PAPER NAME

senti.pdf

AUTHOR

suswadi suswadi

WORD COUNT

5089 Words

CHARACTER COUNT

27514 Characters

PAGE COUNT

9 Pages

FILE SIZE

500.7KB

SUBMISSION DATE

Feb 27, 2024 11:52 AM GMT+7

REPORT DATE

Feb 27, 2024 11:53 AM GMT+7

● **4% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- Crossref Posted Content database
- 4% Submitted Works database

● **Excluded from Similarity Report**

- Internet database
- Crossref database
- Quoted material
- Small Matches (Less then 10 words)

- Publications database
- Bibliographic material
- Cited material

Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine

Suswadi [a,1]; Moh. Erkamim [a,2,*]

[a] *Universitas Tunas Pembangunan, Jalan Balekambang Lor 1 Manahan Surakarta 57139, Indonesia*
[1] suswadi@lecture.utp.ac.id; [2] erkamim@lecture.utp.ac.id
* Corresponding author

Abstract

During the COVID-19 outbreak, Indonesian marketplaces were significantly impacted including Shopee app. It is necessary to evaluate the features and services of the Shopee application by looking at the feedback given by the public in Google Play Store reviews. This is what prompted research to be conducted from Kaggle data in the form of Shopee reviews. From this data, sentiment analysis is carried out utilizing the Support Vector Machine (SVM) and Random Forest methods. This method are used to classify reviews based on positive and negative sentiments. The results showed that the level of classification accuracy in the Random Forest model is 82.21%. While the SVM model provides a higher level of accuracy of 84.71%. Data exploration on positive and negative sentiment classes is used to find insight into this problem. In positive sentiment, words that often appear such as "*belanja*", "*aplikasi*", and "*barang*" are found. As for the negative sentiments, namely "*ongkir*", "*kirim*", "*aplikasi*". These words can be used to be a quality improvement or evaluation for the Shopee company.

Introduction

The Covid-19 pandemic has had a significant impact on various sectors in Indonesia, one of which is the industrial and business sectors [1], [2]. The reduction in buying and selling activities has directly caused business and industry players to adapt by developing their businesses through the marketplace [3]. One of the marketplaces in Indonesia is Shopee. Based on the results of the 2019-2020, internet survey by the Association of Indonesian Internet Service Providers (APJII), Shopee ranks first in e-commerce that is often used by Indonesians with a percentage of 27.4% [4]. The Shopee application can be downloaded via the App Store, Google Play Store, or App Gallery. The more Shopee users, the better and bad reviews or feedback from the public regarding this application. Community reviews are important for evaluating application features and services so they can get better. In addition, it is also a consideration for the public to download the Shopee application. For the review to become useful information, it is necessary to analyze it. One way to analyze reviews is to do a sentiment analysis [5]–[7].

Sentiment analysis is defined as machine learning to extract subjective information and opinions regarding emotions, attitudes, and moods, and order to calculate the polarity of the text by the text [8]. Sentiment analysis can be used to determine a value. This value can be either positive or negative. This value can be used as a parameter in decision-making [9]–[11]. Sentiment analysis can be performed using several methods such as Random Forest, Support Vector Machine (SVM), and Naïve Bayes [5], [6], [12]. Random Forest is an ensemble learning algorithm whose classification is based on the results obtained from a set of decision trees [13]–[15]. The number of trees can reduce the variance in the overall model and control overfitting. And the more trees there are in the random forest, the more accurate the results [10], [16]. In the past few years, Random Forest has become more popular for classifying data because it works better than SVM, Naïve Bayes, and other machine learning algorithms [5], [17]. SVM is supervised learning algorithm utilized for classification, regression, and the discovery of anomalies or outliers. This SVM model requires a parameter for implementation. Due to the accuracy of the model is very dependent on these parameters, therefore the optimal value will be obtained from the appropriate grid-search technique [18]. According to [19], the SVM approach generates higher accuracy compared to other methods, has high classifying speed, and can tolerate irrelevant attributes.

