

# Sentiment Analysis of JNE User Perception using Naïve Bayes Classifier Algorithm

Annisa Uswatun Khasanah<sup>1</sup> and Adelia Febriyanti<sup>1</sup>

<sup>1</sup> Department of Industrial Engineering

Islamic University of Indonesia, Yogyakarta

email : [annisa.uswatun@uii.ac.id](mailto:annisa.uswatun@uii.ac.id)

two: <https://doi.org/10.31315/opsi.v15i1.7179>

Received: 13<sup>rd</sup> June 2022; Revised: 14<sup>th</sup> June 2022; Accepted: 15<sup>th</sup> June 2022;

Available online: 18<sup>th</sup> June 2022; Published regularly: June 2022

## ABSTRACT

The logistics industry is growing very rapidly. One of big industry in Indonesia is PT. Tiki Line Nugraha Ekakurir (JNE), which has been established for 29 years. This company has an extensive network in all cities in Indonesia, with service points of 1,500 locations. JNE has an application called my JNE on Google Play, which received more than 86,000 reviews and since December 2019 only got a rating of 2.4 stars out of a total rating of 5 stars. This study is obtained to analysis JNE user review data from Google Play. The reviews used in this study totaled 1,876 classified into positive and negative sentiment classes using the Naïve Bayes Classifier algorithm and word associations were also implemented. Classification with naïve bayes classifier with 90% training data and 10% test data had the best accuracy of 85.87%. Furthermore, for the text association, information is obtained that JNE users are talking about "send", "package", "courier", "good", "application", "fast", "service", "receive", "help", and "star". Whereas in the class of negative sentiment users often talk about "send", "package", "courier", "disappointed", "service", "service", "bad", "application", "severe", and "slow".

**Keywords:** Sentiment Analysis, Word Associations, Fishbone Diagram, JNE, Google Play, Naïve Bayes Classifier

## 1. INTRODUCTION

Information technology is growing rapidly. Humans are required to constantly develop technology. This demand is carried out to support the process of processing data and information so that it can be carried out more effectively and efficiently (Josi et al., 2014). The rapid development of technology has brought human civilization to an all-digital era. One of the impacts of the digital era is the development of internet technology with users that continue to grow every year. Based on a survey conducted by the Indonesian Internet Service Providers Association (APJII) in March-April 2019, it is known that there was an increase in users by 27 million people to 171.17 million users.

The increase in the number of internet users has made various companies compete in marketing their products through the internet media. Based on a survey conducted by APJII (2018), it is known that selling online is included in the top 10 reasons why respondents in Indonesia use the internet. Whereas buying online is included in the top 15 reasons for

internet use. Therefore, there are many online buying and selling sites also known as ecommerce.

The growth of e-commerce is certainly supported by the logistics industry, namely shipping companies. The Central Statistics Agency explained that economic growth in Indonesia throughout 2017 reached 5.19% with the transportation and warehousing sector being one of the highest. The shipping company will send the order from the seller to reach the consumer. In the delivery of this item, everything is controlled through information technology, starting from the delivery of goods to the expedition officer, updating the latest status of the location of the goods, until the goods are finally received by consumers. The development of logistics services is also evidenced through the large number of companies engaged in freight forwarding services. One of the well-known and most widely used shipping and logistics service companies is PT. Tiki Line Nugraha Ekakurir (JNE). Based on Top Brand Indonesia data related to the most widely used courier services

in 2015 to 2019, it shows that JNE has always been at the top of the list (Top Brand Award, 2020).

