Comparison of Machine Learning Methods for Classifying User Satisfaction Opinions of the PeduliLindungi Application

Putu Tisna Putra, Anthony Anggrawan, Hairani Hairani
Universitas Bumigora, Mataram, Indonesia

ABSTRACT

Since the emergence of the Covid-19 virus, the Indonesian government urged people to study, work, and worship or work from home. The social restriction policy has changed people's behavior which requires physical distance in social interaction. The government developed an application to minimize the spread of Covid-19, namely the PeduliLindungi application. The PeduliLindungi application is a tracking application to prevent the spread of Covid-19. The government's policy of implementing the PeduliLindungi application during Covid-19 aroused pros and cons from the public. The volume of PeduliLindungi application review data on Google Play was increasing, so manual analysis could not be done. New analytical approaches needed to be carried out, such as sentiment analysis. This research aimed to analyze user reviews of the PeduilLindungi application using classification methods, namely Support Vector Machine (SVM), Random Forest, and Naïve Bayes. The methods used were Synthetic Minority Oversampling Technique (SMOTE), Random Forest, SVM, and Naïve Bayes. SMOTE was used to balance user review data on the PeduliLindungi application. After the data had been balanced, classification was carried out. The results of this study showed that the Random Forest method with SMOTE got better accuracy than the SVM and Naïve Bayes methods, which was 96.3% based on the division of training and testing data using 10-fold cross-validation. Thus, using the SMOTE method could improve the accuracy of classification methods in classifying opinions of user satisfaction with the PeduliLindungi application.

Corresponding Author:

Hairani,+6287839793970,
Faculty of Engineering and Computer Science Undergraduate,
Universitas Bumigora, Mataram, Indonesia,
Email: hairani@universitasbumigora.ac.id

How to Cite:
P. Tisna Putra, A. Anggrawan, and H. Hairani, "Comparison of Machine Learning Methods for Classifying User Satisfaction Opinions of the PeduliLindungi Application", *MATRIK: Jurnal Managemen,Teknik Informatika, dan Rekayasa Komputer*, Vol. 22, No. 3, pp. 431-442, Jul, 2023.
This is an open access article under the CC BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

1. INTRODUCTION

Covid-19 has been a public concern since its appearance was detected in the city of Wuhan, China, in 2020. The emergence of Covid-19 has caused the order of human life to change, especially social life. Since the beginning of the emergence of the covid-19 virus, the Indonesian government has urged to study, work, and worship from home or work from home. People's habits, such as gathering and shaking hands, must be accustomed to social restrictions. The social restriction policy has changed people's behavior which requires physical distance in the process of social interaction. The government itself developed an application that is used to minimize the spread of covid-19, namely the PeduliLindungi application [1]. The PeduliLindungi application is a tracking application to prevent the spread of covid-19 [2].

The features that can be used in the PeduliLindungi application are the location of vaccinations and the community sharing location data for traveling so that the contact history with Covid-19 patients can be known. In addition, the application lets users know when they are in the red zone [3]. The policy of implementing the use of the PeduliLindungi application by the government during covid-19 arouses the pros and cons of the public. Public complaints about the use of the PeduliLindungi application are (1) frequent errors, (2) vaccine certificates do not appear, (3) vaccine status does not change, and (4) concerns about data leakage [4]. In addition, the volume of PeduliLindungi app review data on Google Play is increasing, so manual analysis cannot be done, so a new analytical approach needs to be taken, such as sentiment analysis [5]. Sentiment analysis quickly emerged as an automated process to examine semantic relationships and meanings in reviews. One of the sentiment analysis techniques on social media data that can be used is the machine learning approach [5]. The machine learning approach utilizes algorithms to extract and detect sentiment from an opinion on social media data. Therefore, this research proposes a machine learning approach for sentiment analysis of user satisfaction in the PeduliLindungi application, such as satisfied, neutral, and dissatisfied automatically.

