Jurnal Mantik, 6 (3) (2022), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Product Review Sentiment Analysis At Online Store Jiniso Official Shop Using Naive Bayes Classifier (Nbc) Method

Efta Riana Putri[1], Herny Februariyanti[2]
[1,2]Universitas Stikubank Semarang, Indonesia

A R T I C L E I N F O

Article history:

Received Jul 2, 2022
Revised Jul 18, 2022
Accepted Aug 10, 2022

Keywords:

Naïve Bayes Classifier
Python
Text Mining
Preprocessing Data
Shopee

ABSTRACT

It is undeniable that online shopping is the choice of many people during the pandemic, because online shopping is an easy and safe solution for people who are required to carry out social distancing by the government. Jiniso Official shop is one of the online stores that currently sell through Shopee. This study was conducted to categorize and analyze customer views of the product by using Jiniso product review data from the comments column on the Shopee application. In this study, researchers will utilize the text mining process using classification techniques, the algorithm that will be used is the Naive Bayes algorithm. Naive Bayes allows classification based on the assumption of separate conditions between the predicted attributes of a particular class. For this reason, the Naive Bayes Classifier is a very competent classification, it works quite well in classification tasks so many researchers are trying to improve the performance of Naive Bayes [3]. The results of sentiment analysis using NBC produces an accuracy rate of 0.941747572815534 or 94%. From the results of this study, positive sentimental reviews can be used as a reference to maintain things that make customers feel satisfied. while negative reviews can be used as motivation to improve services and products.

Corresponding Author:

Efta Riana Putri,
Information Technology and Industry Faculty,
Universitas Stikubank Semarang,
Jl. Tri Lomba Juang, Mugassari, Kec. Semarang Sel., Kota Semarang, Jawa Tengah 50241, Indonesia
Email: eftarianaputri@mhs.unisbank.ac.id

1. INTRODUCTION

JINISO Official Shop is Jiniso's official account on Shopee. Jiniso is one of Indonesia's local brands with the slogan "JINISO Active Jeans". In one month Jiniso was able to reap billions of sales and sales could penetrate up to thousands of packages in a day, Jiniso has crowned the brand with sales of jumbo-sized jeans at number no. 1 in the market. For information, this local brand of jeans is indeed famous for its very complete size variations, from the smallest to the largest sizes. Online shopping or what is known as an online store often colors virtual world trade, is a place to buy goods and services via the internet, or as a form of electronic commerce, it is undeniable that online shopping is the choice of many people during a pandemic, because Online Shopping is an easy and safe solution for people who are required to do social distancing by the government. In doing online

shopping, people only need an internet subscription fee to be able to access all their shopping needs. Before people decide to buy a product, of course, it is important to know the history of the seller and how the seller offers the product, buyers can find out by looking at reviews about the product that can be read. by prospective buyers through product reviews on the website. online shop site.

This study was conducted to categorize and analyze customer views on products by using Jiniso product review data from the comments column on the Shopee application. In this study, researchers will utilize the text mining process, which is the stage of the data analysis process in the form of text where the data source is obtained from a document in the form of sentence data. The concept of text mining is usually used in the classification of textual documents where the document will be classified according to the topic of the document (Darwis et al., 2021), the algorithm that will be used is the Naive Bayes algorithm. Naive Bayes is one of the algorithms in text mining that is used to classify text using probability and statistical calculations proposed by Thomas Bayes. Naive Bayes allows classification based on the assumption of separate conditions between the predicted attributes of a particular class. For this reason, the Naive Bayes Classifier is very competent, performing well enough in the classification task that many researchers are trying to improve the performance of Naive Bayes (Humairah et al., 2020).

Based on the previous analysis, the Naive Bayes classification (NBC) is used to analyze the level of customer satisfaction at Baksi President Restaurant, the results of which can be used as an option in conducting aspect level sentiment analysis for customer satisfaction assessment of President Malang. meatballs (Parasati et al., 2020). The previous Naive Bayes classification analysis was carried out using Rapid Miner to analyze sentiment on Twitter. Based on the test results, the classification accuracy level using the Nave Bayes Classifier method has the highest value in the fourth experiment with 66.79% with a data comparison of 60% for training data. and 40% for test data. The lowest level of accuracy was obtained in the first experiment of 62.98% with a comparison of data sharing, namely 20% for training data and 80% for test data (Darwis et al., 2021). The Naïve Bayes classification is used to carry out a sentiment analysis on Indonesian marketplace customer opinion posts on Twitter, the results of the clustering process are made up of two data classes, namely positive data class and negative data class with a total of 1200 data. . To determine the performance of the Naive Bayes classifier in the classification process, an experiment was conducted using 10 times of cross-validation (10 times of cross-validation). The experimental results obtained an average accuracy of 93.33% (Artanti et al., 2018).