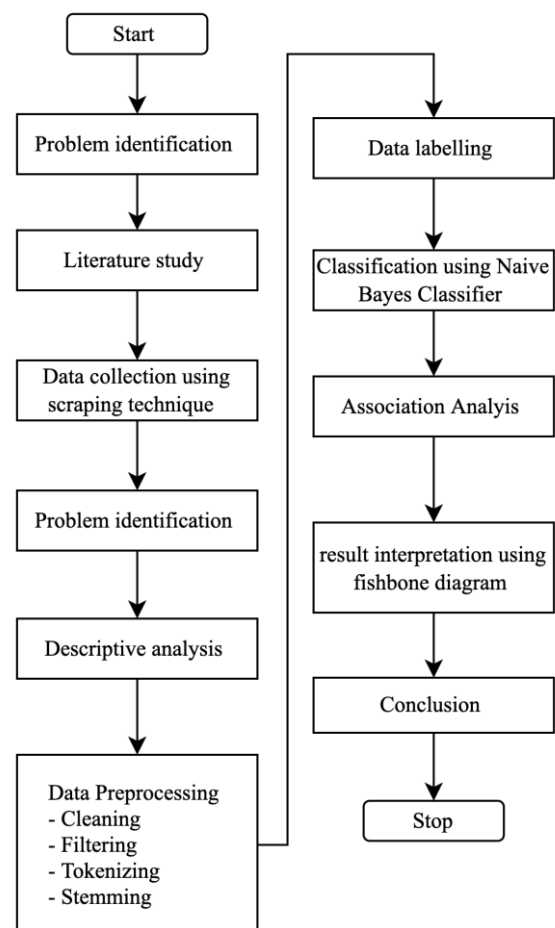
JNE is a logistics service company that has been established for 29 years. JNE has an extensive network in all cities in Indonesia, with service points of 1,500 locations and a total of more than 40,000 employees. JNE's communication system has been connected online with more than 150 locations. JNE has an app called My JNE which was released on January 27, 2016, on Google Play. As for now, my JNE application has been downloaded by more than one million users and has received more than 86,000 reviews covering not only the application but also JNE's overall services. This review includes complaints of a negative nature as well as positive suggestions.

User reviews are one of the effective and efficient media to find information about the image of a company. This is because consumers see the perspective of word-of-mouth recommendations (opinions) of other users before using a product or service (Fanani, 2017). The rating obtained by JNE on *Google Play* until December 2019 is quite low, which is only 2.4 stars out of a total of 5 stars. This study aims to conduct further analysis related to JNE user reviews with the Naïve Bayes Classifier algorithm which has a minimum error rate compared to other classification algorithms (Liu, et al., 2016). Jumeilah (2018) conducted study on sentiment analysis using Naïve Bayes and obtained 85% of accuracy. Ibrohim and Budi (2018) also conducted study by comparing four methods and Naïve Bayes obtained the best accuracy of 70,06%. While Azalia (2019) also conducted study using naïve bayes and obtained 82,63% of accuracy. Classification is carried out on positive and negative sentiments and information extraction using descriptive analysis and associations between words to be able to find what words often appear in JNE user reviews. Furthermore, an analysis was carried out using a fishbone diagram on negative sentiment to find out the causal factors, so that JNE can find out what needs to be done to improve its services.

This study aims to find out the general perception of JNE users through the My JNE application, find out the performance of the Naïve Bayes Classifier algorithm, and factors improvements based on negative reviews.

## 2. METHODS

The subject of this study is JNE users' perceptions of JNE services through reviews on My JNE on the Google Play website. The data used is the latest version of My JNE user review data, which is 2.3.12, from September 19, 2019, to December 31, 2019, with a total sample of 1,876 comments. The data is taken online with scraping techniques using an extension in Google Chrome, namely Data Scraper version 3.299.84. The data taken include the time (when reviews are uploaded), reviews, and ratings. The flow of the study can be seen in Figure 1.



**Figure 1.** Research flow

*Preprocessing* of data is carried out before the classification process. *Preprocessing* is obtained to select and convert data into structured data. This is also the initial stage to change the content structure of the data to become an appropriate format so that it can be processed by the *Naïve bayes Classifier algorithm*. The process of *this preprocessing* is

carried out using *R* and *Python* software. The preprocessing stages carried out are *cleaning*, *filtering*, *tokenizing*, and *stemming*.

The classification method used is Naïve Bayes Classifier be the simplest method of probability classification, has a very high level of accuracy when applied to databases with big data (Wati, 2016). Naïve Bayes has been widely used in sentiment analysis problems such as Dey et al. , (2016), Liu et al. (2016), Raksanagara et al. (2016), Barfian et al., (2017), Mukherjee and Bala (2017), El-Masri et al. , (2017), Jumeilah (2018), Kunal et al. (2018), Ibrohim and Budi (2018), Mahadzir et al. (2018), Sánchez-Franco et al., (2019) and Azalia et al., (2019).

