

Sentiment analysis on myindihome user reviews using support vector machine and naïve bayes classifier method

Sulton Nur Hakim*, Andika Julianto Putra, Annisa Uswatun Khasanah

Department of Industrial Engineering, Universitas Islam Indonesia, Yogyakarta, Indonesia

sultonthakim99@gmail.com; andikaputra6789@gmail.com; 145220102@uii.ac.id

*Corresponding Author: sultonthakim99@gmail.com

ARTICLE INFO

Keywords

Sentiment Analysis;
Support Vector Machine;
Naïve Bayes Classifier.

Article history

Received:
July 12, 2021
Revised:
August 30, 2021
Accepted:
August 30, 2021
Available online:
August 31, 2021

ABSTRACT

In the era of globalization, the internet has become a human need in doing various things. Many internet users are an opportunity for internet service providers, PT Telekomunikasi Indonesia (Telkom). One of PT Telkom's products is IndiHome. As the only state-owned enterprise engaged in telecommunications, PT Telkom is expected to meet the needs of the Indonesian people. However, based on the rating obtained by IndiHome products through the myIndiHome application on Google Play, it is 3.5 out of 87,000 more reviews. The reviews focus on how important the effect of word-of-mouth is on choosing and using internet provider products. The review data was collected on November 1, 2020 to December 15, 2020, with a total of 2,539 reviews as a sample. The sentiment analysis process that has been carried out shows that the number of reviews included in the negative sentiment class was 1.160 reviews, and the positive class was 1.374 reviews out of a total of 2,539 reviews. The results indicate that service errors in IndiHome services are still quite high, reaching 46.7% as indicated by the number of negative reviews. The classification results show that the average value of the total accuracy of the Support Vector Machine (SVM) method is 86.54% greater than Naïve Bayes Classifier (NBC) method which has an average total accuracy of 84.69%. Based on fishbone diagram analysis, there are 12nd problems on negative reviews that classify problems 5P factors: Price, People, Process, Place, and Product.

1. Introduction

Technological development is something that cannot be avoided in this life. In this era of globalization, mastery of technology and the internet is prestige and an indicator of a country's progress and holds the key to the future (Anaeto et.al., 2016). It is the reason for various countries to give significant attention to the development of technology, with the expectations to act as a leading relying on technology (Beumer, 2019), and the internet improvement, especially in Indonesia as it has become both leisure tools and medium of communication (Dogruer et al, 2011)

Based on the survey results of the Indonesian Internet Service Providers Association (IISPA) in June 2020, it shows that there is a penetration of internet users in Indonesia of 73.7% as shown in Figure 1. This value increased by 8.9%, with a total increase of more than 25 million internet users. Currently, the total internet users are 196 million, more than the total Indonesian population of 266 million. The number of internet users will continue to grow, develop and expand network infrastructure in Indonesia (Indonesian Internet Service Providers Association , 2020).

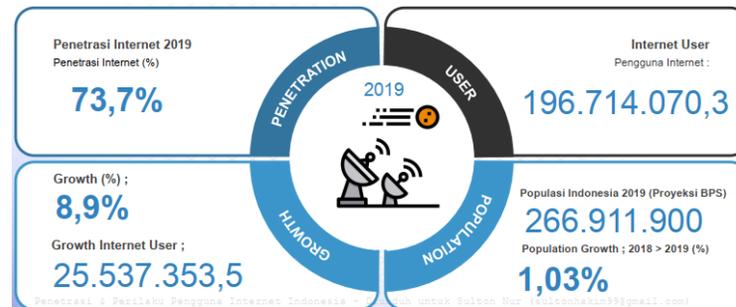


Figure 1. Internet User Penetration on 2019

In Indonesia, the number of internet users is a significant market opportunity for internet service provider companies (Chauhan C., 2017). Various internet service provider companies compete to provide the best service for the community by developing good internet services and following the community's needs (Garcia & Berton, 2020). Private companies and the government also participate in business competition in the telecommunications sector, namely PT Telekomunikasi Indonesia Tbk (PT Telkom). One of the products of PT Telkom, which has the largest market, is IndiHome. IndiHome is a digital service using optical fiber by offering Triple Play services consisting of interactive TV, landlines, and internet with access speeds of up to 100 Mbps. As shown in Table 1, IndiHome has become the most used Internet Service Provider (ISP) brand for five consecutive years.

