App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm

App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm

Dany Pratmanto[1*], Rousyati Rousyati[2], Fanny Fatma Wati[3], Andrian Eko Widodo[4], Suleman Suleman[5], Ragil Wijianto[6]

[1,4,6]Program Studi Sistem Informasi, STMIK Nusa Mandiri Jakarta, Indonesia
[2,3,5]Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Indonesia

E-mail: dany.dto@nusamandiri.ac.id

Abstract. An online marketplace site is a shopping place that is currently popular with the community because it offers a variety of convenience and one of the marketplace apps is Shopee. Some people are satisfied with the service provided by the Shopee app. But unisex some people who give complaints about this application. User-provided response to Shopee app in the Comments field of Shopee Google Play Store can be analyzed for negative and positive sentiments. This research aims to assist Shopee's management of the positive or negative opinions of application users and can provide empirical evidence for related theories so that it can be used as a donation of thought for the development of theories Next. With the number of reviews shown, you need an analysis that can classify these reviews into positive or negative classes. The method used for the sentiment analysis of Shopee app reviews is the Naive Bayes algorithm obtaining an accuracy yield of 96,667%

1. Introduction

The online marketplace shopping site or e-commerce is experiencing rapid development. One reason to grow the marketplace or e-commerce is because of the ease and comfort in the use and have a lot of advantages that can buy goods without having to come to the store so can buy anywhere without any restrictions and can be done for 24 hours. Also, another advantage of using the marketplace is that there are many discounts, promo and even free postage to the most popular choice of the community for shopping [1].

One of the most popular marketplace apps at the moment is Shopee. In 2020 the Indonesian shopee branch was ranked first among the most clicked e-commerce sites in Indonesia [2]. Shopee is an application engaged in buying and selling online and can be accessed easily by using a mobile phone. Shopee is present in the form of Moblie application that allows its users to do shopping activities online so that it can be accessed anytime and anywhere. Although there are many advantages in Shopee application, there are also some shortcomings that arise such as goods sold not by the picture, packaging or goods damaged during delivery, the seller sends the goods that are not suitable Booking transactions, and even many scams happening due to the many market things or e-commerce. From the benefits or disadvantages gained from the application Shopee then not very few people especially who give reviews through social media about something based on their opinion. To help management see the sentiment of the community, especially users of the Shopee app, there needs to be an analysis method that can summarize the reviews,

positive or negative. The sentiment classification aims to classify user reviews into positive or negative opinions.

In sentiment analysis, the classification is divided into three tiers, i.e. document level, sentence level, and aspect level. In this study, sentiment analysis took sentence rates, using the Naïve Bayes Classifier method. Sentiment analysis is referring to feelings, emotions, attitudes, and opinions. Because the textual data is used when there is a need to analyze the concept of expressing sentiment and calculating insights to explore the business. Sentiment analysis is a process for determining whether a review is expressed to be positive, negative or neutral. The mining opinion does not observe the topic of the text but is more focused on the expression described from the opinion text [1]. This determines the review or review with regards to the Shopee app on the Google Play Store including positive, negative or neutral opinions. Sentiment Analysis Research for Online Shopping Site Service Assessment Using Naïve Bayes Algorithm, data set taken from public tweets and re-tweets on Twitter e-commerce sites in Indonesia, namely Lazada, Bukalapak, Blibli, and Tokopedia. In this research using Naïve classifying machine algorithm Bayes Finally, we measure the performance of the classification engine using 10-fold cross-validation. Evaluation results show an average accuracy of 93.33% [1]. Based on research analysis of the level of popularity of e-commerce in Indonesia based on social media sentiments using the Naïve Bayes method with the percentage of the final result of the popularity of each e-commerce in Indonesia, namely e-commerce 1 = 91.2%, e-commerce 2 = 76.4%, e-commerce 3 = 37.8% E-commerce 4 = 33.8% where the percentage resulted from the difference between positive and negative sentiments [3].

This research will analyze issues related to some reviews of the Shopee app in the Google Play Store and determine the accuracy of the results of the sentiment analysis Algortma Naïve Bayes. The benefit of this research is to help Shopee management about the positive or negative opinions of the application user and can provide empirical evidence for related theories so that it can be used as a donation of thought to Development of subsequent theories.

2. Theoretical Basis

The Naive Bayes algorithm is a classification algorithm based on probabilities in statistics proposed by Thomas Bayes that predicts future opportunities based on past opportunities (Bayes theorem). This method is then combined with "naive" where the conditions between the attributes of each other are not related to each other [4].