The method proposed by researchers is the application of Random Forest and SVM techniques to carry out sentiment analysis on reviews of the Shopee application. There are to methods that focus on the selection of algorithms

or data analysis techniques used to achieve research objectives. In this context, researchers compare the performance of the two methods and choose the best model based on the results of this comparison. However, it is important that good research also requires more in-depth methodological steps, such as data collection, visualization and exploration, feature extraction, model validation, performance evaluation, and interpretation of results. Apart from that, this research also has a clear objective, namely helping the Shopee company improve the quality of its services based on insights from sentiment analysis. Therefore, apart from applying the method, this research method also includes in-depth data analysis and applying the results to provide useful recommendations for the company.
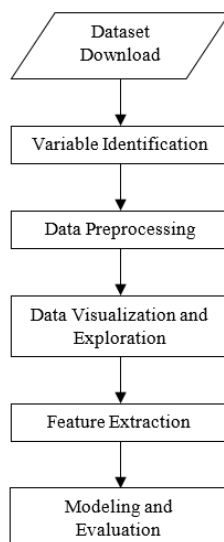
Method

Figure 1. System Main Flowchart

Figure 1 shows the procedures used in this study started with dataset input, variable identification, data preprocessing, data visualization and exploration, feature extraction, modeling, and evaluation.

A.　Dataset Download

The dataset is downloaded via Kaggle regarding Shopee reviews on the Google Play Store from January-March 2021 [12], [16]. This data is in the form of Comma Separated Values (CSV), which is a database where the characteristics of each record are separated by a comma (,) or semicolon (;). The data has variable identification as shown in Table 1.

Table 1. Data Identification

Variable Name	Description
User	User or reviewer name
Review	A collection of words or sentences that delve into the Shopee application, containing satisfaction, compliment, and support, while also encompassing negative sentiments when expressing complaints, criticisms, and disappointments regarding the Shopee app
Label	The characteristics of the reviews, 1 for positive reviews and 0 for negative reviews.
Date	The recorded time of the user when submitting a review
Case Folding	The collection of sentences that have been converted to lowercase.
Punctuation	The set of sentences that have had their punctuation removed
Remove Emoji	The collection of sentences that have had emoticons removed

B.　Data Preprocessing

1. Removing unused columns, namely column user, date, case folding, punctuation, and remove emoji. Therefore, the remaining column of reviews and labels is used for the next stage.

2. Cleaning data by removing HyperText Markup Language (HTML) links, emoji, mentions, usernames, hashtags, and punctuation and removing strips of white space such as redundant spaces, tabs, and newlines.

3. Conducting case folding and stop-word removal, namely changing all letters to lowercase and removing unnecessary words. Then stemming or changing the basic words to words based on the literary package is carried out.

4. Removing duplicate data, therefore, data is not biased in modeling.

5. Converting imbalanced data into balanced data using a random sampling method to equalize the number of positive and negative reviews.

6. Doing tokenization, namely changing reviews in the form of sentences or paragraphs into certain words or parts.

C. Data Visualization and Exploration

Performing data visualization using the most word frequent and word cloud plots is carried out to find out the words that appear the most in positive or negative sentiment. This process is carried out to get an insight that can be useful for the Shopee application.

D. Feature Extraction

Feature extraction with Bag of Words (BoW) with the CountVectorizer function is performed to convert text data to numbers in the form of sparse matrices. Next, the weighting is done using the TF-IDF for the process of assigning weights to the association between a phrase and a document based on the concept of the frequency occurrence of words in a document and the inverse of the frequency of documents containing that term.

E. Modeling and Evaluation

1. Splitting the data into training data and test data with a ratio of 80% training data and 20% test data using the train-test split function of the sci-kit-learn package.

2. Modeling the classification of positive and negative sentiments using the Random Forest and SVM methods.

3. Performing hyperparameter tuning by selecting the best model in the GridSearch function.

4. Performing cross-validation of data to reduce model bias.

5. Evaluating the classification method using accuracy metrics

Results and Discussion

In this study, the dataset was collected via Kaggle regarding Shopee reviews from January-March 2021 [16]. There were two classes, namely positive sentiment reviews, and negative sentiment reviews. Positive sentiment reviews contain praise and support. While negative sentiment reviews contain complaints, satire, and disappointment towards the Shopee application.