Several previous studies have discussed user reviews of the PeduliLindungi application using various methods, such as research conducted by [6] using the Naïve Bayes method for analyzing user reviews of the Protect application with a total of 1179 datasets and the accuracy obtained is 83.3%. Research [7] compared the Naïve Bayes and SVM methods for sentiment analysis of the PeduliLindungi application using a dataset of 4636 instances. Based on the results of his research, the SVM method has higher accuracy than Naïve Bayes, which is 91% for SVM and 90% for Naïve Bayes. Research [8] used SVM, Naïve Bayes, and KNN methods to classify community responses to the PeduliLindungi application using a dataset of 6000 instances. Based on the test results of this research using 80% training data and 20% testing data, the SVM method got the best accuracy compared to the Nave Bayes and KNN methods of 76.5%. Research [9] utilized the Naïve Bayes method for sentiment analysis of the Peduli Lindungi application with a dataset of 1000 instances. Based on the test results of this research using 80% training and 20 testing data, the Naïve Bayes method obtained an accuracy of 73%. Research [10] used the Long Short Term Memory (LSTM) method for user sentiment classification in the PeduliLindungi application with review data from as many as 3000 instances divided into positive and negative reviews. The results of his research obtained an accuracy of 82.44%, a precision of 78.66%, and a recall of 87%.

Research [11] used the Naïve Bayes method to classify user reviews of the Peduli Lindungi application, with 496 instances of review data divided into positive and negative reviews. The study results obtained an accuracy of 85%, precision of 77.7%, and recall of 98%. Research [12] combined the Naïve Bayes method with Particle Swarm Optimization (PSO) feature selection to classify user reviews on the PeduliLindungi application with review data of 1364 instances divided into positive and negative reviews. The results of the combination of Naïve Bayes and PSO methods get an accuracy of 93% and AUC of 97.7%. Research [13] used the Convolutional Neural Network (CNN) method for sentiment analysis of PeduliLindungi application user reviews based on aspects of Visual Experience, Scan, Vaccine Certificate, eHac, Covid-19 Test, Login, Performance, and Security. The data used were 2320 instances, and the results of his research obtained an f1 score of 95.13%. Research [14] used the SVM method for sentiment analysis of user reviews on the PeduliLindungi application with an accuracy of 64%. Research [15] used the C4.5 method to classify the nutritional status of toddlers with an accuracy of 95%.

Research conducted by [16] used SVM and Naïve Bayes methods to analyze online news media application reviews on Google Play with a dataset of 5000 instances with an accuracy of 88% for SVM and 87% for Naïve Bayes. Research conducted by [17] used the SVM method for sentiment analysis of electronic money service customers with a dataset of 3852 instances obtained from Twitter with an accuracy of 98%. Finally, research conducted by [18] used the random forest method to analyze comments on the relocation of the country's capital with a dataset of 1639, and the accuracy obtained was 76%. The difference between this research and previous research is this research handles the problem of unbalanced data on PeduliLindungi application reviews using SMOTE. After the data is balanced, the PeduliLindungi application user reviews are classified using three methods, namely SVM, Random Forest, and Naïve Bayes. The novelty of this research is the combination of the SMOTE method with several machine learning methods such as SVM, Random Forest, and Naïve Bayes for classifying user opinion satisfaction in the PeduliLindungi application. Therefore, this study aims to analyze user reviews of the PeduliLindungi application using classification methods, namely SVM, Random Forest, and Naïve Bayes.

2. RESEARCH METHOD

Figure 1 shows the flow of this research. Based on Figure 1, this research consists of several stages: collecting datasets by scraping user reviews on the PeduliLindungi application on Google Play. Next is labeling user review data on the PeduliLindungi application, divided into satisfied, neutral, and dissatisfied reviews based on the rating given. Rating ratings of more than three are considered satisfied reviews, ratings of three become neutral, and ratings of less than three become dissatisfied. The next stage is the text preprocessing stage. The text preprocessing section processes raw data into quality data by case folding, tokenization, filtering, and stemming. The results of text preprocessing are then carried out text weighting using the TF-IDF method. The next stage is text classification using Nave Bayes, Random Forest, and SVM methods based on the division of training and testing data using 10-fold cross-validation. The last stage is performance testing based on accuracy using the confusion matrix table.
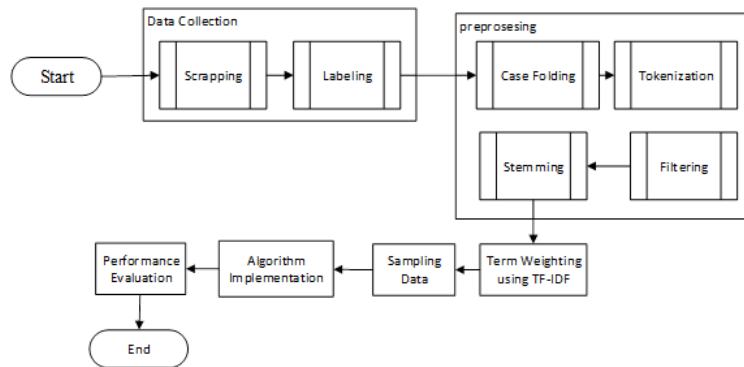