After the classification is carried out, the association analysis is implemented to identify important patterns associated within words. And the last step is the identification of the negative sentiment class review with the fishbone diagram.

### 3. RESULTS AND DISCUSSION

#### 3.1. Descriptive Analysis

Descriptive analysis is carried out to find out a general description of JNE services. On Image 2 there are user reviews based on time as follows.

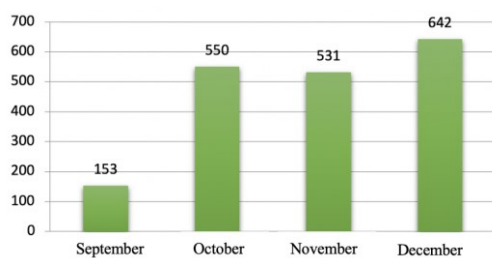


Figure 2. Number of reviews

Based on Figure 1, with the time span of data collection during September 19, 2019 to December 31, 2019, the results of the graph tend to increase. Meanwhile, in September there were only 153 reviews because the data was taken according to the upgrade date of the latest version of the my JNE application, which starts on September 19. In October there were 550 reviews, while in November it decreased to 531 reviews. Furthermore, in December reviews increased to 642, this is estimated to happen because in that month there was a National Online Shopping Day (HARBOLNAS) which

increased the number of e-commerce sales so that the use of shipping services including JNE also increased.

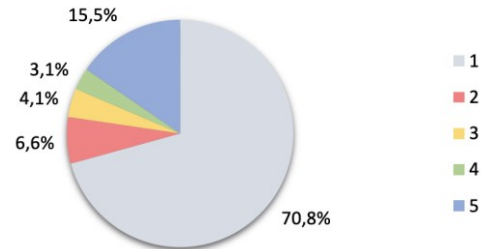


Figure 3. Review rating percentage

Based on Figure 3, it can be seen that out of a total of 1,876 reviews, as many as 70.8% or as many as 1,328 users gave a rating of 1 or not good. Furthermore, the second highest percentage is 15.5% or 290 users which gives a rating of 5 or very good. Then followed by 2 ratings of 6.6% with 123 users, 3rd ratings of 4.1% with 77 users, and 4th ratings of 3.1% with 58 users.

#### 3.2. Sentiment Class Labeling

Before labeling the sentiment class, preprocessing the data is first carried out using Python and R Studio tools. Furthermore, the labeling of the sentiment class is done using R studio software, with the results of which can be seen in The Figure 4.

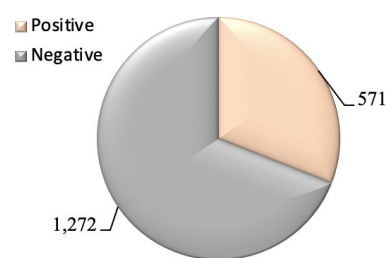


Figure 4. Review rating percentage

#### 3.3. Classification with Naïve Bayes Classifier

While other studies (Jumeilah,2018; Ibrohim and Budi, 2018; and Azalia, 2019) only one type division of training and testing data, in this study the classification is carried out by dividing the training data and test data into five comparisons, 60% 40%, 70% 30%, 80% 20%, 85% 15%, and 90% 10%. After classification with five types of differences in training and test

data, a comparison of accuracy results can be carried out which can be seen on Figure 5 :

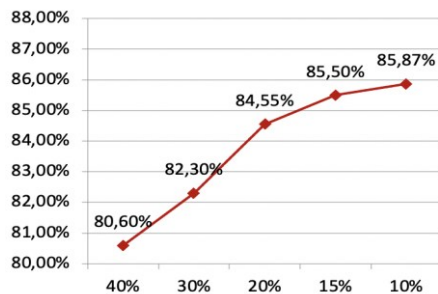


Figure 5. Accuracy comparison

Based on Figure 5, it is known that the average classification accuracy with the Naïve Bayes Classifier algorithm will tend to rise as the test data used in the test decreases. The maximum results were obtained in the comparison of test data of 10% and training data of 90%, with the highest accuracy of 85.87% and a precision of more than 80% for positive and negative classes. These results are quite better if compared with Jumeilah (2018), Azalia (2019). The overall value of precision and recall has increased as the test data used as can be seen in the Figure 6.