A. Data Collection

After downloading the dataset from [16], 2937 review data in Indonesian were obtained about Shopee, of which 1668 were positive sentiment reviews labeled 1 and 1269 were negative sentiment reviews labeled 0. Here are some sample review data from the Shopee application shown in **Table 2**.

Table 2. Multiple Sample Data Review

No	Review	Label
1.	*sangat membantu skali suka bgt makasihmakin kesini makin mantap â¤ï¸â¤ï¸â¤ï¸â¤ï¸ â¤ï¸â™¥ï¸â™¥ï¸â™¥ï¸ â™¥ï¸ mah voucer gratis ongkir tambahin dong buat member platinumulasan saya kemaren kemaren mantap terus tapi kok sdh seminggu ini voucer saya batasi sehari cuma bisa dua sekarang malah udh pake voucer alasa limit sdh terpenuhi kaya susah bgt order pake voucer pdhal vocer bkn suka*	Positive
2.	*pilihan jasa kirim sekarang ribet jasa kirim tentuin pihak alhasil beli apa krn sebagian brang tdk memilih jasa kirim â€â™€ï¸Haduhaduhð Ÿ˜*	Negative

B. Data Preprocessing

The purpose of preprocessing the data is to eliminate noise, clarify features, improve precision, and convert the original data to be processed according to research requirements. All machine learning algorithms, whether supervised or unsupervised, begin with preprocessing before data processing and analysis [20]. Text preprocessing consists of case folding, punctuation removal, stop-words removal, tokenization, and stemming. Here are some steps in data preprocessing:

1. Removing unused columns, namely column user, date, case folding, punctuation, and remove emoji. Therefore, the remaining column of reviews and labels is used for the next stage.

2. Cleaning review :

a. Removing HTML links in each review and emoji icon. When someone gives a review regarding an application or an item, intentionally or not, they sometimes use an emoji icon describing the feelings they are experiencing at that time. Review after removing the HTML link and emoji icon shown in **Table 3**.

Table 3. Review After Remove HTML Link and Emoji Icon

No	Before	After
1.	*sangat membantu skali suka bgt makasihmakin kesini makin mantap â¤ï¸â¤ï¸â¤ï¸â¤ï¸â¤ï¸ â™¥ï¸â™¥ï¸â™¥ï¸â™¥ï¸ mah voucer gratis ongkir tambahin dong buat member platinumulasan saya kemaren kemaren mantap terus tapi kok sdh seminggu ini voucer saya batasi sehari cuma bisa dua sekarang malah udh pake voucer alasa limit sdh terpenuhi kaya susah bgt order pake voucer pdhal vocer bkn suka*	*sangat membantu skali suka bgt makasihmakin kesini makin mantap mah voucer gratis ongkir tambahin dong buat member platinumulasan saya kemaren kemaren mantap terus tapi kok sdh seminggu ini voucer saya batasi sehari cuma bisa dua sekarang malah udh pake voucer alasa limit sdh terpenuhi kaya susah bgt order pake voucer pdhal vocer bkn suka*
2.	*pilihan jasa kirim sekarang ribet jasa kirim tentuin pihak alhasil beli apa krn sebagian brang tdk memilih jasa kirim â€â™€ï¸Haduhaduhð*Ÿ˜,*	*pilihan jasa kirim sekarang ribet jasa kirim tentuin pihak alhasil beli apa krn sebagian brang tdk memilih jasa kirim Haduhaduh*

b. Removing mentions, usernames, hashtags, numbers, punctuation, and dashes of white space. Review after removing mentions, usernames, hashtags, numbers, punctuation, and white space strips shown in **Table 4**.