Figure 1. Research Flowchart

2.1. Dataset Collection

he data used are user reviews about the PeduliLindungi application on Google Play obtained by scraping data. However, the results of scraping user review data do not yet contain review categories such as satisfied, neutral, and dissatisfied, so it is necessary to label the review category based on the rating value given by the user. For example, a user score rating ≥ 4 becomes Satisfied, a user score rating ≤ 3 becomes Dissatisfied, and a score rating = 3 becomes Neutral.

2.2. Text Preprocessing

Text preprocessing is used to process raw data into quality data by performing case folding, tokenization, filtering, and stemming [19]. The commonly used stages in the preprocessing stage are shown in Figure 2. In Figure 2, the preprocessing stage begins with case folding. Case folding is the process of changing capital letters into lowercase letters. In contrast, tokenization is the process of separating sentences into words. Then, filtering is used to remove less essential words or keep important words. For example, common words that usually appear in Indonesian are "yang," "dan," "di," and "from," so they need to be removed. Finally, Stemming is the process of forming base words.
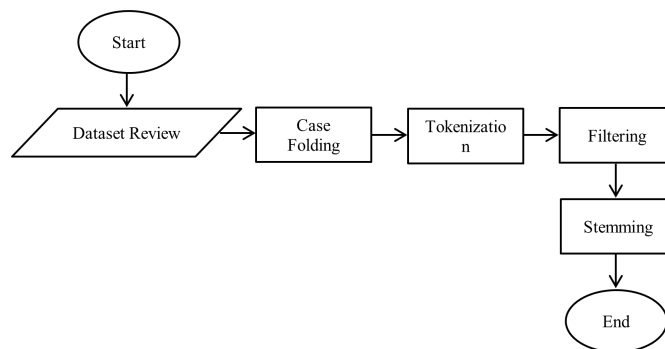

Figure 2. Preprocessing Stages

2.3. Word Weighting

The data from text preprocessing that is still in the form of words will be converted to numerical form with a word weighting process that aims to get the weight of each word used as a feature. The result of word weighting with TF-IDF is the multiplication of TF and IDF values which will produce the weight of each word using Equation (1). Tf is term frequency, W is TF-IDF weight, and idf is Inverse Document Frequency.

$$W = Tf \times idf \tag{1}$$

2.4. Data Sampling

The data used contains class imbalance, which can affect the performance of the classification method. Therefore, unbalanced data must be handled to improve the classification method's performance. This research employed the SMOTE method to balance the data by adding instances to the minority class based on the nearest neighbor. Figure 3 displays the various stages of the SMOTE method.
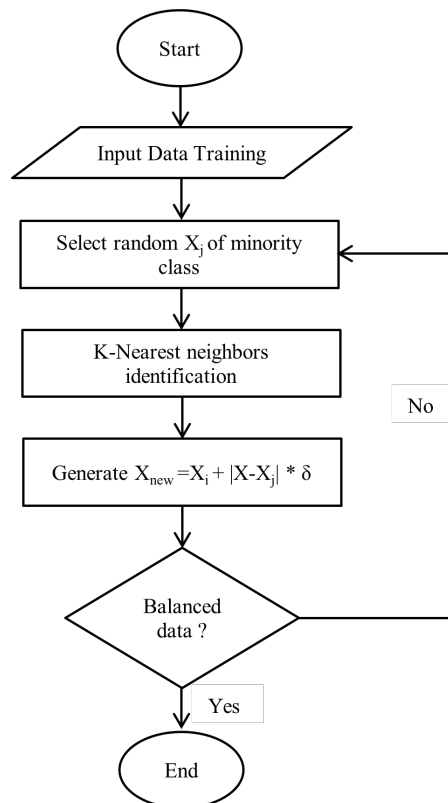

