International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 2, No. 2, September 2023, pp. 87∼98
ISSN: 2828-5611
DOI: 10.30812/IJECSA.v2i2.3406

Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods

I Nyoman Switrayana[1], Diki Ashadi[1], Hairani Hairani[1], Afrig Aminuddin[2]
[1]Universitas Bumigora, Mataram, Indonesia
[2]Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia

Article Info

Article history:

Received September 22, 2023
Revised September 26, 2023
Accepted September 28, 2023

Keywords:

Latent Dirichlet Allocation
Sentiment Analysis
SMOTE-Tomek Links
Support Vector Machine

ABSTRACT

Kitabisa is an Indonesian application that functions to raise funds online. Users can easily support various types of campaigns and donate funds to various social causes through the app. User reviews of the application are very diverse, and it is not sure whether user reviews of the application tend to be positive, neutral, or negative. This research aimed to analyze the sentiment of the Kitabisa application by modeling topics using Latent Dirichlet Allocation (LDA) and classifying user reviews using a Support Vector Machine (SVM). The scrapped dataset showed imbalanced dataset problems, so the SMOTE-Tomek Links oversampling technique was proposed. The results of this study show that using LDA produces five topics often discussed in 750 reviews. Then, the performance of SVM without using SMOTE-Tomek Links was 72% accuracy, 76% precision, 72% recall, and 64% f1 score. Meanwhile, using SMOTE-Tomek Links could significantly improve the performance, namely 98% accuracy, 98% precision, 98% recall, and 98% f1 score. Based on this research, the application of SVM achieved high performance for user sentiment classification, especially when the dataset was in a balanced state. Therefore, the SMOTE-Tomek Links oversampling technique is recommended for dealing with unbalanced sentiment datasets.

Corresponding Author:

Hairani Hairani,
Universitas Bumigora, Mataram, Indonesia.
Email: Hairani@universitasbumigora.ac.id

How to Cite: I. N. Switrayana, D. Ashadi, H. Hairani, and A. Aminuddin, Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods, *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 2, no. 2, pp. 87 - 98, Sep. 2023. doi: 10.30812/ijecsa.v2i2.3406.

1. INTRODUCTION

In the current digital era, using applications to make donations and fundraise is increasingly popular. One of the well-known fundraising platforms in Indonesia is Kitabisa. Kitabisa is a company engaged in social enterprise that utilizes the latest technology as a financial manager for digital-based fundraising (crowdfunding). Kitabisa is an application that allows individuals, organizations, and charitable institutions to raise funds online [1].

Kitabisa application is a digital platform that was established in 2013. This application has reached over 1 million downloaders and received a high rating of 4.6/5.0 on Google Playstore. This achievement shows that many users are satisfied with its services. In carrying out its operations, Kitabisa has committed to operating according to applicable rules. The application has licenses and legality from several government agencies, including the Ministry of Law and Human Rights (Kemenkumham), the Ministry of Communication and Information (KOMINFO), the National Amil Zakat Agency (BAZNAS), and the Ministry of Social Affairs (KEMENSOS). The Kitabisa platform has facilitated 2.5 million monthly donation transactions and supported more than 28,000 social fundraisers every month by 2023. The increasing number of KitaBisa application users can affect user satisfaction with the application's performance. Therefore, it is necessary to conduct a sentiment analysis of Kitabisa application reviews due to the high participation of users in donating. Sentiment analysis involves user reviews on the Google Play Store, including app features, performance, and overall experience.

Previous research has widely used the SVM method to perform sentiment analysis. Some of them are Research [2] comparing the use of the SVM method with Nave Bayes on sentiment analysis about the COVID-19 vaccine. The results showed that the SVM algorithm performs better in accuracy, precision, and recall, with values of 90.47%, 90.23%, and 90.78%. The performance of the Nave Bayes algorithm is 88.64%, 87.32%, and 88.13%, with a difference of 1.83% accuracy, 2.91% precision, and 2.65% recall. Performance improvement is enhanced by optimizing both SVM and Nave Bayes algorithms using particle swarm optimization. SVM's excellent performance was adopted in this study to analyze sentiment. Research [3] compared SVM, Random Forest, and Naive Bayes methods on user sentiment about the PeduliLindungi application. The results show that the Random Forest method has better accuracy than SVM and Naive Bayes, which is 96.3%. The Random Forest method gets very high accuracy due to the use of SMOTE in balancing the user review data of the PeduliLindungi application in the Play Store.