Table 1. Indonesia Top Brand Index ISP (Indonesia Top Brand Award, 2020)

Brand	Top Brand Indonesia (TBI)				
	2016	2017	2018	2019	2020
IndiHome	48.1%	50.3%	42.1%	39.8%	36.7%
First Media	18.6%	17.3%	22.4	29.9%	23.1%
Biznet	4.7%	2.1%	6.4%	8.9%	8.2%

PT Telkom has developed a mobile application-based service, namely the MyIndiHome application. Through this application, IndiHome customers do not need to come to the nearest Telkom Plaza to get new installations, additional services, billing information, and service interruption reports. Currently, MyIndiHome on Google Playstore has been downloaded more than 5 million with a 3.6 and 98 thousand reviews rating (Indriati & Ridok, 2016). The low rating accompanied by various positive and negative reviews indicates that the services provided by PT

Telkom have not fully met the expectations of IndiHome users. Therefore, it is necessary to evaluate to improve the quality of service to the community. (Diba, 2020).

Reviews of a product/service can be in the form of positive and negative reviews (Sari & Kalender, 2020). The increasing number of MyIndiHome application reviews on the Google Play site makes it difficult for companies to get comprehensive information from these reviews. There needs to be a method that can be used to get information on reviews effectively and efficiently, one of which is Text Mining. One of the Text Mining analysis techniques is sentiment analysis (Ardiansyah, et al., 2020). With sentiment analysis, the company can find out information about the responses and attitudes of a group or individual towards a topic of contextual discussion of the entire document. (Alwasi'a, 2020)

Based on the description above, it is necessary to further analyze the MyIndiHome application reviews on the Google Play site to find out user opinions. Furthermore, opinions will be classified into positive or negative reviews (Fransiska, Rianto, & Gufroni, 2020). It is imperative to conduct this research as it will help understand the effect and impact of word-of-mouth on customers' confidence before using or even buying the product (Kundu & Rajan, 2016). The reviews were evaluated using sentiment analysis using a classification approach because classification is a model that predicts an unknown value (Muhammad & Yan, 2015). Two methods are used, namely the Super Vector Machine (SVM) and Naïve Bayes Classifier (NBC) methods. The negative sentiments will be carried out using a fishbone diagram to determine the factors causing it so that PT Telkom can use the results as a basis for service improvement. This research is expected to present a good and appropriate classification of sentiment analysis to provide useful information for evaluating the performance of IndiHome services.

2. Literature Review

In their research, Garcia & Berton [5] developed the Support Vector Machine, Random Forest, and Logistic Regression methods to see topics that are often discussed and classify emotions regarding the Covid-19 pandemic. The data used is data from social media Twitter which is in English and Portuguese. This study indicates that ten topics are often discussed during the Covid-19 pandemic: economic impacts, case reports, politics, and entertainment (Baid, Gupta, & Chaplot, 2017). In the negative class classification, the three methods have the same F-1, but in the positive class classification, the Linear SVM shows the best performance with an F-1 Score value of 0.66 compared to the other two methods. (Redhu, Srivastava, Bansal, & Gupta, 2018)

Patel & Passi (Patel & Passi, 2020) compared the Naïve Bayes, Support Vector Machine, Random Forest, and K-Nearest Neighbor methods to classify reviews about the 2014 World Cup into three class: positive neutral, and negative. Data taken from Twitter in June-July 2014 and obtained about 2 million tweets. The results showed that the Naïve Bayes method had the best performance because it had the highest accuracy and AUC values compared to other methods.

In a study written by Al-Smadi (Al-Smadi, 2018) conducted a sentiment analysis on hotel review data in Arabia. There is a comparison between two classification methods, namely Recurrent Neural Network (RNN) and Support Vector Machine (SVM) with a dataset of 24,028 reviews. The results showed that the SVM method had a better performance than RNN based on the F-1 value and accuracy with values of 89.9% and 95.4%, respectively.