Based on reviews given by Shopee application users, it is hoped that there will be feedback for developers to improve service quality (Qamal, 2021). The results of this analysis will be in the form of positive and negative sentiments, the results of the sentiment analysis are then visualized using Wordcloud to see the words that appear most often for each sentiment. From the results of this study, positive sentiment can be used as a reference for the Jiniso Official Shop store to maintain the quality of product sales services and the results of negative sentiment can be used as evaluation material so as to improve product quality. sale. better service at the Jiniso Official Shop online store.

2. RESEARCH METHOD

Text mining can be interpreted as the discovery of new information not previously known by computers, by extracting information automatically from various sources. The key to this process is to combine information that has been extracted from multiple sources (Ratnawati, 2018). Sentiment analysis is a computational research of opinions, sentiments, and emotions expressed textually (Setian & Seprina, 2019). The data analysis method used to analyze product review sentiment at the Jiniso Official Shop online store uses the Naïve Bayes classifier algorithm. The algorithm is widely used in analysis, especially sentiment analysis to generate sentiment data both positive and negative (Nugroho, Agung, Dzulatkha, Alfatan, 2020).The Naïve Bayes model has a high level of accuracy when implemented on large data in the database discovered by Thomas Bayes in the 19th century.

20-18. The basic concept is the Bayes theorem, which is calculating the probability of classification (Taufik, 2017). The stages carried out in this research can be seen in the following figure:

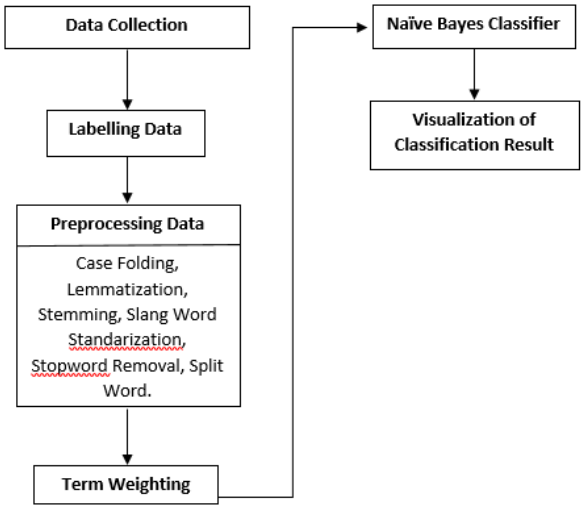

Figure 1. Research Method Diagram

2.1 Data Collection

Data collection is determined based on the data to be processed in the form of positive reviews and negative reviews (Nugroho, 2018). Customer review data collection using the scraping method using the Instant Data Scraper. This tool can be perfectly compatible with all the websites to be scratched, to use it you need to visit the web store on Google Chrome and then add it as an extension. In use, the results of the scratched data can be exported into excel and CSV forms which can then be downloaded.

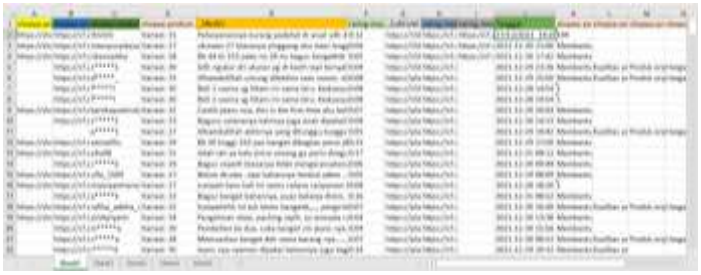

Figure 2. Scraping Result Data

2.2 Labelling Data

A total of 1.207 data were obtained from scraping which was taken in December 2021, from the final results of the data labeling process 977 positive reviews and 50 negative reviews were obtained.

```
Positif    977
Negatif     50
```

Figure 3. Number of Data Labels

2.3 Preprocessing Data

After labeling the data, the data will go through the preprocessing stage. The data processing process or text preprocessing serves to convert unstructured text data into structured data (Sari &

Wibowo, 2019). At this stage it is done by using a library in the Python programming language. Data preprocessing is carried out with the stages of Case Folding, Lemmatization, Stemming, Slang Word Standardization, Stopword Removal, and Split Word to produce clean data and ready to be processed at a later stage. The process carried out is as follows.