Research [4] uses the Naive Bayes method to classify thesis topic types based on abstracts. The results show that the Nave Bayes method produced 88.69% accuracy, 89.76% precision, and 90.4% sensitivity. Research [5] uses the LDA method to model student perceptions of online learning. The research modeling results illustrate that students' perceptions/views of online learning still dominate about internet network and quota problems. Research [6] conducted an application review using SVM and Naive Bayes methods. Based on the results of their research, the SVM method is more accurate than the Naive Bayes method, which is 92.86% accurate.

Based on several previous studies, there are differences with this research, namely modeling the topics that appear in Kitabisa reviews and used the SVM method combined with SMOTE-Tomek Links to overcome unbalanced data which has never been carried out in previous research. By balancing the data, SVM is expected to achieve higher performance. The data used comes from Kitabisa user reviews on the Google Play Store. This research aims to analyze the sentiment of the Kitabisa application by modeling topics using Latent Dirichlet Allocation (LDA) and classifying user reviews using a Support Vector Machine (SVM). The scrapped dataset shows imbalanced dataset problems, so the SMOTE-Tomek Links oversampling technique is proposed. This research is also expected to provide valuable insights for developing and improving the quality of Kitabisa based on user reviews on the Google Play Store.

2. RESEARCH METHOD

This research has six stages of research methodology using the CRISP-DM methodology. The CRISP-DM methodology is shown in Figure 1.

2.1. Business Understanding

At this stage, identification of the business objectives and problems to be solved by scraping data in the form of user reviews of the Kitabisa application is carried out; then, we explore an in-depth understanding of the context and needs of the research. The main objectives of the research are to gain a thorough understanding of the application and the importance of user reviews.

2.2. Data Understanding

In this stage, data collection relevant to the research is carried out by taking user reviews of the Kitabisa application through the Google Play Store. Then, analyze and comprehend the characteristics of the data in the form of user reviews in Indonesian

2.3. Data Preparation

During this stage, data cleaning or preprocessing is carried out on the data that has been collected. Preprocessing is a stage in processing data that can be used to process an analysis [7]. The preprocessing used in this study goes through 5 stages. The preprocessing stages are shown in Figure 2.

Figure 1. CRISP-DM Methodology

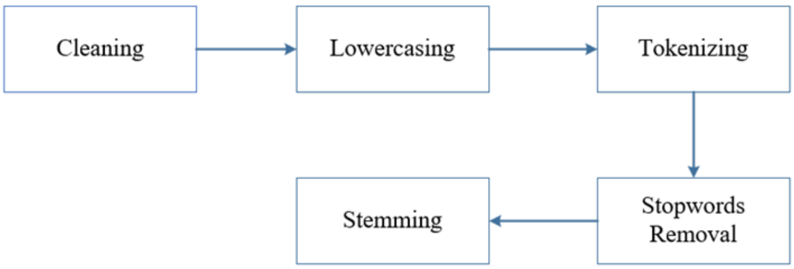

Figure 2. Preprocessing Stages

1. **Cleaning:** Aims to clean the data from unwanted or irrelevant characters.
2. **Lowercasing:** Aims to convert all letters in the text into lowercase letters for consistency in data processing. This process avoids differences or variations in writing caused by differences in capitalization in the text so that it can reduce the number of features that will be used for training.
3. **Tokenizing:** Separating the text into separate units called tokens, in which case the token or smallest unit is the word.
4. **Stopwords Removal:** Removing common words in Indonesian that do not contribute significantly to the text analysis. Words removed in the stopwords stage usually include connecting words such as ”dan,” ”atau,” and ”di,” then common words such as ”saya,” ”kamu,” and ”mereka,” as well as other words that frequently appear in a particular language and tend not to carry any special information about the text itself.
5. **Stemming:** Converting words in the text into basic forms or root words. Words in the text will be trimmed or cut to leave only the basic form or root word. For example, words such as ”berdonasi,” ”memudahkan,” and ”membantu” will be converted into ”donasi,” ”mudah,” and ”bantu”..