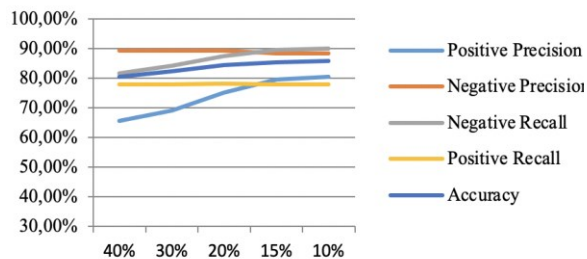


Figure 6. Accuracy comparison

### 3.4. Visualizations and Associations

Visualizations and Associations are performed on The Sentiment Class of Positive Reviews, Negative Reviews, and Reviews as a whole:

#### - Positive Reviews

Based on the barplot visualization of positive reviews in Figure 7, it is known that there are 10 words that often appear in positive reviews, including: send, package, courier, good, application, fast, service, receive, help, and star.

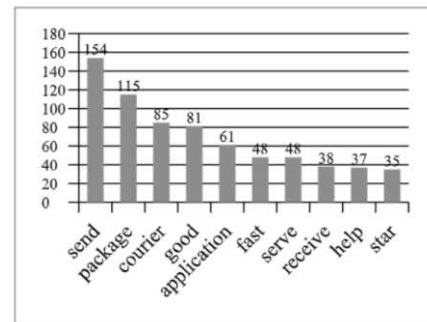


Figure 7. Positive review sentiment visualization

As for the association of positive reviews, it can be seen in Figure 8 :



Figure 8. Visualization of Positive Sentiment

#### - Negative Reviews

Based on the barplot visualization of negative reviews in Figure 9, it is known that there are 10 words that often appear in positive reviews, including: send, package, courier, disappointed, service, service, bad, application, severe, and slow. As for Figure 10 , there are associations between words in negative reviews:

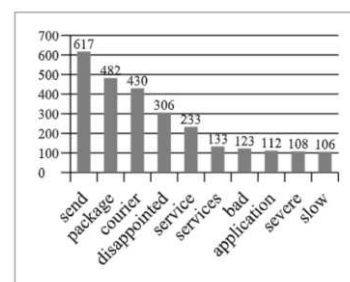


Figure 9. Negative review sentiment visualization

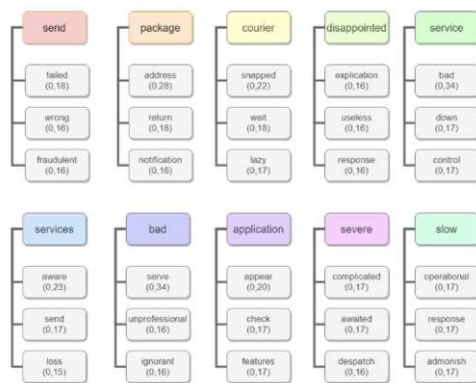


Figure 10. Visualization of Negative Sentiment

Based on the visualization and association of positive and negative reviews, it is known that there are words that appear simultaneously both positively and negatively, which are send, package, courier, service, application, and service. Furthermore, an analysis of the word is carried out by making associations as a whole.



Figure 11. Word Association of Overall review

### 3.5. Diagram Fishbone

Based on negative reviews and overall associations, information was obtained regarding problems that occurred in JNE which was analyzed with a fishbone diagram based on 6 factors, namely Process, People, Price, Promotion, Place, and Product in Figure 11.