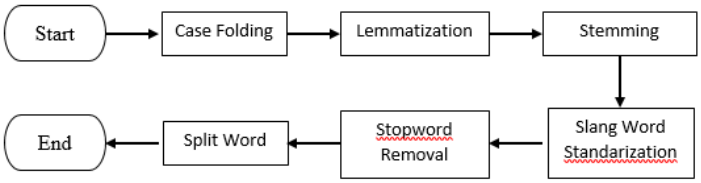

Figure 4. Preprocessing Stage

2.4 Term Weighting

Term Frequency is a weighting concept by looking for how often (frequency) a term appears in a document (Ruger et al., 2021) which is used to assign a weight value to each word contained in the dataset by counting the number of occurrences that will affect the determination of document keyword linkages. The term weighting process is carried out using the Term Frequency – Inverse Document Frequency (TF-IDF) method.

$$idf_j = log\ (\frac{D}{df_j}) \quad (1)$$

2.5 Naïve Bayes Classifier

The Naive Bayes algorithm is a classifier method used in probability and statistical calculations (Sari & Wibowo, 2019). In using the Naïve Bayes algorithm, we need a value (weight) for each word contained in the dataset document. This value is obtained by calculating the probability (probability) of being positive or negative for the words in each document.

Probability calculation aims to calculate the value (weight) of each word in the dataset document. The data needed in this calculation is the number of words that are categorized as positive or negative, the occurrence of each word in the document which is categorized as positive or negative, and all words. This value will later be used in the Naïve Bayes Classification which will affect the determination of accuracy, recall, and precision which are the parameters of the results in this study.

Classification using Naïve Bayes is done by comparing the value (weight) of each word contained in the dataset. If the value (weight) of the resulting word opportunity (probability) is in a positive category, the sentiment is positive, and vice versa.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (2)$$

Description:

P(C | X) : Probability of class against the document
P(X}C) : The probability of a document occurring in a class
P(C) : Classification class probability
P(X) : Probability of document occurrence

2.6 Visualization Of Classification Result

The classification results will be visualized in the form of a word cloud which aims to display words that often appear in every positive and negative sentiment. Wordcloud is a form of data visualization in the form of text that describes a large collection of words. Wordcloud is built using

the Wordcloud library and PIL (Python Imaging Library), which is a library for the Python programming language. From this process, the more words used, the larger the size of the words displayed in the word cloud.

3. RESULTS AND DISCUSSIONS

3.1 Preprocessing Data

The purpose of text processing is that the data obtained will be more structured so that it is easier to perform data processing (Mahardika & Zuliarso, 2018).

Case Folding is the process of generalizing the use of capital letters so that they become lowercase while eliminating characters other than a-z.

	ulasan	label
0	Pelanyanannya kurang padahal dr awal udh dijel...	Negatif
1	ukuraan 27 biasanya pinggang aku masi longgar ...	Positif
2	Bb 44 tb 153 pake no 28 itu bagus bangettttt a...	Positif
3	Udh ngukur dri ukuran yg di kasih tapi ternyat...	Positif
4	Alhamdulillah untung dilebihin satu nomor, dan...	Positif

Figure 5. Before The Case Folding Stage

```
0    pelanyanannya kurang padahal dr awal udh dijel...
1    ukuraan biasanya pinggang aku masi longgar ini...
2    bb tb pake no itu bagus bangettttt agak kepanj...
3    udh ngukur dri ukuran yg di kasih tapi ternyat...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 6. After The Case Folding Stage

Lemmatization is a stage that aims to normalize text/words based on their basic form, namely lemma, lemma is the basic form of a word that has a certain meaning.

```
0    pelanyanannya kurang padahal dr awal udh dijel...
1    ukuraan biasanya pinggang aku masi longgar ini...
2    bb tb pake no itu bagus bangettttt agak kepanj...
3    udh ngukur dri ukuran yg di kasih tapi ternyat...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 7. Before The Lemmatization Stage

```
0    pelanyanannya kurang padahal dr awal udh dijel...
1    ukuraan biasanya pinggang aku masi longgar ini...
2    bb tb pake no itu bagus bangettttt agak kepanj...
3    udh ngukur dri ukuran yg di kasih tapi ternyat...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 8. After The Lemmatization Stage

Stemming is the process of finding the root of a word and removing the affixes in the word. Stemming aims to reduce variations of words that have the same root. For example, the word "mendengarkan", and "didengarkan" will be changed to "dengar". At the stemming stage, it is done using the help of a library in the Python programming language, namely Sastrawi.