Classifying Naïve Bayes is very simple, efficient and a popular machine learning technique for text classification, and has good performance in many domains. However, Naïve Bayes has the disadvantage of being very sensitive to too many features, which results in low classification accuracy [5]

The method used is the Naïve Bayes method which has stages including: data cleaning process (Data Cleaning), data integration (Data Integration), data selection (Data Selection), data transformation (Data Transformation), mining process, pattern evaluation (Pattern Evaluation), Knowledge Presentation [6].

Naïve Bayes classifier is a classification using simple statistics based on the Bayes theorem which assumes that the existence or absence of a class with other features. As a probability model, naïve Bayes classifier can be trained efficiently as supervised learning [7]. Partitioning is the separation of data sets carried out using the KNIME Partition node by applying a default random seed. The use of random seeds gives reproducible results after node re-execution [8].

GATA Framework is an alternative in Indonesian text pre-processing, the application also provides an application program interface (API) feature for sending data from external applications. Whereas the GATA framework is a framework based on the PHP programming language developed under the name MTG Framework in 2012 and changed its name to the GATA Framework in 2017. The GATA framework has been able to overcome various external

problems, namely usability, ability, response, security, presence, and reliability, as well as internal factors, namely ease of syntax or code that is easy to use and has used the Model View Controller (MVC) [9].

3. Research Methods

Data will be processed using the Knime Application. Text processing that researchers use Tokonize, Transform Case, Stopword (Dictionary). Tests carried out by selecting a Naive Bayes Selection selection to get the accuracy value. From the accuracy value of the confusion matrix model the proportion of the number of predicted values obtained and the ROC Curve so that we can see the ROS graph with the AUC (Area Under Curve) value. The experimental research method was conducted in research using the framework of thought proposed in the study as follows:

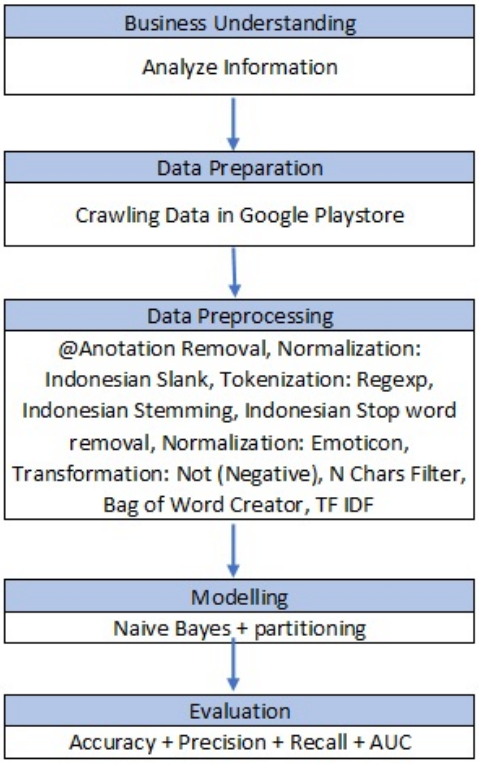

Figure 1. Framework for Thinking

Figure 1, explains that the framework in this study begins by analyzing information about e-commerce shopee then taking a review of the Shopee's application in PlayStore, from review the text processing is done using @Anotation Removal, Normalization: Indonesian Slank, Tokenization: Regexp, Indonesian Stemming, Indonesian Stop word removal, Normalization: Emoticons, Transformation: Not (Negative), N Chars Filters, Bag of Word Creators, TF IDF. After that, the modeling stage uses Naive Bayes + partitioning. Then evaluate the accuracy, precision, recall, and AUC.

4. Results and Discussion

Research conducted on sentiment analysis Review by using the Naive Bayes Algorithm in the Shopee application using the Knime Application in measuring the accuracy of experimental data.

To conduct the experiment the researcher used comment review data on the shopee application in the Play Store. The data taken were 200 review data consisting of 100 positive review data and 100 negative review data.

The results obtained from the data preprocessing stage using the GATA Framework and KNIME @Anotation Removal, Normalization: Indonesian Slank, Tokenization: Regexp, Indonesian Stemming, Indonesian Stop word removal, Normalization: Emoticon, Transformation: Not (Negative), N Chars Filter, Bag of Word Creator, TF IDF as follows:

Table 1. Text Processing using GATA Framework and KNIME

Text	@Anotation Removal	Tokenization: Regexp	Normalization: Indonesian Slank	Indonesian Stemming	Indonesian Stop word Removal
.Bagus banget @shopee . aplikasinya gue beli barang pagi sore dikirim besok aplikasinya bagus y	Bagus banget aplikasinya gue beli barang pagi sore dikirim besok aplikasinya bagus y	bagus banget aplikasinya gue beli barang pagi sore dikirim besok aplikasinya bagus y	bagus banget aplikasinya aku beli barang pagi sore dikirim besok aplikasinya bagus y	bagus banget aplikasi aku beli barang pagi sore dikirim besok aplikasi bagus y	bagus banget aplikasi beli barang pagi sore dikirim besok aplikasi bagus y
Text	@Anotation Removal	Tokenization: Regexp	Normalization: Indonesian Slank	Indonesian Stemming	Indonesian Stop word Removal
bagus banget aplikasi beli barang pagi sore dikirim besok aplikasi bagus y	bagus banget aplikasi beli barang pagi sore dikirim besok aplikasi bagus y	bagus banget aplikasi beli barang pagi sore dikirim besok aplikasi bagus	bagus[] banget[] aplikasi[] beli[] barang[] pagi[] sore[] besok[] aplikasi [] bagus	0,091 0,030 0,061 0,061 0,030 0,030 0,030 0,030 0,030 0,061	1,0 0,9 0,7 0,7 0,6 183,0 183,0 0,8 200,0 0,8

4.1. Determine the results of research using the Naive Bayes algorithm on Knime as follows
Modeling Naive Bayes uses Partitioning, which is an input table divided into two partitions. The first data partition, namely training, consists of 140 data, the second data partition, namely testing, consists of 60 data.

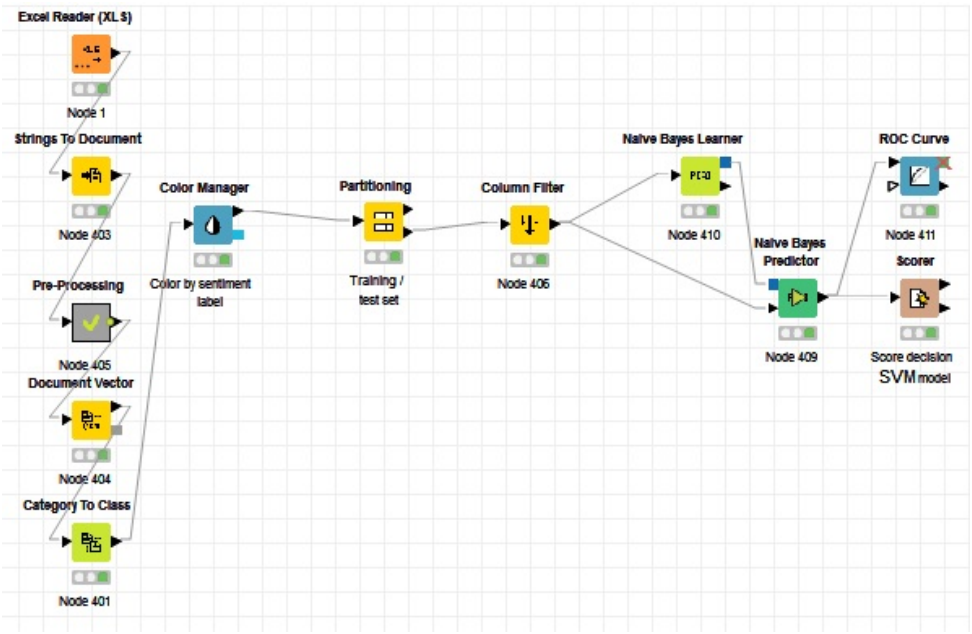

Figure 2. Modeling Naive Bayes with Partitioning

4.2. In this test we can get an accuracy value of 96,667% as shown in the figure below:
Figure 3 From 60 testing data (second partition) data on positive and negative reviews about the peer to peer lending application shopee application. Positive review data that fits predictions is 28 data. The negative review data included in the positive prediction are 2 data. Positive news data included in the negative prediction is 0 data and negative news data that fits the prediction is 30 data so that it can be seen that the accuracy value obtained is 96.667%, precision 100%, and recall 93.33%.