Table 4. Review After Removing Mentions, Usernames, Hashtags, Numbers, and White Space Strips

No	Before	After
1	*sangat membantu skali suka bgt makasihmakin kesini makin mantap mah voucer gratis ongkir tambahin dong buat member platinumulasan saya kemaren kemaren mantap terus tapi kok sdh seminggu ini voucer saya batasi sehari cuma bisa dua sekarang malah udh pake voucer alasa limit sdh terpenuhi kaya susah bgt order pake voucer pdhal vocer bkn suka*	*sangat membantu skali suka bgt makasihmakin kesini makin mantap mah voucer gratis ongkir tambahin dong buat member platinumulasan saya kemaren kemaren mantap terus tapi kok sdh seminggu ini voucer saya batasi sehari cuma bisa dua sekarang malah udh pake voucer alasa limit sdh terpenuhi kaya susah bgt order pake voucer pdhal vocer bkn suka*
2	*pilihan jasa kirim sekarang ribet jasa kirim tentuin pihak alhasil beli apa krn sebagian brang tdk memilih jasa kirim Haduhaduh*	*pilihan jasa kirim sekarang ribet jasa kirim tentuin pihak alhasil beli apa krn sebagian brang tdk memilih jasa kirim Haduhaduh*

3. Case folding, stemming and filtering with stop-word.

A list of stop-words is a compilation of words that are often used yet have been deemed to be devoid of significance because of their frequency. This deletion serves the purpose of reducing the total number of words as well as eliminating noise. Review after case folding, stemming and removing stop-words shown in **Table 5**.

Table 5. Review After Case Folding, Stemming, and Removing Stop-words

No	Before	After
1	*sangat membantu skali suka bgt makasihmakin kesini makin mantap mah voucer gratis ongkir tambahin dong buat member platinumulasan saya kemaren kemaren mantap terus tapi kok sdh seminggu ini voucer saya batasi sehari cuma bisa dua sekarang malah udh pake voucer alasa limit sdh terpenuhi kaya susah bgt order pake voucer pdhal vocer bkn suka*	*sangat bantu suka makasihmakin kesini makin mantap voucer gratis ongkir tambah buat member platinumulasan saya kemaren kemaren mantap terus minggu voucer saya batas hari cuma bisa dua sekarang malah voucer alasa limit penuh kaya susah order voucer vocer suka*
2	*pilihan jasa kirim sekarang ribet jasa kirim tentuin pihak alhasil beli apa krn sebagian brang tdk memilih jasa kirim Haduhaduh*	*pilih jasa kirim sekarang ribet jasa kirim tentu pihak hasil beli bagi brang pilih jasa kirim*

4. Removing duplicate data

After dropping duplicates, 1835 out of 2937 reviews contained positive or negative sentiments regarding Shopee reviews in Indonesia.

5. Converting imbalance data to balance data

In this study, the data which originally consisted of 1668 positive reviews and 1269 negative reviews were converted into balanced data with a data ratio of 801:801 in which the collection was carried out using the random sampling method so that the number of positive reviews was the same as negative. Pie charts of imbalanced data are shown in **Figure 2**.
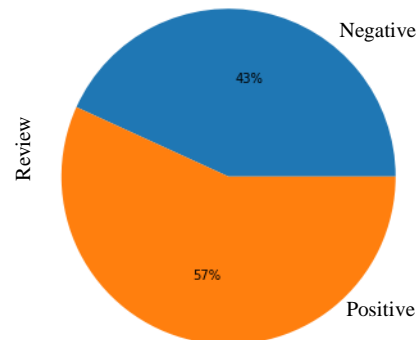

Figure 2. Pie Chart Data Imbalance

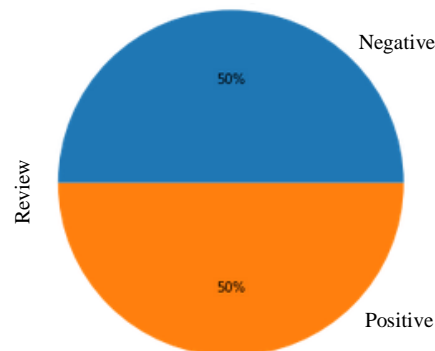The following is a pie chart resulting from a random sampling of data that becomes a balance shown in **Figure 3**