2.4. Modeling

This stage begins by weighting the data using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF is a method for determining the relative frequency of each word or token where each word will be given a weighting in the form of a value based on whether or not a word is important in the document. This is based on the number of occurrences of words in a document and measures these words against all existing documents [8]. The TF-IDF method monitors the occurrence of tokens contained in the set of text [9] to become a vector. This stage is also called feature extraction or vectorization. Text data is transformed into a vector to be trained on a learning model. In order to obtain TF-IDF, it is necessary to calculate Term Frequency (TF) and Inverse Document Frequency (IDF). TF output is the proportion of the frequency of occurrence of words in the document to the number of all words in the document. TF is obtained by equation (1). IDF is the logarithm result of dividing the number of documents by the number of documents that contain the term/word. IDF can be obtained by Equation (2). TF-IDF is calculated by multiplying TF and IDF as in Equation (3). Where n = the number of times the term (word) appears in a document, m = the total number of terms (words). Where N = number of documents, df = number of documents with term (word).

$$TF = \frac{n}{m} \tag{1}$$

$$IDF = log\frac{N}{df} \tag{2}$$

$$TF - IDF = TF * IDF \tag{3}$$

After performing feature extraction with TFIDF, the Latent Dirichlet Allocation (LDA) method is implemented to perform topic modeling on user reviews of the Kitabisa application. The advantage of the LDA method is that it can extract topics accurately on a large enough data set [10]. LDA works by receiving input from individual documents and several parameters. Then, LDA produces an output as a model consisting of weights that can be normalized according to probability. There are two types of relevant probabilities: the probability that a particular document yields a particular topic and the probability that a particular topic yields specific words from a vocabulary set. In the topic modeling stage using Latent Dirichlet Allocation (LDA), there are various flows in the modeling stage, and Figure 3 shows the flow of topic modeling. Topic Modeling Validation is used to evaluate and measure the quality of the topic modeling model. This is done by calculating the coherence score for each topic the model generates. The coherence score describes the degree of congruence and relationship of keywords in a topic. The coherence score can indicate the extent to which the topics generated by the model have consistent clarity of document content and are interrelated.