The results of previous studies show that the performance of the two methods is quite good. However, there has been no research using the Support Vector Machine and Naïve Bayes Classifier methods for sentiment analysis on internet service providers (Oktaviani, Warsito, Yasin,

& Santoso, 2021). Therefore, this research is expected to illustrate and evaluate the performance and services provided, especially the IndiHome services offered by PT Telekomunikasi Indonesia.

3. Research Methodology

3.1 Population and Sample Research

The population in this study is all reviews or reviews of IndiHome service users from the Google Play website database through its application, namely MyIndiHome. As for the sample used is the MyIndiHome application review on November 1, 2020 - December 15, 2020 because that time is the period of using the application version 3.85.005.

3.2 Type and Data Source

The type of data used in this study is primary data. Primary data is data sources that directly provide data to data collectors (Sugiyono, 2015) or data obtained from the first hands. The data was obtained using the scraping method from the MyIndiHome website using the Google Chrome extension, namely Data Scraper. The data obtained are 2,539 reviews.

3.3 Research Variable

Research variables are a trait, characteristic, or phenomenon that can show something to be observed or measured whose values are different (Silaen, 2018). In this study, two variables are used: date (the time of the review collected) and review (the content of user reviews).

3.4 Data Analysis Method

In this study, R Studio software version 1.4.1103 and Microsoft Excel 2016. There are several data analysis methods used in this research, among others (Pratmanto, et al.):

- a. Descriptive analysis is to provide an overview of MyIndiHome reviews on the Google Play site.
- b. Sentiment analysis is to label data into positive and negative sentiment classes (Aaputra, Rosiyadi, Gata, & Husain).
- c. Machine learning methods, namely the Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC), classify positive and negative reviews.
- d. Wordcloud is used to visualize the words that appear most often used in reviews.
- e. The Fishbone diagram identifies the factors that cause problems obtained from negative reviews to solve the problems encountered.

4. Results and Discussion

The data processing in this study was carried out by analyzing the descriptive results of 2,539 reviews of myIndiHome users on the Google Play site. Then the data is processed at the pre-processing stage, which includes translating a foreign language, spelling normalization, case-folding, tokenizing, and filtering. Then, the data can be used in the labeling and text association steps.

4.1. Descriptive Analysis

Descriptive analysis was conducted to find out the general description of the response data about IndiHome that had been obtained. Two aspects are the number of reviews based on a

certain period and the comparison of the number of reasons users are categorized into two categories, namely positive and negative reviews.

The most reviews were obtained in the third week of November, namely on 15-22 November 2020, with a total of 499 reviews. It is estimated because this period is the initial period after the IndiHome service bill payment deadline. After the payment process, the customer hopes that the service obtained can be maximized. But the service that customers get in that period is still the same, and some are even worse than the previous period. Therefore, many customers write their complaints in the comment column of the myIndiHome application.

4.2. Pre-processing Data

Myindihome review data needs to be pre-processing process first because the data that has been obtained is still not structured, there is a lot of noise, and there are foreign languages that need a translation. The pre-processing stage is carried out with the help of the Rstudio-1.4.1103 and Microsoft Excel application. There are no definite provisions regarding what must be done at this stage. It depends on the researcher's needs. Several steps are carried out in this study, such as foreign language translation review, spelling normalization, case-folding, tokenizing, and filtering (Kuntoro, Asra, Pratama, Effendi, & Ocanitra, 2020). After this process is done, the data will become more structured, and it is easier to get information from the existing data.

4.3. Weighing and Labeling of Sentiment Class

The next stage after the pre-processing process is to calculate the weight or sentiment score of each review. One of the commonly used algorithms is lexicon-based. Lexicon-based can extract opinion sentences with very high precision (Azhar, 2017) and fast because it is done automatically. The Lexicon-based method works by creating a dictionary of opinion words (Lexicon) first. The words contained in the dictionary will be used to identify positive and negative comments in a sentence. Word weight is obtained from the frequency of occurrence of words in the document. The more often a word appears in a review, the greater its weight or in other words, the weight of the word is directly proportional to its appearance in user reviews.

