Naïve Bayes algorithm's categorization based on the outcomes of model training completed prior to testing. The model obtained during the training process will influence the classification results during testing. The results of the analysis carried out from a total of 2000 data which are evaluations of the Shopee app in the Play Store in 2024 show a distribution of 1198 negative sentiments, 583 positive sentiments and 219 neutral sentiments. Sentiment analysis regarding the use of the Shopee application in the Play Store which was completed via the Naïve Bayes technique obtained results of 71% accuracy, 74% precision, 87% recall and 80% f1-score.

References

[1] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[2] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.

[3] A. Fauzi *et al.*, *Metodologi Penelitian*, I. Purwokerto: CV Pena Persada, 2022.

[4] M. S. Hasibuan, Suhardi, and T. A. B. Sembiring, "Analisis Sentimen Kebijakan Vaksin Covid-19 Menggunakan SVM dan C4.5," Universitas Islam Negeri Sumatera Utara, 2021. [Online]. Available: http://repository.uinsu.ac.id/18678/1/Laporan Akhir Penelitian B5 TTD.pdf

[5] Samsir, Ambiyar, U. Verawardana, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.

[6] Y. Julianto, D. H. Setiabudi, and S. Rostianingsih, "Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine," *J. Infra*, vol. 10, no. 1, pp. 1–7, 2022.

[7] M. P. R. Putra and K. R. N. Wardani, "Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 5, no. 1, pp. 63–71, 2020, doi: 10.32767/jutim.v5i1.791.

[8] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.

[9] Farhan, Triase, and A. M. Harahap, "Penggunaan Algoritma Naive Bayes Dalam Text Mining Untuk Klasifikasi Pasal UU ITE," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 6, pp. 314–322, 2023.

[10] N. Faridhotul Hidayah, K. Paranita Kartika R., and S. Nur Budiman, "Penerapan Metode Naive Bayes Dalam Analisis Sentimen Aplikasi Sentuh Tanahku Pada Google Play," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 679–683, 2022, doi: 10.36040/jati.v6i2.5610.

[11] H. Setiawan and I. Zufria, "Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1003–1012, 2023,

[12] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[13] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.

[14] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.

[15] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.

[16] K. D. Indarwati and H. Februariyanti, "Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Gojek Menggunakan Metode Naive Bayes Classifier," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 1, 2023, doi: 10.35957/jatisi.v10i1.2643.

[17] P. P. O. Mahawardana, I. A. P. F. Imawati, and I. W. Dika, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," *J. Manaj. dan Teknol. Inf.*, vol. 12, no. 2, pp. 50–56, 2022, [Online]. Available: https://ojs.mahadewa.ac.id/index.php/jmti/article/view/2111http://dx.doi.org/10.24843/JTRTI.2022.v03.i01.p17