Figure 3. Pie Chart Data Balance

6. Tokenization. Review after tokenization is shown in **Table 6**.

No	Before	After
1	*sangat bantu suka makasihmakin kesini makin mantap voucer gratis ongkir tambah buat member platinumulasan saya kemaren kemaren mantap terus minggu voucer saya batas hari cuma bisa dua sekarang malah voucer alasa limit penuh kaya susah order voucer vocer suka*	"sangat", "bantu", "suka", "makasih", "makin", "kesini", "makin", "mantap", "voucer", "gratis", "ongkir", "tambah", "buat", "member", "platinum", "ulasan", "saya", "kemaren", "kemaren", "mantap", "terus", "minggu", "voucer", "saya", "batas", "hari", "cuma", "bisa", "dua", "sekarang", "malah", "voucer", "alasa", "limit", "penuh", "kaya", "susah", "order", "voucer", "vocer", "suka"
2	*pilih jasa kirim sekarang ribet jasa kirim tentu pihak hasil beli bagi brang pilih jasa kirim*	"pilih", "jasa", "kirim", "sekarang", "ribet", "jasa", "kirim", "tentu", "pihak", "hasil", "beli", "bagi", "brang", "pilih", "jasa", "kirim"

C. *Visualization and Data Exploration*

Data visualization and exploration aim to describe the characteristics or general picture of the results of the analysis to gain insights that can be useful for the Shopee application and the public in choosing the best marketplace.

1. Data Visualization with the Most Word Frequently, shown in **Figure 4**.
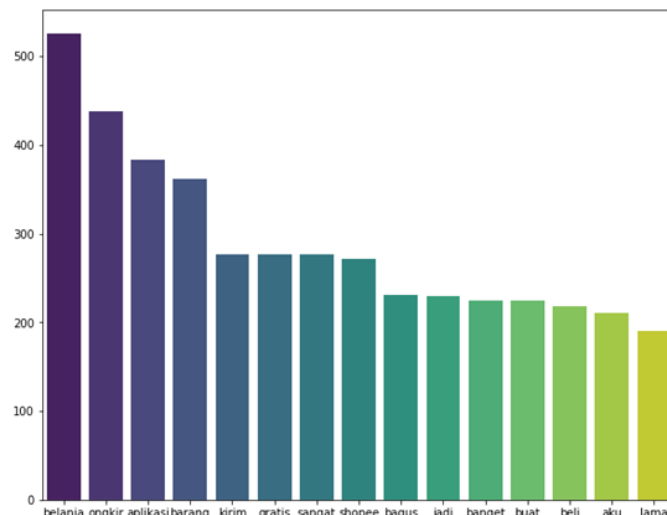
Figure 4. Plot of Most Word Appeared Frequently

2. Data Visualization with Word Cloud

   a) Positive Sentiment

   In the positive sentiment word cloud visualization, it was found that three words appeared the most frequently besides keywords, namely "*belanja*", "*aplikasi*", and "*barang*" with respectively appearing 394, 231, and 217 frequencies. This shows that "*belanja*" is one of the keywords that trigger positive sentiment. A positive sentiment word cloud is shown in **Figure 5**.

Figure 5. Positive Sentiment Word Cloud

   The word "*belanja*" meaning shopping is one of the dominant words in positive sentiment. In other words, the Shopee application provides convenience in shopping. Because the features in it are very varied, ranging from shopping for food products, cosmetics, pulses, and even the newest feature, namely the Shopee Food feature. This feature makes it easy for people to order food online. The word "*aplikasi*" meaning application is the word that has the second most frequency after the word "*belanja*". It can be seen that the public already understands and believes that Shopee is an application engaged in the business sector which provides convenience in terms of goods transactions. So it is hoped that Shopee will maintain public trust by improving the quality of the application system. The third highest frequency is the word "*barang*" meaning goods. In the Shopee application, users can find various kinds of goods needed in everyday life. It is hoped that with the customer's trust in the ease of buying goods, Shopee can continue to enhance the standard of service in the process of buying and selling goods and daily necessities.