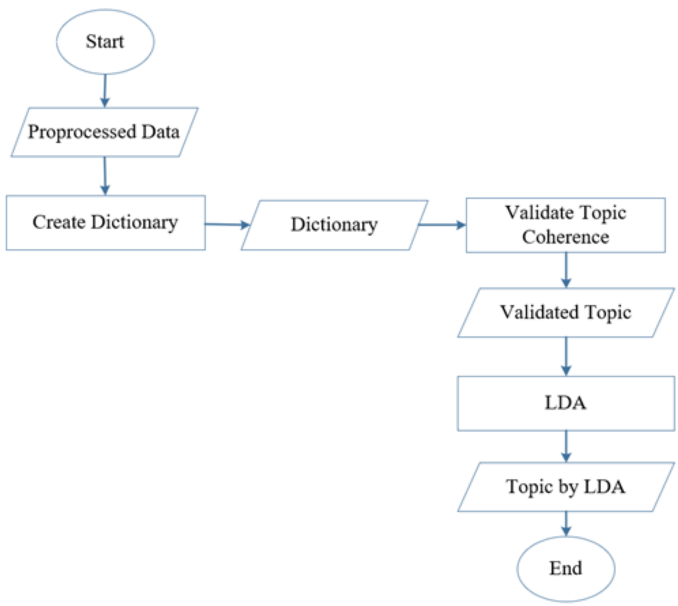

Figure 3. Flow of Topic Modeling by LDA

The next sentiment classification process uses a Support Vector Machine (SVM). SVM is one of the methods that can be used in the classification or regression process. This SVM method aims to find the best hyperplane that separates two classes in the input space [11]. The principle of the SVM method is the linear classification that can be separated. The concept of kernels in the high-dimensional workspace has been incorporated into the SVM method to solve non-linear problems. Two scenarios exist in training and testing SVM models with imbalanced and balanced datasets. Data balancing here uses SMOTE-Tomek Links. SMOTE-Tomek Links is a combination of SMOTE and Tomek Links, which is included in the over-under sampling category where the oversampling technique uses SMOTE, and the undersampling technique is Tomek Links [12]. Figure 4 shows the flow of sentiment classification modeling using SVM with/without SMOTE-Tomek Links.

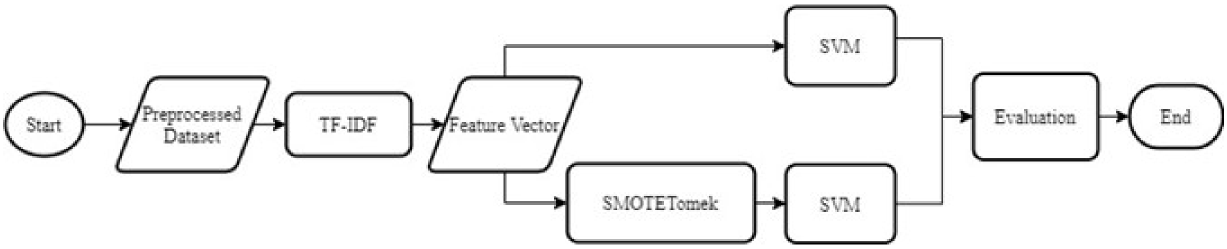

Figure 4. Learning model architecture using SVM with/without SMOTE-Tomek Links

2.5. Evaluation

Evaluation is conducted to see the performance of the proposed model. Evaluation of the results of training and testing SVM models using evaluation metrics: accuracy, precision, recall, and f1-score. SVM performance will be compared with input data in an unbalanced data state with input data in a balanced state. The dataset division is 80% training data and 20% test data.

2.6. Deployment

Implement the research results in a practical context by saving the learning model with the best performance and developing it in the form of a web-based application so that it will be easy for users in sentiment classification with new input data.

3. RESULT AND ANALYSIS

3.1. Dataset

The steps involved data preparation using the Jupyter Notebook application and the Python programming language. The data in this study was obtained by scraping data obtained through user reviews from the Google Play Store. The amount of data taken amounted to 750 data. After successfully scraping the data, the next step is to sort the data, which only displays the ”ulasan” and ”rating” used in the study. ”Ulasan” is text input, and ”rating” results from giving user satisfaction on a scale of 1-5. The dataset that has been saved is then recalled for class/labeling based on the rating. The rating is then categorized into three parts: positive, neutral, and negative. Rating 1-2 is categorized as ”negatif”, Rating 3 is categorized as ”netral”, and rating 4-5 is categorized as ”positif”. So, from 750 data, 497 positive reviews, 75 neutral reviews, and 178 negative reviews will be obtained. Figure 5 shows the sample data labeled based on the rating.

Preprocessing is an important stage in analyzing data before starting the analysis process. Preprocessing is done to avoid imperfect data, disturbances in the data, and inconsistent data [13]. The purpose of preprocessing is to remove noise in the data to improve data quality. Data cleaning is used to clean data from invalid or incomplete values. Figure 6 shows the results of data cleaning on reviews. Lowercasing is used to convert all letters into lowercase letters to eliminate differences in writing. This also affects the number of features that will be reduced because words with the same meaning but different writing will become one entity. Figure 7 shows the result of lowercasing from the previous text.