Figure 3. The SMOTE Process

2.5. Implementation Model

At this stage, the algorithms used in classification will be implemented, namely the SVM, Random Forest, and Naïve Bayes methods. The Naïve Bayes classifier is a classification method based on Bayes' theorem. The Naïve Bayes method uses probability and statistical methods to predict opportunities based on previous experience. The Naive Nayes method's classification process uses Equation (2) [20]. P(H|X) represents the probability that the hypothesis is true for the observed sample data X. P(X|H) represents the probability of the sample data X if hypothesis H is true. P(H) represents the probability of hypothesis H, and P(X) represents the probability of the observed sample. The SVM method works by separating two different classes using Equation (3). W is the weight, x is the input data, y is the class, and b is the bias. The Random Forest algorithm is an ensemble learning-based classification method. The advantages of Random Forest are that it has good accuracy results, is robust against outliers and noise, and is faster than bagging and boosting [21]. The stages of the Random Forest method are shown in Figure 4.

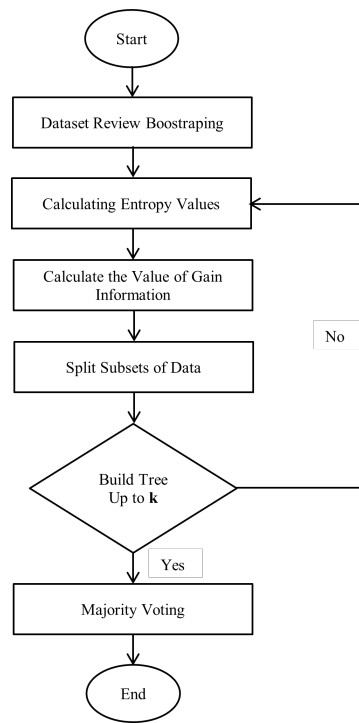$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{2}$$

$$w.x + b = 0 \tag{3}$$

Figure 4. Random Forest Flowchart

2.6. Performance Testing

At this stage, testing is carried out on the classification method used. This research uses accuracy testing to measure performance on the classification method using Equation (4). Accuracy measurement uses a confusion matrix table that consists of True Positive, True Negative, False Positive, and False Negative. True Positive is the actual positive class predicted correctly. True Negative is the actual negative class predicted correctly, False Positive is the actual negative class, but the positive class is predicted. Finally, false Negative is the actual positive class, but the negative class is predicted.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

3. RESULT AND ANALYSIS

This section describes the research results obtained under the stages in Figure 1. PeduliLindungi application user review data is 4005 instances that do not yet have labels, so labeling needs to be done based on user score ratings. For example, a user score rating ≥ 4 becomes Satisfied, a user score rating ≤ 3 becomes Dissatisfied, and a score rating = 3 becomes Neutral. The results of the data labeling that has been done are shown in Table 1. The data that has been obtained is text-processed to make it more qualified with case folding, tokenization, filtering, and stemming. The processed data can be shown in Table 2. The data processed in the previous stage is still in the form of words, so it needs to be converted to numeric form. Changing words into numeric form is required to obtain the weight of each word so that it can be processed by the classification method. The results of word weighting using the TF-IDF method are shown in Table 3.

Table 1. Dataset Labelling Result

No	Content	Score	Opinion
1	aplikasi tolong susah cek hasil sertifikasi vaksin	1	Dissatisfied
2	fitur isi data tanggal lahir repot tulis manual	2	Dissatisfied
3	min ijin scan barcode muncul notifikasi ijin	2	Dissatisfied
4	biar masuk pabrik	4	Satisfied
5	aplikasi gak suruh upload sertifikat vaksin	1	Dissatisfied
…	…	…	…
4001	bagus	5	Satisfied
4002	Update berat forced closed	1	Dissatisfied
4003	Masuk aplikasi minta kirim kode verifikasi	1	Dissatisfied
4004	Manfaat aplikasi bikin ribet masuk mall	1	Satisfied
4005	Bermanfaat butuh butuh	3	Neutral