   b) Negative Sentiment

   In the depiction of negative sentiment word cloud, it was found that three words appeared the most frequently other than keywords, namely "*ongkir*", "*kirim*", and "*aplikasi*" with respectively appearing 253, 185, and 162 frequencies. This shows that "*ongkir*" meaning delivery cost is one of the keywords that trigger negative sentiment. The negative sentiment word cloud is shown in **Figure 6**.
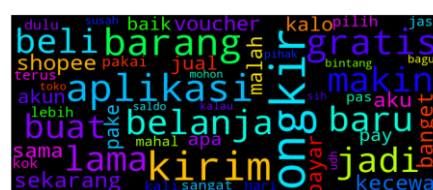

Figure 6. Negative Sentiment Word Cloud

The word "*ongkir*" is one of the dominant words in negative sentiment. It can be said that the community is very concerned about postage issues in terms of choosing a marketplace. Expensive, complicated shipping costs or the absence of vouchers can cause people to be reluctant to use the Shopee application and can switch to other marketplaces. The word "kirim" meaning delivery is the word that has the second highest frequency of negative sentiment. It can be said that reviews regarding the Shopee application talk a lot about shipping issues. Problems related to shipping that is now being managed by Shopee, such as long delivery times and no other delivery service options. This often makes Shopee users uncomfortable. As a result, users provide negative reviews about this, so it is hoped that Shopee can review shipping again so that Shopee users do not switch to other marketplaces. The third highest frequency of negative sentiment is the word "*aplikasi*". This word is also found in positive sentiments, but in this case, the word "*aplikasi*" can have a negative meaning. The Shopee application makes it easy to buy and sell goods, but some people give negative reviews regarding the application system which often has errors. This can be a constructive review for Shopee so that they can immediately take action on a system that often experiences errors.

D. Doing Bag-of-Words (BoW) and Weighting with TF-IDF

This research utilized the BoW model where based on the review data sample, a list is formed for each document shown in **Table 7**.

Table 7. List in Document Form

Review	Term
sangat membantu suka, ... voucer gratis ongkir ... voucer ... suka	*"sangat", "bantu", "suka", ..., "voucer", "gratis", "ongkir", ... "voucer", ..., "suka"*
sangat membantu suka, ... voucer gratis ongkir ... voucer ... suka	*"sangat", "bantu", "suka", ..., "voucer", "gratis", "ongkir", ... "voucer", ..., "suka"*

Then for each word/term, the frequency is calculated and mapped back into the document. Word frequency in documents is shown in **Table 8**.

Table 8. Word Frequency in Document

Sangat	*Bantu*	*Suka*	*Voucer*	*Gratis*	*Ongkir*
1	1	2	2	1	1

Evaluation to determine the significance of a word or phrase in a corpus or collection of documents is also a step in text mining and information retrieval. This highlights the terms in documents that recur frequently in the corpus as a 1602 x 3782 Sparse matrix. In addition, the variable reduction method employed in this study is TF-IDF weighing.

E. Model Classification and Evaluation

The method proposed in this research involves combining Random Forest and SVM classification methods using a pipeline. This process involves hyperparameter tuning stages to optimize the performance of the two methods. The optimum parameters were selected by considering the use of the BoW gram range (1,1) and (1,2), as well as the use of the TF-IDF feature with True and False Boolean settings. In the context of the Random Forest method, experiments were also carried out using the classifier criterion Gini and entropy, while the SVM technique used a linear kernel. To maximize model quality, this research also includes the use of GridSearchCV with a combination of 10 times cross-validation and 1 verbose to monitor iterations. As the main evaluation metric, researchers chose accuracy scores as the main indicator to compare the performance of these two classification models. The results of this comparison show that the best model was found using the SVM method, with an accuracy level of 84.71%. The research also includes a comparison of accuracy values from cross-validation experiments between Random Forest and SVM, which are presented in **Table 9** as qualitative indicators that help in evaluating the performance of the two classification methods.