```
0    pelanyanannya kurang padahal dr awal udh dijel...
1    ukuraan biasanya pinggang aku masi longgar ini...
2    bb tb pake no itu bagus bangettttt agak kepanj...
3    udh ngukur dri ukuran yg di kasih tapi ternyat...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 9. Before The Stemming Stage

```
0    pelanyanannya kurang padahal dr awal udh dijel...
1    ukuraan biasa pinggang aku mas longgar ini nge...
2    bb tb pake no itu bagus bangettttt agak panjan...
3    udh ngukur dri ukur yg di kasih tapi nyata tet...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 10. After The Stemming Stage

At this stage the process of standardizing words from words that are not standard but is often used in general by the community, this process requires a slang dictionary in CSV format, therefore researchers use a dictionary to standardize slang words.

```
0    pelanyanannya kurang padahal dr awal udh dijel...
1    ukuraan biasa pinggang aku mas longgar ini nge...
2    bb tb pake no itu bagus bangettttt agak panjan...
3    udh ngukur dri ukur yg di kasih tapi nyata tet...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 11. Before The Slang Word Standardization Stage

```
0    pelanyanannya kurang padahal dari awal sudah d...
1    ukuraan biasa pinggang aku mas longgar ini nge...
2    bb tb pakai no itu bagus banget agak panjang d...
3    sudah ngukur dari ukur yang di kasih tapi nyat...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 12. After The Slang Word Standardization Stage

Stopword removal is done to eliminate stopwords, this process is carried out to take important words from the token results and remove words that often appear and are general in nature, not showing their relevance to the text.

```
0    pelanyanannya kurang padahal dari awal sudah d...
1    ukuraan biasa pinggang aku mas longgar ini nge...
2    bb tb pakai no itu bagus banget agak panjang d...
3    sudah ngukur dari ukur yang di kasih tapi nyat...
4    alhamdulillah untung dilebihin satu nomor dan ...
```

Figure 13. Before The Stopword Removal Stage

```
0    pelanyanannya dijelasin salah size jinisonya b...
1    ukuraan pinggang longgar ngepres bangettr coba...
2    bb tb pakai no bagus dikit pinggang besar diki...
3    ngukur ukur kasih nyata gede enggak apa-apa ba...
4    untung dilebihin nomor longgar dikit engenggak...
```

Figure 14. After The Stopword Removal Stage

At the word split stage is done to break the string so as to produce a list of strings.

```
0    pelanyanannya dijelasin salah size jinisonya b...
1    ukuraan pinggang longgar ngepres bangettr coba...
2    bb tb pakai no bagus dikit pinggang besar diki...
3    ngukur ukur kasih nyata gede enggak apa-apa ba...
4    untung dilebihin nomor longgar dikit engenggak...
```

Figure 15. Before The Split Word Stage

```
0    [pelanyanannya, dijelasin, salah, size, jiniso...
1    [ukuraan, pinggang, longgar, ngepres, bangettr...
2    [bb, tb, pakai, no, bagus, dikit, pinggang, be...
3    [ngukur, ukur, kasih, nyata, gede, enggak, apa...
4    [untung, dilebihin, nomor, longgar, dikit, eng...
```

Figure 16. After The Split Word Stage

3.2 Naïve Bayes Classifier

Before performing sentiment analysis using nave Bayes classification, the data will be divided into training data and test data. At this stage, the training data and test data will be divided by a ratio of 80:20, which means that 80% of the total data will be used as training data and 20% of the total data will be used as test data. Next, the nave Bayes classification process will be carried out using a library in the Python programming language called scikit-learn to process the classification. The scikit-learn libraries used include MultinomiaNB, confussion_matrix, f1, score, accuracy_score, recall_scpre, precision_score, classification_report.

```
MultinomialNB Accuracy: 0.941747572815534
MultinomialNB Precision: 0.0
MultinomialNB Recall: 0.0
MultinomialNB f1_score: 0.0
confusion matrix:
 [[  0  11]
 [  1 194]]
-----------------------------------------------

              precision    recall  f1-score   support

     Negatif       0.00      0.00      0.00        11
     Positif       0.95      0.99      0.97       195