Table 2. Text Preprocessing Result

No	Content	Score	Opinion
1	[aplikasi, tolong, susah, cek, hasil, sertifikasi, vaksin]	1	Dissatisfied
2	[fitur, isi data, tanggal, lahir, repot, tulis, manual]	2	Dissatisfied
3	[min, ijin, scan, barcode, muncul, notifikasi, ijin]	2	Dissatisfied
4	[biar, masuk, pabrik]	4	Satisfied
5	[aplikasi, gak, suruh, upload, sertifikat, vaksin]	1	Dissatisfied
…	…	…	…
4001	[bagus]	5	Satisfied
4002	[update, berat, forced, closed]	1	Dissatisfied
4003	[masuk, aplikasi, minta, kirim, kode, verifikasi]	1	Dissatisfied
4004	[manfaat, aplikasi, bikin, ribet, masuk, mall]	1	Dissatisfied
4005	[bermanfaat, butuh, butuh]	3	Neutral

Table 3. Term Weighting Results Using TF-IDF Method

Index	Term	TF-IDF Values
(1, 6482)	aplikasi	0,4
(1, 6185)	tolong	0,25
(1, 5822)	susah	0,29
(1, 5405)	cek	0,53
(1, 2701)	hasil	0,31
(1, 2249)	sertifikasi	0,35
(1, 1344)	vaksin	0,26
…	…	…
(4005, 1079)	bermanfaat	0,73
(4005, 738)	butuh	0,68

The data used still contains unbalanced data that affects the performance of the classification method. For example, the number of dissatisfied opinion labels (2999 instances) is more than satisfied opinions (506 instances) and neutral (500 instances), so it needs to balance the opinion label data. Therefore, this research uses the SMOTE method to oversample the data so that the data becomes balanced. The amount of unbalanced and balanced data can be seen in Figure 5. The classification methods used in this research are Naïve Bayes, Random Forest, and SVM. The three methods' performance testing is based on accuracy, precision, and specificity using the confusion matrix table. Based on testing the Naïve Bayes, Random Forest, and SVM methods using 10-fold cross-validation, the results are obtained as a confusion matrix table as in Figure 6 for the Naïve Bayes, Figure 7 for the SVM, and Figure 8 for the Random Forest results. Figure 6 shows that without SMOTE, the Naïve Bayes method can correctly classify the Satisfied label in 0 instances, Neutral in 88 instances, and Dissatisfied in 2992. With SMOTE, the Naïve Bayes method can classify the Satisfied label in 2625 instances, Neutral in 2583 instances, and Unsatisfied in 2136. According to Figure 7, the SVM method without SMOTE can correctly classify the Satisfied label in 17 instances, Neutral in 258, and Dissatisfied in 2950. However, with SMOTE, the SVM method can classify the Satisfied label in 2776 instances, Neutral in 2916 instances, and Dissatisfied in 2942. Figure 8 shows that without SMOTE, the Random Forest method can correctly classify the Satisfied label in 11 instances, Neutral in 238 instances, and

Dissatisfied in 2958. With SMOTE, however, the Random Forest method can classify the Satisfied label in 2866 instances, Neutral in 2913 instances, and Dissatisfied in 2888 instances.
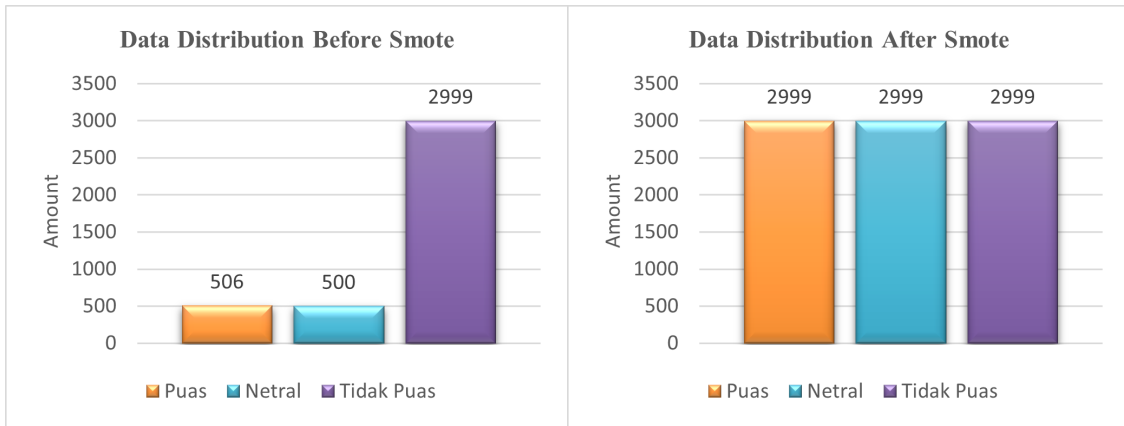