Tokenizing is involved in splitting text into the smallest units (words). Text data (sentences or paragraphs) are cut based on separators such as spaces or punctuation marks. This separation into the smallest units will be useful for performing TF-IDF. Words are the features that will be used in training and testing the model. Figure 8 shows the tokenizer result from the input data of the previous process. Stopwords removal is utilized to remove words that have no special meaning. This process cleans the text, reduces the number of features, and improves the dataset's quality. Figure 9 shows the result of the removal of stopwords. Stemming serves to convert words in the text into basic forms or root words. The words in the text will be trimmed or cut so that they only leave the basic form or root word and then are converted into a string. Root words from stemming may not have meaning or are not found in the Indonesian dictionary because stemming forms the root word by "cutting" prefixes and affixes. Figure 10 shows the result of stemming.

	Ulasan	Rating	Hasil
0	Appnya bagus, mudah dipahami dan sangat memban...	4	Positif
1	Aplikasi yang sungguh luar biasa. Terimakasih ...	5	Positif
2	Terakhir isi kantong donasi via transfer bank,...	1	Negatif
3	Sedekah subuh otomatis nya tdk sesuai waktunya...	5	Positif
4	aneh, smalem udh daftar dan mengajukan penggal...	2	Negatif
5	Saya mengajukan penggalangan dana utk pembangu...	4	Positif
6	Semakin mudah dan ngga repot2 lagi untuk berba...	5	Positif
7	Aplikasi ini sangat bermanfaat baik bagi pembe...	5	Positif
8	Angel wes angeel... habis apdet, pngen baca de...	2	Negatif
9	Tidak bisa membaca deskripsi cerita yg bersang...	2	Negatif

Figure 5. Sample dataset of scraping results

	Ulasan	Rating	Hasil
0	Appnya bagus mudah dipahami dan sangat membant...	4	Positif
1	Aplikasi yang sungguh luar biasa Terimakasih u...	5	Positif
2	Terakhir isi kantong donasi via transfer bank ...	1	Negatif
3	Sedekah subuh otomatis nya tdk sesuai waktunya...	5	Positif
4	aneh smalem udh daftar dan mengajukan penggala...	2	Negatif
5	Saya mengajukan penggalangan dana utk pembangu...	4	Positif
6	Semakin mudah dan ngga repot lagi untuk berbag...	5	Positif
7	Aplikasi ini sangat bermanfaat baik bagi pembe...	5	Positif
8	Angel wes angeel habis apdet pngen baca deskri...	2	Negatif
9	Tidak bisa membaca deskripsi cerita yg bersang...	2	Negatif

Figure 6. Cleaned data

	Ulasan	Rating	Hasil
0	appnya bagus mudah dipahami dan sangat membant...	4	Positif
1	aplikasi yang sungguh luar biasa terimakasih u...	5	Positif
2	terakhir isi kantong donasi via transfer bank ...	1	Negatif
3	sedekah subuh otomatis nya tdk sesuai waktunya...	5	Positif
4	aneh smalem udh daftar dan mengajukan penggala...	2	Negatif
5	saya mengajukan penggalangan dana utk pembangu...	4	Positif
6	semakin mudah dan ngga repot lagi untuk berbag...	5	Positif
7	aplikasi ini sangat bermanfaat baik bagi pembe...	5	Positif
8	angel wes angeel habis apdet pngen baca deskri...	2	Negatif
9	tidak bisa membaca deskripsi cerita yg bersang...	2	Negatif

Figure 7. Lowercasing results

```
Tokenizing result :

0     [appnya, bagus, mudah, dipahami, dan, sangat, ...
1     [aplikasi, yang, sungguh, luar, biasa, terimak...
2     [terakhir, isi, kantong, donasi, via, transfer...
3     [sedekah, subuh, otomatis, nya, tdk, sesuai, w...
4     [aneh, smalem, udh, daftar, dan, mengajukan, p...
5     [saya, mengajukan, penggalangan, dana, utk, pe...
6     [semakin, mudah, dan, ngga, repot, lagi, untuk...
7     [aplikasi, ini, sangat, bermanfaat, baik, bagi...
8     [angel, wes, angeel, habis, apdet, pngen, baca...
9     [tidak, bisa, membaca, deskripsi, cerita, yg, ...
```