    accuracy                           0.94       206
   macro avg       0.47      0.50      0.48       206
weighted avg       0.90      0.94      0.92       206
```

Figure 17. Naïve Bayes Classifier

Based on Figure 17 shows the accuracy value that can be generated is 0.941747572815534 or 94%. The Precision value indicates the amount of data for the positive category that is correctly classified from the total data that is classified as positive. The recall shows what percentage of data in the positive category were correctly identified by the system. The f1-score value is the average weight of precision and recall harmonics. The precision value that can be generated is 0.0 or 0%, the recall value that can be generated is 0.0 or 0% and the f1-score value is 0.0 or 0%.

Figure 17 shows a confusion matrix in the form of a matrix with a size of 2x2 which represents each positive and negative classification class. In the confusion matrix, there are 4 representational terms in the confusion matrix classification process, including True Positive (TP) which is positive data that is predicted to be correct, and True Negative (TN) is negative data that is predicted to be correct. , False Positive (FP) is negative data but predicted as positive data, and finally, False Negative (FN) is positive data but predicted as negative data.

From the confusion matrix in Figure 17 it can be explained that the model classifies True Positive as 0 and 194 as True Negative. Then for the False Positive classification 11 and 1 are False Negatives. Based on the description of the confusion matrix, the accuracy value is 94% by calculating:

Accuracy = ((0+194)/(0+11+1+194))
= (194)/(206)
= 94%

3.3 Visualization Of Classification Result

Visualization aims to display the words that most often appear in a sentiment. In this study, visualization is done using a word cloud for words that often appear in positive and negative sentiments. Wordcloud was created using the word cloud library contained in the python programming language.

Figure 18. Wordcloud Positive

Based on Figure 18, it can be seen that in positive sentiment there are several words that stand out such as "bagus", "suka", "beli", "ukur", "jiniso". This shows that several reasons that make customers happy are good quality, preferred products, and the right size so that customers leave positive reviews.

Figure 19. Wordcloud Negative

Based on Figure 19, it can be seen that in negative sentiment there are several words that look like “bahan”, “size”, “gede”. This shows that several factors that influence customers in giving negative reviews lie in the size or size that is too large.

4. CONCLUSION

Based on the results of the sentiment analysis on the Jiniso Official Shop product review using the Naïve Bayes algorithm, it can be concluded that the classification using the Naïve Bayes method produces an accuracy rate of 0.941747572815534 or 94%. From the sentiment analysis of the Jiniso Official Shop product reviews, it produces positive and negative sentiments, positive sentiment reviews that can be used as a reference to maintain things that make customers feel satisfied. while negative reviews can be used as motivation to improve services and products.

REFERENCES

Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). *Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes*. 8–9.

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, *15*(1), 131. https://doi.org/10.33365/jtk.v15i1.744

Humairah, Darmawan, I., & Pratiwi, O. N. (2020). Analisis Sentimen Ulasan Produk Toko Online Rubylicious Untuk Peningkatan Layanan Menggunakan Algoritma Naive Bayes. *E-Proceeding of Engineering*, *7*(2), 7026–7034.

Mahardika, Y. S., & Zuliarso, E. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes. *Prosiding SINTAK 2018*, *2015*, 409–413.

Nugroho, Agung, Dzulatkha, Alfatan, V. 10 N. 3 M. 2020 I. : 2407-3903. (2020). PENERAPAN INTERNET OF THINGS (IOT) PADA SISTEM KONTROL DAN MONITORING LAMPU GEDUNG MENGGUNAKAN NODEMCU BERBASIS TELEGRAM. *SIGMA - Jurnal Teknologi Pelita Bangsa 167*, *10*(September), 167–172.

Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *2*(2), 200. https://doi.org/10.30645/j-sakti.v2i2.83

Parasati, W., Abdurrachman Bachtiar, F., & Setiawan, N. Y. (2020). Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *4*(4), 1090–1099. http://j-ptiik.ub.ac.id

Qamal, M. (2021). Analisis Sentimen Toko Online Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Teknologi Terapan and Sains 4.0*, *1*. https://ojs.unimal.ac.id/tts/article/view/6771%0Ahttps://ojs.unimal.ac.id/tts/article/download/6771/3301

Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, *3*(1), 50. https://doi.org/10.35314/isi.v3i1.335

Ruger, A. H., Suyanto, M., & Kurniawan, M. P. (2021). Sentimen Analisis Pelanggan Shopee di Twitter dengan Algoritma Naive Bayes. *Journal of Information Technology*, *1*(2), 26–29. https://doi.org/10.46229/jifotech.v1i2.282

Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, *10*(2), 681–686.

Setian, D., & Seprina, I. (2019). ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET LAZADA MENGGUNAKAN TEXT MINING DAN ALGORITMA NAIVE Bina Darma Conference on Computer Science P (c | X) = P (x | C) P (c) P (x). *Bina Darma Conference on Computer Science*, 998–1004.

Taufik, A. (2017). Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Komputer*, *III*(2), 40–47.