After getting the sentiment score, data processing is continued by labeling the sentiment class. Labelling is done by dividing the review data into three classes of sentiment, namely positive sentiment (score > 0), neutral (score = 0) and negative (score < 0) (Buntoro, 2017). However, in this study, there were only two classes used, namely positive and negative classes. This is because the neutral sentiment class provides less information and input for the company (Putri, Khasanah, & 'Azzam, Sentiment Analysis on Grab User Reviews Using Support Vector Machine and Maximum Entropy Methods, 2019).

In this study, class reduction was carried out by manually classifying the sentiment class into positive, negative, and neutral sentiment classes. If the neutral sentiment class is not identified, the positive and negative sentiment words will be included in the positive sentiment class, and if the neutral sentiment contains the balanced positive and negative sentiment words, it will be included in the negative sentiment class (Gumilang, 2018). Previously this was done by considering that negative information can be extracted more easily to be translated as user complaints or dissatisfaction. It can be input for Telkom to make improvements in a better direction. Based on Figure 2., it is obtained that the result of labeling the sentiment class with positive reviews has more frequency than negative reviews. From a total of 2,539 reviews, there were 1374 positive reviews (54.22%) and 1160 negative reviews (45.78%).

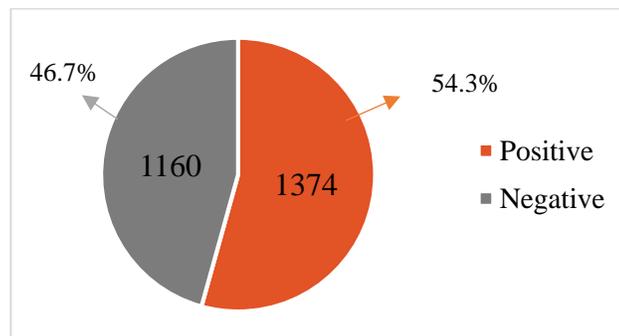


Figure 2. Sentiment Class Labeling After Reduction

4.4. Classification Analysis

After carrying out the class labeling stage, the data processing is carried out by classification analysis. The results of the labeling data are divided into two, namely training data and testing data. The classification process uses two algorithms, namely Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC).

The review data will be divided into two data, namely training data and testing data. Training data is used to form a classification model. This model is a knowledge representation that will be used to predict new data classes. Then, the testing data is used to measure the performance of the model that has been obtained. Based on the Pareto Principle, the ratio commonly used for training data and testing data is 80:20. However, it is possible in a study not only to use these comparisons. It is because the amount of training data will affect accuracy. The more training data, the more the model will learn so that the accuracy will be better (Arrofiqoh & Harintaka, 2018)

In this study, three comparative values of training and testing data were used. The following is a comparison of the amount of training data and testing data as presented in Table 2, Table 3, and Table 4.

- a. Comparison of training data is 70%, and testing data is 30%.

Table 2. First Comparison

Data Type	Percentage	Total
Training Data	70%	1773.8 \approx 1.774
Testing Data	30%	760,2 \approx 760

- b. Comparison of training data is 80%, and testing data is 20%.

Table 3. Second Comparison

Data Type	Percentage	Total
Training Data	80%	2027.2 \approx 2.027
Testing Data	20%	506.8 \approx 507

- c. Comparison of training data is 90%, and testing data is 10%:

Table 4. Third comparison

Data Type	Percentage	Total
Training Data	90%	2280.6 \approx 2.281
Testing Data	10%	253.4 \approx 253

In the SVM method, several kernels are employed, such as the Linear kernel, Polynomial, Radial Basis Function (RBF), and Sigmoid. From each kernel, the best accuracy will be sought which will later be used to compare accuracy with the NBC method. Comparison of the kernels that have been tested as shown in Table 5.