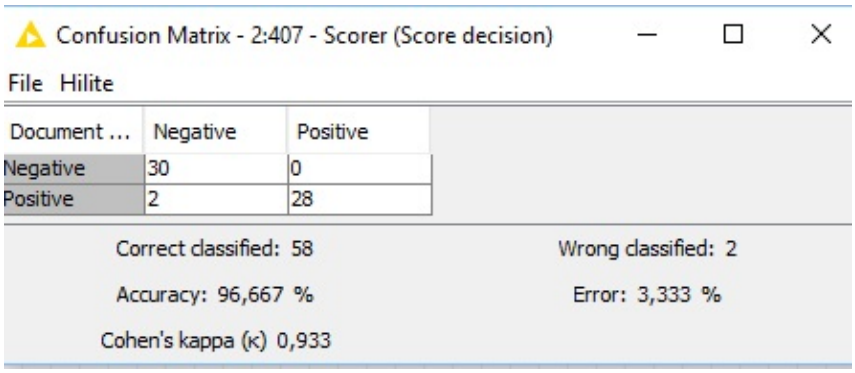

Figure 3. Accuracy Results

4.3. The ROC graph of sentiment analysis using the Naive Bayes algorithm + partitioning technique:
.

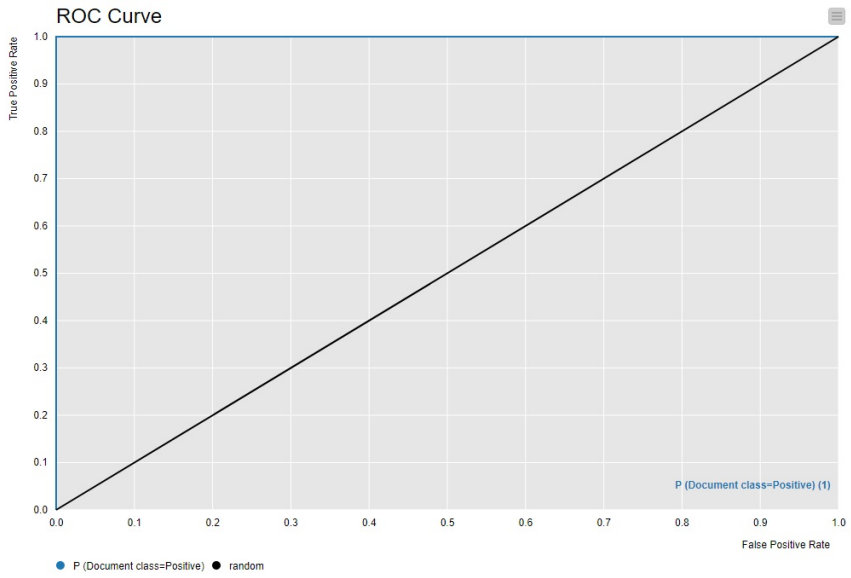

Figure 4. ROC curve

The results of testing the accuracy or accuracy using the Naive Bayes algorithm using partitioning techniques obtained an AUC value: 1.00 (positive class: Positive) so that it included excellent classification.

5. Conclusion

From the test results using the Naive Bayes Algorithm by using sentiment analysis review data on the Shopee application in Playstore, from 200 review data taken consisting of 100 positive reviews and 100 negative reviews. The application of data mining can help categorize shopee application sentiment reviews. Naive Bayes algorithm with partitioning technique produces Accuracy 96,667%,precision 100% , recall 93.33% , and AUC 1.00 so that it is included in the excellent classification. Naive Bayes sentiment analysis with proven effect in producing high accuracy values. So it can be concluded that the application of partitioning with the Naive Bayes algorithm can provide a solution to the problem of sentiment classification of Shopee applications.

References

[1] Muljono Artanti D P Syukur A Prihandono A and Setiadi D R I M, 2018 Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes Konf. Nas. Sist. Inf. 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018 Anal. p. 8–9.
[2] Müller J, 2020, Top 10 e-commerce sites in Indonesia as of 1st quarter 2020, statista.com. [Online]. Available: https://www.statista.com/statistics/869700/indonesia-top-10-e-commerce-sites/.
[3] Widagdo A S A B S W and Nasiri A, 2020 Analisis tingkat kepopuleran e-commerce di indonesia berdasarkan sentimen sosial media menggunakan metode naïve bayes J. Inf. Politek. Indonusa Surakarta 6, 1 p. 1–5.
[4] Rachmat C A and Lukito Y, 2016 Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes J. Inform. dan Sist. Inf. Univ. Ciputra 02, 02 p. 26–34.
[5] Muthia D A, 2014 Analisis Sentimen Pada Review Buku Menggunakan Algoritma Naive Bayes J. Paradig. vol XVI no.1 Maret 2014 XVI, 1 p. 8–16.
[6] Heryono H and Kardianawati A, 2018 Implementasi Metode Naive Bayes Untuk Klasifikasi Kredit Motor JOINS (Journal Inf. Syst. 3, 1 p. 10–21.

[7] Lorosae T A Prakoso B D Saifudin and Kusrini, 2018 Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan Naive Bayes p. 25–30.
[8] Melagraki G and Afantitis A, 2013 Enalos KNIME nodes: Exploring corrosion inhibition of steel in acidic medium Chemom. Intell. Lab. Syst. 123 p. 9–14.
[9] Kurniawan S Gata W Puspitawati D A Parthama I K S Setiawan H and Hartini S, 2020 Text Mining Pre-Processing Using Gata Framework and RapidMiner for Indonesian Sentiment Analysis IOP Conf. Ser. Mater. Sci. Eng. 835, 1.