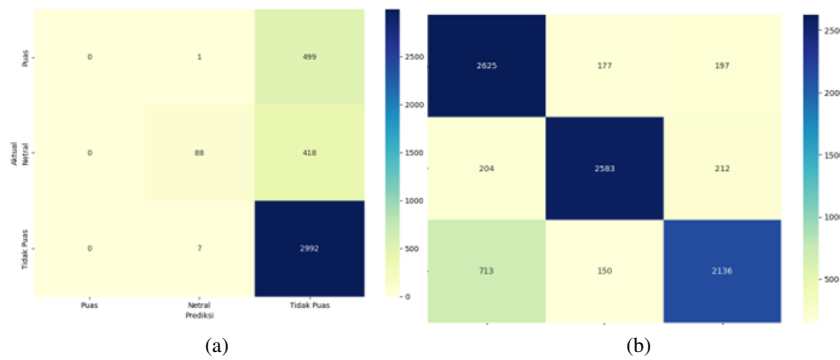Figure 5. Data Distribution of Satisfaction Opinion on the Use of PeduliLindungi Application

Figure 6. (a) Naïve Bayes Method Confusion Matrix Results Before SMOTE, (b) Naïve Bayes Method Confusion Matrix Results After SMOTE
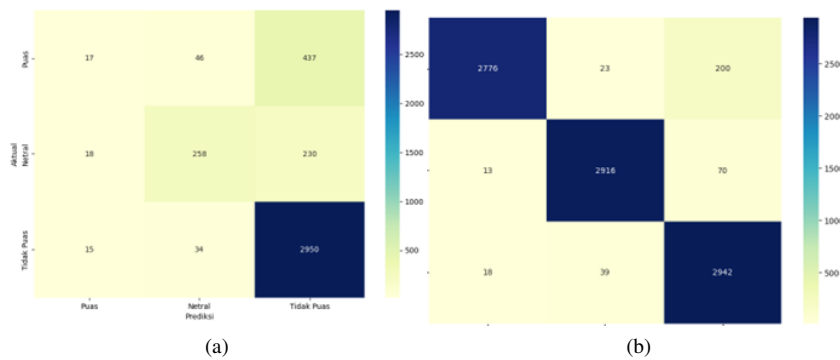

Figure 7. (a) SVM Method Confusion Matric Results Before SMOTE, (b) SVM Method Confusion Matric Results After SMOTE
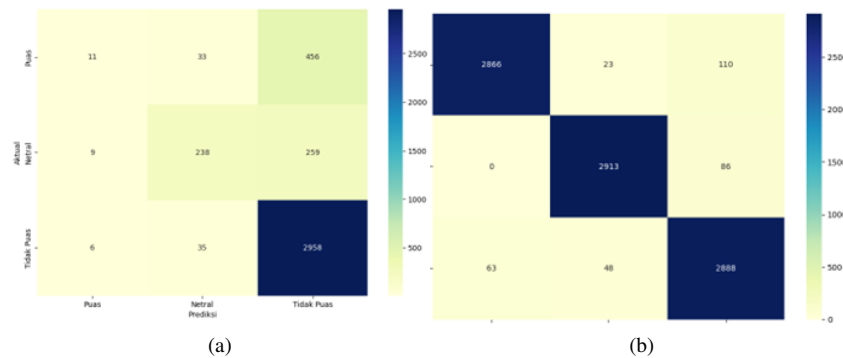
Figure 8. (a) Confusion Matric Result of Random Forest Method Before SMOTE, (b) Confusion Matric Result of Random Forest Method After SMOTE