Table 5. Comparison of Accuracy Results on SVM Classification

Kernel	Accuracy
Linear	88.95%
Polynomial	64.10%
RBF	87.97%
Sigmoid	85.75%

Table 4 presents the Linear kernel that has the highest accuracy compared to other kernels. Therefore, the Linear kernel will be used in the SVM method classification process for the next classification process. The classification process is carried out by making a confusion matrix to determine the level of accuracy, recall, and precision. This matrix is used to evaluate the performance of the model formed by each classification algorithm. In assessing the model, five trials of the dataset were carried out to get the best accuracy value (Putri, 2019).

In this study, five trials were conducted for each data set. The difference in each experiment is in the number of times the data set is randomized by making pseudorandom numbers n times using the `set.seed(n)` formula in R software. Summary of the accuracy values of each experiment in this study as shown in Table 6.

Table 6. Accuracy Comparison of SVM and NBC Model

Trial	Model Accuracy	
	SVM	NBC
Comparison 70:30		
Trial 1	87.24%	84.61%
Trial 2	87.24%	84.34%
Trial 3	87.5%	83.95%
Trial 4	87.37%	83.68%
Trial 5	87.37%	84.47%
Average	87.34%	84.21%
Comparison 80:20		
Trial 1	89.35%	86.59%

Trial	Model Accuracy	
	SVM	NBC
Trial 2	89.15%	84.81%
Trial 3	88.95%	84.22%
Trial 4	88.95%	84.81%
Trial 5	89.15%	82.84%
Average	89.11%	84.65%
Comparison 90:10		
Trial 1	86.17%	90.11%
Trial 2	86.17%	83%
Trial 3	86.17%	84.98%
Trial 4	86.17%	85.38%
Trial 5	86.17%	82.6%
Average	86.17%	85.2%
Total Average	86.54%	84.69%

Based on Table 5, the accuracy value with the SVM method is greater than the NBC method for all trials except for the first experiment with a ratio of 90:10. In addition, the average value of the total accuracy of the SVM method is 86.54% higher than the NBC method, which has an average total accuracy of 84.69%. Therefore, it can be concluded that the SVM algorithm has a better performance in classifying myIndiHome review data than the NBC method.

4.5. Visualization and Text Associations

Visualization is done to get information about topics that users in the myIndiHome comment column often discuss. In addition, this study also searched for text associations that most often appear together to strengthen and clarify the information obtained in the visualization process. The following is the result of text association for positive and negative reviews.

a. Positive Review

In the classification results of positive reviews of the myIndiHome application as shown in Figure 3, from the number of positive reviews as many as 1.379 reviews, it was obtained that several words appeared the most, such as the word "application" with a frequency of 505 times, "good" 252 times, "help" 181 times, "great" 162 times, "fast" 124 and so on.

The following is information obtained based on the results of text associations in positive reviews. myIndiHome has an excellent, complete, and convenient interface to use when opening it. During the pandemic, this application is beneficial for users in requesting services, from installation to complaints of disturbances. Once updated, the app has a nice and neat design. In addition, users rate employees have done an excellent job which is indicated by fast and courteous service. The application also provides complete information such as billing details, usage, service changes, and rewards. The speed of IndiHome also increased and was compatible after a complaint was made.

a. Negative Review

In the negative reviews, it was found that several words that appeared the most with topics considered relevant as negative sentiments, such as the word "application" with a frequency of 264 times, "payment" 218 times, "network" 191 times, "wifi" as many as 174 times. times, "slow" 148 times and so on (see Figure 4).