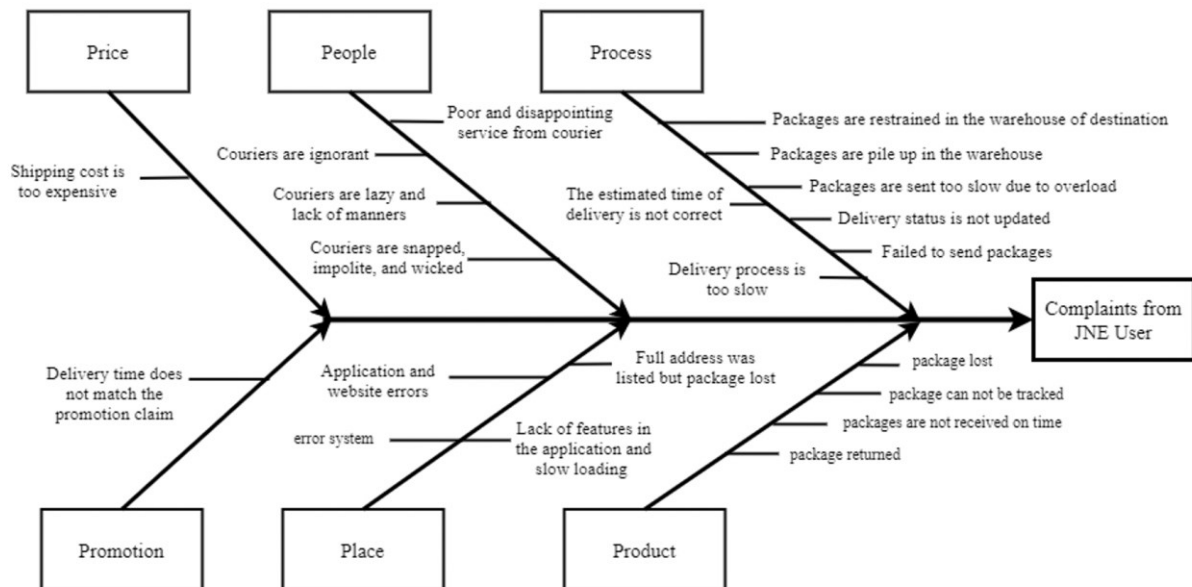
### 3.6. Problem Solving Plan

**Process and Product Aspects:** (1) Grouping packages from the beginning so that nothing is scattered and improving the warehouse layout which is arranged based on the type of package and destination address and marking packages with a certain code according to the destination address so that it is easy to identify; (2) Packages that have been given a certain code according to the address are sent from the warehouse to each Drop Point spread across several regions; (3) If there is a problem in the warehouse related to the obstruction of shipping via air or cargo, JNE should contact the customer so that the customer is aware of the problems that occur; (4) Minimize unnecessary processes to speed up transit times; (5) Provide a minimum daily target that must be met for package delivery; (6) The developer immediately improves the receipt check system on both the website and application so that the delivery status is always updated and tightens the system so that there is no more delivery failure due to the address that is considered incomplete.

**People Aspects:** (1) Create SOPs and conduct training for couriers and customer service to serve customers well; (2) Add an assessment feature to the application in the form of ratings, criticisms and suggestions for couriers who deliver packages so that customers can provide an assessment of the courier's performance; (3) Add information features related to couriers who deliver packages in which there is a telephone number or courier chat access directly; (4) Provide a database in the form of a telephone number from each JNE agent in the application to make it easier for customers to find information on the whereabouts of packages; (5) Tighten the controlling or supervising process of courier and CS performance and carry out performance evaluations; (6) Provide strict penalties for couriers and CS who do not provide good service.

**Price Aspect:** Reassessing shipping costs and providing discounts on shipping costs for customers who are disappointed because the package did not arrive on time.

**Promotion Aspect:** JNE must be more consistent in the implementation of delivery in accordance with existing promotional claims and provide information to customers if there are obstacles that cause the package not to arrive on time.



**Figure 10.** Research Flow Interpretation of Fishbone Diagrams Negative Reviews

**Place Aspects:** (1) Regularly update the system both on the application and on the website and test the application until no errors are detected; (2) Add the necessary features to support JNE's performance and maintain applications and websites at least once a year.

#### 4. CONCLUSIONS

Based on the analysis that has been done, 69% of users have a positive perception while 31% of users have a negative perception. Based on the analysis that has been done, 69% of users have a positive perception while 31% of users have a negative perception. The application of the Naïve Bayes Classifier method was carried out using five types of comparison of training data and test data. The results obtained that the comparison of training data and test data of 90% : 10% resulted in the most optimal level of accuracy, precision, and recall with an accuracy of 85.87%. Based on the results of the classification and association carried out, information was obtained that in general users discuss six things, namely sending, packages, couriers, services, applications, and services. While Based on the fishbone diagram, problems were found based on six factors, which are Process, People, Price, Promotion, Place, and Product.