In this section, we will analyze the results obtained by the Random Forest, Naïve Bayes, and SVM methods based on accuracy, as shown in Figure 9. Based on Figure 8, the Random Forest method with SMOTE gets the best accuracy compared to the Naïve Bayes and SVM methods. The accuracy obtained by the Random Forest method with Smote is 96.3%. Additionally, using SMOTE can increase the accuracy of classification methods supported by research [22–26]. The increase in accuracy in the classification method used after using SMOTE is that the Random Forest method increased by 16.2%, the SVM method increased by 15.4%, and the Naïve Bayes method increased by 4.7%. A comparison of the research results obtained with the results of previous research is shown in Table 4. Table 4 shows that the proposed method gets better accuracy than some previous research results using the same case study.
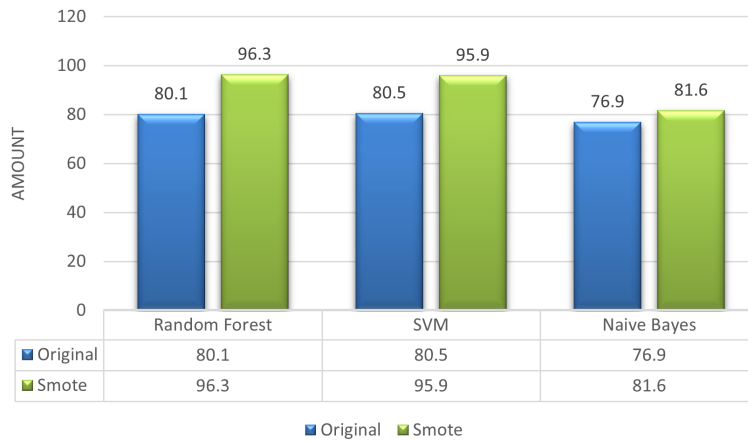

Figure 9. Accuracy of the Classification Methods Used

Table 4. Comparison Results of this Research with Previous Research

No	Researcher(s)	Methods	Case Study	Accuracy
1	Aritonang, et al. [13]	CNN		95.13%
2	Mustapa, et.al [12]	Nave Bayes + PSO		93%
3	Saputra, et.al [11]	Nave Bayes		85%
4	Lustiansyah, et al [10]	LSTM		82%
5	Rais, et.al [9]	Naive Bayes	PeduliLindungi Application Review	73%
6	Salma, et.al [8]	SVM		76.50%
7	Locarso, et.al [6]	Nave Bayes		83.30%
8	Darusman [14]	SVM		64%
9	Proposed Method	Random Forest + Smote		96.30%

4. CONCLUSION

This research dataset consists of user reviews of the PeduliLindungi application on Google Play, with a total dataset of 4005 reasons for the PeduliLindungi application, divided into 506 Satisfied reviews, 500 Neutral reviews, and 2999 Dissatisfied reviews. The data contains unbalanced data, so it is necessary to balance it using the SMOTE method. After SMOTE, the total dataset became 8997 reasons for the PeduliLindungi application, divided into 2999 Satisfied reviews, 2999 Neutral reviews, and 2999 Dissatisfied reviews. Based on the test results, the Random Forest method with Smote gets better accuracy than the Naïve Bayes and SVM methods. Therefore, it can be concluded that the effect of using SMOTE in solving unbalanced data on public opinion reviews about the PeduliLindungi application can improve performance in the classification method used. Future research can use embedding methods other than TF-IDF, such as Word2Vectorizer, BERT, or Bag of Words, and feature selection methods to improve accuracy in classification methods.

REFERENCES

[1] L. Ellyanti, Y. Ruldeviyani, L. E. Pradana, and A. Harjanto, "Sentiment Analysis of Twitter Users to the PeduliLindungi Using Naïve Bayes Algorithm," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 414–421, 2023.

[2] M. I. Nurmansyah, C. Rosidati, Yustiyani, and N. M. Nasir, "Measuring the Success of PeduliLindungi Application Use for Supporting COVID-19 Prevention: A Case Study among College Students in Jakarta, Indonesia," *Kesmas : Jurnal Kesehatan Masyarakat Nasional*, vol. 17, no. 1, pp. 11–16, 2022.

[3] W. Sudiarsa and G. B. Wiraditya, "Heuristic Evaluation Usability Analisys on Information and Tracking Covid-19 Application Peduli Lindungi Using Heuristic Evaluation," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 3, no. 2, pp. 354–364, 2020.

[4] R. D. Novita and M. A. Razak, "Personal Data Protection in Falsification of Covid-19 Vaccination: A Juridical Review," *YURIS (Journal of Court and Justice)*, vol. 1, no. 3, pp. 25–37, 2022.

[5] Z. Drus and H. Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review," *Procedia Computer Science*, vol. 161, pp. 707–714, 2019.

[6] G. K. Locarso, "Analisis Sentimen Review Aplikasi Pedulilindungi Pada Google Play Store Menggunakan NBC," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 2, pp. 353–361, 2022.