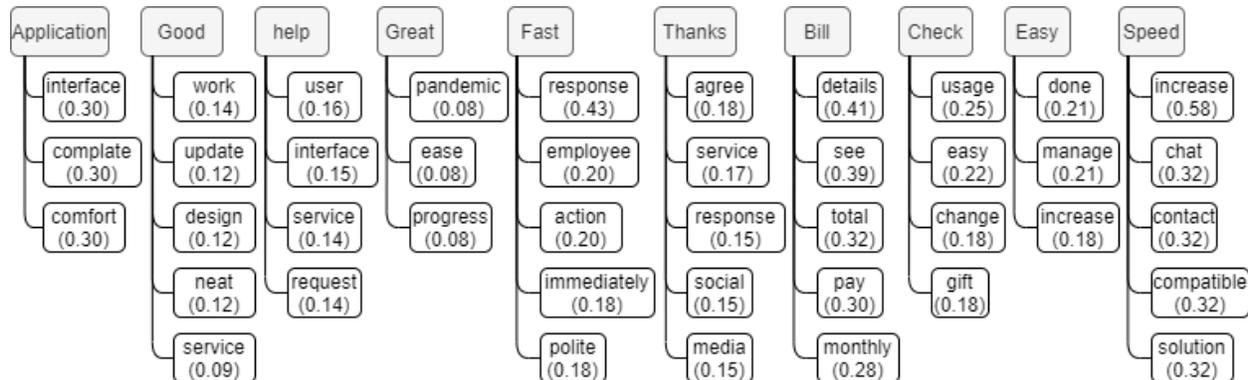


Figure 3. Text Association on Positive Reviews

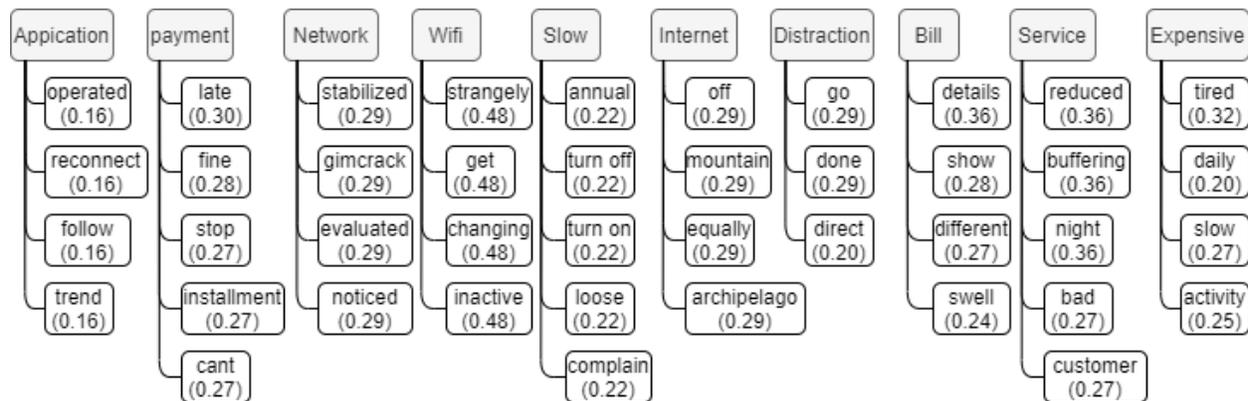


Figure 4. Text Association on Negative Reviews

The following is information obtained based on the results of text associations in negative reviews. When operated by the user, the application cannot be connected and is considered cheap because of frequent errors. If the user is late in making a payment, a fine will be imposed, even if the network is terminated immediately and the fine must be paid in cash. The users' network get is cheap and unstable, so it needs to be stabilized, cared for, and evaluated. Users get wifi speeds that alternate or don't match what was promised and sometimes get inactive wifi. Even though it has been a subscription for years, the network speed is still slow. There is a loose cable. The network is disconnected. When in the mountains, it shows that the IndiHome network is not evenly distributed throughout the archipelago. If there is a network interruption, the user will come to the plaza to immediately resolve the disturbance. It is because the complaint in the application has not received a response. The service provided has been reduced, which is

indicated by a bad network or buffering at night. Users give up using IndiHome because they paid a lot, but the internet connection they get for a day is still slow and impacts customer activities.

4.6. Factors for Improving Negative Review myIndiHome Problem

The information obtained from the negative review association is used to determine the problems that caused myIndiHome to get a negative review. Analysis of the problem using a Fishbone Diagram (FD). FD can make complex systems organized, analyzing the causes of risk qualitatively (Luo, Wu, & Duan, 2017). The results obtained 12 problems classified into 5P factors: price, people, process, place, and product. Then the problem-solving plan is given based on the existing issues, as shown in Table 7. It shows the proposed improvements for each problem. For instance, problem bills swell at the Price factor. The company can provide periodic warnings when approaching the payment limit.