Table 9. Comparison of Random Forest and SVM Accuracy Values Using Cross-Validation

Iteration	Random Forest	SVM
0	0.697674	0.767442
1	0.875000	0.859375
2	0.828125	0.843750
3	0.820312	0.843750
4	0.906250	0.875000
5	0.828125	0.867188

Iteration	Random Forest	SVM
6	0.812500	0.851562
7	0.804688	0.812500
8	0.820312	0.867188
9	0.828125	0.882812
Mean	0.822111	0.847057

From **Table 9** it is known that the accuracy value of the model with 10-level cross-validation. Based on the average accuracy for each model, it is obtained that the Random Forest model has an average accuracy value of 0.822111 or 82.21%. While the SVM model produces an average higher accuracy value of 0.847057 or 84.71%. From the results of calculating the level of accuracy, the SVM model is the best model with an accuracy of 84.71%, the reviews are classified correctly as in the following confusion matrix of SVM shown in **Table 10**.

Table 10. Confusion Matrix of SVM

Iteration	Positive (prediction)	Negative (prediction)
Positive (actual)	136	17
Negative (actual)	35	133

From **Table 10**, 136 positive reviews were classified correctly and 17 positive reviews were not classified correctly. Furthermore, for negative reviews 133 reviews are classified correctly and the remaining 35 are not classified correctly. This shows that the SVM model is already good at classifying Shopee application reviews.

Conclusion

Based on the sentiment analysis of the Shopee application reviews, it was found that the classification using SVM method was superior with an accuracy score of 84.71%, while the accuracy score results from the Random Forest method were obtained at 82.21%. Therefore, the sentiment analysis classification model for Shopee application reviews using SVM method can be utilized by marketplace companies (Shopee) to find out what people think about the Shopee application on the Google Play Store.

In this study, it was found that positive and negative sentiments from Shopee application users can provide insight or new knowledge for Shopee users and companies. According to the most word frequent diagram, the most frequently used words are "*belanja*", "*aplikasi*", and "*barang*". The three words in positive sentiment that often appear are "*belanja*", "*aplikasi*", and "*barang*". In this case, the Shopee company is expected to be able to maintain and even improve the convenience of shopping and existing features so that people continue to believe that Shopee is a marketplace that is worth using. While the three words that have the most frequency of negative sentiments are "*ongkir*", "*kirim*", and "*aplikasi*". This illustrates that people are still complaining about things related to postage costs, delivery service problems, and application systems that still have errors. The results of this analysis can provide input and suggestions to Shopee regarding services, especially shipping costs. With this analysis, it is hoped that the public will gain new knowledge about the good and bad features of the Shopee application service. Therefore, it can be a consideration for the community in downloading and using the Shopee application.

References

[1]   M. I. Purba, D. C. Y. Simanjutak, Y. N. Malau, W. Sholihat, and E. A. Ahmadi, "The effect of digital marketing and e-commerce on financial performance and business sustaina-bility of MSMEs during COVID-19 pandemic in Indonesia," *Int. J. Data Netw. Sci.*, vol. 5, no. 3, pp. 275–282, 2021, doi: 10.5267/j.ijdns.2021.6.006.

[2]   R. Rosita, "Pengaruh Pandemi Covid-19 Terhadap Umkm Di Indonesia," *J. Lentera Bisnis*, vol. 9, no. 2, p. 109, 2020, doi: 10.34127/jrlab.v9i2.380.

[3]   H. P. Monalika and R. Septiyanti, "Intervening Effect of Information Technology on Msmes Performance during Covid-19 Pandemic," vol. 3, no. 1, pp. 1–14, 2022.

[4]   A. P. J. I. Indonesia, "Laporan Survei Internet APJII 2019 – 2020 (Q2)," Jakarta, 2020. [Online]. Available: https://apjii.or.id/survei/surveiinternetapjii20192020q2-21072046.

[5]   L. J. Anreaja, N. N. Harefa, J. G. P. Negara, V. N. H. Pribyantara, and A. B. Prasetyo, "Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea Mobile Application Users in Indonesia," *JISA(Jurnal Inform. dan Sains)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.31326/jisa.v5i1.1267.