[7] I. Yunanto and S. Yulianto, "Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes and Support Vector Machine," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 4, pp. 807–814, 2022.

[8] A. Salma and W. Silfianti, "Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors," *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, pp. 158–162, 2021.

[9] Z. Rais, F. Taufani, T. Hakiki, and R. Aprianti, "Sentiment Analysis of Peduli Lindungi Application Using the Naive Bayes Method," *SAINSMAT: Journal of Applied Sciences, Mathematics, and Its Education*, vol. 11, no. 1, pp. 23–29, jun 2022. [Online]. Available: https://jurnal.ahmar.id/index.php/sainsmat/article/view/794

[10] G. A. Lustiansyah, D. Widiyanto, B. T. Wahyono, U. Pembangunan, N. Veteran, K. Depok, L. Short, T. Memory, A. Sentimen, and G. P. Store, "Analisis klasifikasi sentimen pengguna aplikasi pedulilindungi berdasarkan ulasan dengan menggunakan metode long short term memory," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2022, pp. 327–336.

[11] I. Saputra, T. Djatna, R. R. A. Siregar, D. A. Kristiyanti, H. R. Yani, and A. A. Riyadi, "Text Mining of PeduliLindungi Application Reviews on Google Play Store," *Faktor Exacta*, vol. 15, no. 2, pp. 101–108, 2022.

[12] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization," in *2020 5th International Conference on Informatics and Computing, ICIC 2020*, vol. 2, 2020, pp. 1–7.

[13] P. A. Aritonang, M. E. Johan, and I. Prasetiawan, "Aspect-Based Sentiment Analysis on Application Review using CNN (Case Study : Peduli Lindungi Application)," *Ultima Infosys : Jurnal Ilmu Sistem Informasi*, vol. 13, no. 1, pp. 54–61, 2022.

[14] F. S. Darusman, A. A. Arifiyanti, S. Fitri, and A. Wati, "Sentiment Analysis Pedulilindungi Tweet Using Support Vector Machine Method," *Applied Technology and Computting Science Journal*, vol. 4, no. 2, pp. 113–118, 2022.

[15] H. Hairani, L. Nurhayati, and M. Innuddin, "Web-Based Application for Toddler Nutrition Classification Using C4.5 Algorithm," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 1, no. 2, pp. 77–82, 2022.

[16] U. Kusnia and F. Kurniawan, "Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes," *Jurnal Keilmuan & Aplikasi Teknik Informatika*, vol. 5, no. 36, pp. 24–28, 2022.

[17] F. Romadoni, Y. Umaidah, and B. N. Sari, "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 2, pp. 247–253, 2020.

[18] T. C. Herdiyani and A. U. Zailani, "Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia," *Jurnal Teknologi Sistem Informasi*, vol. 3, no. 2, pp. 154–165, sep 2022.

[19] H. Hairani, A. Anggrawan, A. I. Wathan, K. A. Latif, K. Marzuki, and M. Zulfikri, "The Abstract of Thesis Classifier by Using Naive Bayes Method," in *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, 2021, pp. 312–315.

[20] H. P. Doloksaribu and Y. T. Samuel, "Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 16, no. 1, pp. 1–11, 2022.

[21] N. Muna, F. L. Afriansyah, A. B. Suprayogy, R. Medik, P. N. Jember, M. Informatika, T. Informasi, P. N. Jember, and R. Forest, "Penerapan Algoritma Random Forest untuk Analisis Sentimen Komentar di YouTube tentang Islamfobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, pp. 49–54, 2019.

[22] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, "Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model," *IEEE Access*, vol. 9, pp. 78 621–78 634, 2021.

[23] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling , Oversampling , and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information*, vol. 14, no. 1, pp. 1–15, 2023.

[24] H. Hairani, A. S. Suweleh, and D. Susilowaty, "Penanganan Ketidak Seimbangan Kelas Menggunakan Pendekatan Level Data," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 109–116, 2020.

[25] H. Hairani, A. Anggrawan, and D. Priyanto, "Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link," *International Journal on Informatics Visualization*, vol. 7, no. 1, pp. 258–264, 2023.

[26] A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," *International Journal of Information and Education Technology*, vol. 13, no. 2, pp. 289–295, 2023.