Table 7. Plan to Solve The myIndiHome Problems of Negative Review

Factor	Problem	Improvement Proposal
Price	Bills swell	Provide periodic warnings when approaching the payment limit.
	Expensive service price	Provide promos on big days, such as Christmas and New Year.
People	Unilateral disconnection	Make standard operations procedures regarding network disconnection, including warnings when network disconnection will be carried out.
	Bad customer service at night	Create a 24-hour complaint system on the application.
Process	Complaints in the application no response	Adding customer service employees who are in charge of serving customers via applications; The developer immediately fixes the application system so that every complaint can be received and read by officers
	Payment must be cash	Create a payment system that can ease users, such as being able to be paid in stages.
	There are no billing details in the application	The developer must immediately fix the application system. Therefore, the details of each bill can appear in the application and be known to the user.
Place	Loose cable	Control the cable quality regularly, especially after extreme weather.
	The network is not evenly distributed throughout the area	Create a network expansion program in each location; Conducting promos, especially for areas that are still rarely used, to encourage the growth of service users.
Product	Unstable network	Provide time notification if there is maintenance so that consumers can understand; Controlling internet quality regularly, especially after bad weather such as heavy rain and wind.
	The application cannot log in	The developer immediately fixes the application system so that users can log in.

Factor	Problem	Improvement Proposal
	Speed does not match.	Perform periodic control of the speed obtained by the user; Take decisive action against parties who steal users' internet networks.

5. Conclusion

The number of myIndiHome reviews on November 1, 2020 - December 15, 2020. There are 2539 user reviews with a rating 3 out of 5 stars. Based on sentiment class labeling, the number of positive reviews was 1,374, and negative reviews were 1,160. Based on this explanation, information can be obtained that the sentiment analysis results are relevant and have interpreted the rating received by the myIndiHome application. The classification results show that the average value of the total accuracy of the Support Vector Machine (SVM) method uses Linear Kernel is 86.54% greater than the Naïve Bayes Classifier (NBC) method, which has an average total accuracy of 84.69%. This value aligns with several previous studies that showed the SVM method had better performance than the NBC method. Based on the classification and text associations results conducted, most myIndiHome users talk about application, service, speed, connection, and bill. Based on fishbone diagram analysis, there are 12th problems on negative reviews by myIndiHome users. We classify these problems into 5P factors, namely: Price, People, Process, Place, and Product.

References

- Aaputra, S. A., Rosiyadi, D., Gata, W., & Husain, S. M. (2019). Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization. *Jurnal Rekayasa Sistem dan Teknologi Informasi*. 3(3), 377-382.
- Al-Smadi, M. (2018). Deep Recurrent Neural Network vs. Support Vector Machine for Aspect-based Sentiment Analysis of Arabic Hotels' Reviews. *Journal of Computational Science*, 27, 386-393.
- Alwasi'a, A. (2020). *Analisis Sentimen Pada Review Aplikasi Berita Online Menggunakan Metode Maximum Entropy (Studi Kasus: Review Detik.com pada Google Play 2019) [Skripsi]*. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Ardiansyah, A., Aji, S., Pratmanto, D., Kuryanti, S. J., Putra, O. P., & Adiwihardja, C. (2020). The Analysis of Digital Provider Sentiment of 'by.u' on Google Play Which Uses The Support Vector Machine (SVM) Method. *Journal of Physics: Conference Series*. 1641, 1-6.
- Arrofiqoh, E. N., & Harintaka. (2018). Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi. *Jurnal Ilmiah Geomatika*, 24(2), 61-68.