[6]   R. Meifitrah, I. Darmawan, and O. Nurul Pratiwi, "Sentiment analysis of tokopedia application review to service product recommender system using neural collaborative filtering for marketplace in Indonesia," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 909, no. 1, 2020, doi: 10.1088/1757-899X/909/1/012071.

[7]   B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.

[8]   V. Y. Dronov and G. A. Dronova, "Python as an automation tool in IS. Protecting Database Access in Python," *J. Phys. Conf. Ser.*, vol. 2182, no. 1, p. 012093, Mar. 2022, doi: 10.1088/1742-6596/2182/1/012093.

[9]   M. Yang, B. Jiang, Y. Wang, T. Hao, and Y. Liu, "News Text Mining-Based Business Sentiment Analysis and Its Significance in Economy," *Front. Psychol.*, vol. 13, no. July, pp. 1–7, 2022, doi: 10.3389/fpsyg.2022.918447.

[10] U. Rhohmawati, I. Slamet, and H. Pratiwi, "Sentiment Analysis Using Maximum Entropy on Application Reviews (Study Case: Shopee on Google Play)," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 1, pp. 44–49, 2019, doi: 10.26555/jiteki.v5i1.13087.

[11] P. A. Aritonang, M. E. Johan, and I. Prasetiawan, "Aspect-Based Sentiment Analysis on Application Review using Convolutional Neural Network," *Ultim. InfoSysJ. Ilmu Sist. Inf.*, vol. 13, no. 1, pp. 54–61, 2022, doi: 10.31937/si.v13i1.2684.

[12] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," in *Seminar Nasional Official Statistics*, 2022, vol. 2022, no. 1, pp. 563–572, [Online]. Available: https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/1410.

[13] A. Nayak and S. Natarajan, "Comparative study of Naïve Bayes , Support Vector Machine and Random Forest Classifiers in Sentiment Analysis of Twitter feeds," *Int. J. Adv. Stud. Comput. Sci. Eng. IJASCSE*, vol. 5, no. 1, pp. 14–17, 2016.

[14] Q. Lv, W. Feng, Y. Quan, G. Dauphin, L. Gao, and M. Xing, "Enhanced-Random-Feature-Subspace-Based Ensemble CNN for the Imbalanced Hyperspectral Image Classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 3988–3999, 2021, doi: 10.1109/JSTARS.2021.3069013.

[15] Z. Wu, W. Lin, Z. Zhang, A. Wen, and L. Lin, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis," *Proc. - 2017 IEEE Int. Conf. Comput. Sci. Eng. IEEE/IFIP Int. Conf. Embed. Ubiquitous Comput. CSE EUC 2017*, vol. 1, pp. 531–536, 2017, doi: 10.1109/CSE-EUC.2017.99.

[16] Kaggle, *shopee review indonesian*. 2021.

[17] S. Khomsah, A. F. Hidayatullah, and A. S. Aribowo, "Comparison of the Effects of Feature Selection and Tree-Based Ensemble Machine Learning for Sentiment Analysis on Indonesian YouTube Comments," 2021, pp. 161–172.

[18] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.

[19] L. B. Ilmawan and E. Winarko, "Aplikasi Mobile untuk Analisis Sentimen pada Google Play," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 9, no. 1, p. 53, 2015, doi: 10.22146/ijccs.6640.

[20] A. Maulana and H. Pratiwi, "Sentiment analysis of public towards infrastructure development in Indonesia on Twitter media using boosting support vector machine method," *AIP Conf. Proc.*, vol. 2202, no. 2019, 2019, doi: 10.1063/1.5141695.

● **4% Overall Similarity**

Top sources found in the following databases:

- Crossref Posted Content database
- 4% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**Universitas Sembilanbelas November Kolaka on 2022-04-18** Submitted works	**3%**
2	**Universitas Sebelas Maret on 2019-08-25** Submitted works	**<1%**
3	**Universitas Islam Indonesia on 2022-11-01** Submitted works	**<1%**
4	**Sheffield Hallam University on 2021-09-13** Submitted works	**<1%**