Then 21 problems were obtained from those factors.

#### REFERENCES

- Aditya. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweet Dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, 7.
- APJII. (2018). Hasil Survei Penetrasi dan Perilaku Pengguna Internet Indonesia 2018. Retrieved December 10, 2019, from <https://www.apjii.or.id/content/read/39/410/Hasil-Survei-Penetrasi-dan-PerilakuPengguna-Internet-Indonesia-2018>
- Award, T. B. (2020). Top Brand Index. Retrieved from <https://www.topbrandaward.com/top-brand>
- Azalia, F. Y., Bijaksana, M. A., & Huda, A. F. (2019). Name Indexing in Indonesian Translation of Hadith using Named Entity Recognition with Naïve Bayes Classifier. *Procedia Computer Science*, 157, 142–149. <https://doi.org/10.1016/J.PROCS.2019.08.151>
- Barfian, E., Iswanto, B. H., & Isa, S. M. (2017). Twitter Pornography Multilingual



- Content Identification Based on Machine Learning. *Procedia Computer Science*, 116, 129–136. <https://doi.org/10.1016/J.PROCS.2017.10.024>
- Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier. *International Journal of Information Engineering and Electronic Business*, 8(4), 54–62. <https://doi.org/10.5815/ijieeb.2016.04.07>
- El-Masri, M., Altrabsheh, N., Mansour, H., & Ramsay, A. (2017). A web-based tool for Arabic sentiment analysis. *Procedia Computer Science*, 117, 38–45. <https://doi.org/10.1016/J.PROCS.2017.10.092>
- Fanani. (2017). Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen. Universitas Gadjah Mada.
- Ibrohim, M. O., & Budi, I. (2018). A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media. *Procedia Computer Science*, 135, 222–229. <https://doi.org/10.1016/j.procs.2018.08.169>
- Josi, A., Abdillah, L., & Suryayusra. (2014). Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah. *Jurnal Sistem Informasi (SISFO)*, 05.
- Jumeilah, F. S. (2018). Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes. *Jurnal Sistem Informasi Bisnis*, 8(1), 92. <https://doi.org/10.21456/vol8iss1pp92-98>
- Kristiyanti, D. A. (2015). Analisis sentimen review produk kosmetik melalui komparasi feature selection. *Konferensi Nasional Ilmu Pengetahuan Dan Teknologi (KNIT)*, 2(2), 74–81.
- Kunal, S., Saha, A., Varma, A., & Tiwari, V. (2018). Textual Dissection of Live Twitter Reviews using Naive Bayes. *Procedia Computer Science*, 132, 307–313. <https://doi.org/10.1016/J.PROCS.2018.05.182>
- Liu, J., Tian, Z., Liu, P., Jiang, J., & Li, Z. (2016). An approach of semantic web service classification based on naive bayes. *Proceedings - 2016 IEEE International Conference on Services Computing, SCC 2016*, 356–362. <https://doi.org/10.1109/SCC.2016.53>
- Mahadzir, N. H., Omar, M. F., & Nawi, M. N. M. (2018). A Sentiment Analysis Visualization System for the Property Industry. *International Journal of Technology*, 9(8), 1609. <https://doi.org/10.14716/ijtech.v9i8.2753>
- Mukherjee, S., & Bala, P. K. (2017). Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering. *Technology in Society*, 48, 19–27. <https://doi.org/10.1016/j.techsoc.2016.10.003>
- Putri, D. (2016). Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten. Universitas Gadjah Mada.
- Raksanagara, R., Chrisnanto, Y. H., & Hadiana, A. I. (2016). Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes. *Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes*, 19–24.
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). Implementasi Opinion Mining ( Analisis Sentimen ) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Electrical Power, Electronics, Communications, Controls, and Informatics Seminar (EECCIS)*, 6(1), 37–43.
- Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, 101, 499–506. <https://doi.org/10.1016/J.JBUSRES.2018.12.051>
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes Risa. *Jurnal Evolusi Volume*, 4. <https://doi.org/10.1017/CBO9781107415324.004>