- Azhar, Y. (2017). Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia. *Jurnal Nasional Pendidikan Teknik Informatika*, 6(3), 237-243.
- Baid, P., Gupta, A., & Chaplot, N. (2017). Sentiment analysis of movie reviews using machine learning techniques. *International Journal of Computer Applications*. 7(1), 383-385.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. *INTEGER: Journal of Information Technology*, 2(1), 32-41.
- Chauhan C., S. S. (2017). Sentiment Analysis On Product Reviews. *International Conference on Computing, Communication and Automation (ICCCA)*. 26-31.
- Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. (2016). A Novel Text Mining Approach Based on TF-IDF and Support Vector Machine for News Classification. *IEEE International Conference on Engineering and Technology (ICETECH)*. 112-116.
- Diba, S. F. (2020). Implementation of Naive Bayes Classification Method for Sentiment Analysis on Community Opinion to Indonesian Criminal Code Draft. *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*. 474, 186-192.
- Fransiska, S., Rianto, R., & Gufroni, A. I. (2020). Sentiment Analysis Provider by. U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*. 7(2), 203-212.
- Garcia, K., & Berton, L. (2020). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing Journal*. 101, 1-25.
- Gumilang, Z. (2018). *Implementasi Naive-Bayes Classifier dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-commerce Shopee pada Situs Google Play [Skripsi]*. Yogyakarta: Program Studi Statistika FMIPA Universitas Islam Indonesia.
- Handani, S. W., & Saputra, D. I. (2019). Sentiment Analysis for Go-Jek on Google Play Store. *Journal of Physics: Conference Series*. 1196.
- Indonesia Top Brand Award. (2020). *Tob Brand Index*. Retrieved Desember 20, 2020, from https://www.topbrand-award.com/en/top-brand-index-int/?tbi_find=IndiHome
- Indonesian Internet Service Providers Association . (2020, Juni). *Laporan Survey Internet APJII 2019-2020 (Q2)*. Retrieved Desember 12, 2020, from <https://apjii.or.id/survei2019x/download/BRTdtugLoJEG1wWiSrKZxeb0HDCQ5>
- Indriati, I. & Ridok, A. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwkn). *Journal of Environmental Engineering and Sustainable Technology*, 3(1). 23-32.
-

- Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm And Support Vector Machine Based Smote Technique. *Journal of Physics: Conference Series*, 1641.
- Luo, T., Wu, C., & Duan, L. (2017). Fishbone Diagram and Risk Matrix Analysis Method and Its Application. *Journal of Cleaner Production*, 174, 296-304.
- Muhammad, I., & Yan, Z. (2015). Supervised Machine Learning Approach: A Survey. *International Journal of Soft Computing*, 5(3), 946-952.
- Oktaviani, V., Warsito, B., Yasin, H., & Santoso, R. (2021). Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method. *Journal of Physics: Conference Series*, 1943.
- Patel, R., & Passi, K. (2020). Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. *Internet of Things Journal*, 1(2), 218-239.
- Pratmanto, D., Rousyati, R., Wati, F., Widodo, A. E., Suleman, S., & Wijianto, R. (2020). App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, 1641.
- Putri, B. A. (2019). *Analisis Sentimen Data Ulasan Pengguna Grab Menggunakan Metode Support Vector Machine dan Maximum Entropy*. [Skripsi]. Yogyakarta: Fakultas Teknologi Industri Universitas Islam Indonesia.
- Putri, B. A., Khasanah, A. U., & 'Azzam, A. (2019). Sentiment Analysis on Grab User Reviews Using Support Vector Machine and Maximum Entropy Methods. *International Conference on Information and Communications Technology*, 468-473.
- Redhu, S., Srivastava, S., Bansal, B., & Gupta, G. (2018). Sentiment analysis using text mining: a review. *International Journal on Data Science and Technology*.
- Sari, S. & Kalender, M. (2020). Sentiment Analysis and Opinion Mining Using Deep Learning for the Reviews on Google Play. *The Proceedings of the Third International Conference on Smart City Applications*. 4, 126-137.
- Silaen, S. (2018). *Metodologi Penelitian Sosial Untuk Penulisan Skripsi dan Tesis*. Bandung: In Media.
- Sugiyono. (2015). *Metode Penelitian Kombinasi (Mix Methods)*. Bandung: Alfabeta.

This page is intentionally left blank.