Figure 8. Tokenizing results

```
Stopwords result :

0     [appnya, bagus, mudah, dipahami, membantu, dip...
1     [aplikasi, sungguh, terimakasih, tim, kitabisa...
2     [isi, kantong, donasi, via, transfer, bank, st...
3     [sedekah, subuh, otomatis, nya, tdk, sesuai, m...
4     [aneh, smalem, udh, daftar, mengajukan, pengga...
5     [mengajukan, penggalangan, dana, utk, pembangu...
6     [mudah, ngga, repot, berbagi, sesamaapa, donas...
7     [aplikasi, bermanfaat, pemberi, donasi, donasi...
8     [angel, wes, angeel, habis, apdet, pngen, baca...
9     [membaca, deskripsi, cerita, yg, bersangkutan,...
```

Figure 9. Reviews after stopwords removal

```
Stemming :

0     appnya bagus mudah paham bantu baik fitur dona...
1     aplikasi sungguh terimakasih kitabisa moga apl...
2     kantong donasi transfer bank status mutasi has...
3     sedekah subuh otomatis sesuai saldo potong sor...
4     aneh smalem daftar galang dana kucing sakitfan...
5               galang dana bangun jelas approve status
6     mudah ngga repot bagi sesamaapa donasi otomati...
7     aplikasi manfaat beri donasi donasi moga aplik...
8     angel angeel habis apdet pngen baca deskripsi ...
9     baca deskripsi cerita sangkut galang dana buka...
```

Figure 10. Stemming results

3.2. Experiments

1. Topic Modeling

Topic modeling by Latent Dirichlet Allocation (LDA) of user reviews can provide insight into the topics that appear in the processed text data. Before determining the output, it is necessary to validate topic modeling (Topic Coherence). The topic modeling results will be taken from the five best topics with the highest coherence score value. The results of the five topics contained in the coherence score are taken from each keyword given, and if identified, the topics discussed are listed in Table 1. Table 1 shows the top 5 topics discussed from user reviews.

Table 1. Main topics in the review

No	Topic
1	Donate payments through the application using balance transfer, good for helping people.
2	Kitabisa application can make it easy to give alms and has email login access.
3	Kitabisa application makes it easy to donate and provides benefits to people.
4	Donations via the Kitabisa application can help people in need.
5	Kitabisa donation application is good to use to raise funds and provides many benefits.

2. Sentiment Analysis

The preprocessed data will then be used for TF-IDF weighting. TF calculates the occurrence of a word, and IDF will rate a frequently occurring word as less important based on its occurrence in all documents. The smaller the IDF value, the less important the word is, and vice versa. Figure 11 shows the TF-IDF result of each word in the dataset. There are 2185 words as features to be used. The vectors from TFIDF weighting are then used as input data into the SVM model in the scenario of unbalanced data. In addition, SMOTE-Tomek Links will be applied according to the research scenario on balanced data. Applying the SMOTE-Tomek Links method can improve the class imbalance problem in the dataset. This method has produced a dataset with a balanced amount of data between the majority class (positive) and the minority class (negative and neutral). New synthetic samples have been created among the existing minority data, thus increasing the percentage of minority classes. In addition, the Tomek Links method has also been used to remove ambiguous sample pairs between the minority and majority classes. Figure 12 shows the result of dataset balancing.

```
                Term   TF  DF       IDF  TF-IDF
0              aamii  0.0   1  1.000000     0.0
1             aamiin  0.0  21  0.047619     0.0
2          aamiinkan  0.0   1  1.000000     0.0
3        aamiinkarena  0.0   1  1.000000     0.0
4               abai  0.0   1  1.000000     0.0
...              ...  ...  ..       ...     ...
2180         youtube  0.0   1  1.000000     0.0
2181        youtuber  0.0   1  1.000000     0.0
2182           zakat  0.0  26  0.038462     0.0
2183   zakatterimakah  0.0   1  1.000000     0.0
2184           zemua  0.0   1  1.000000     0.0

[2185 rows x 5 columns]
```

Figure 11. TF-IDF result of dictionary dataset

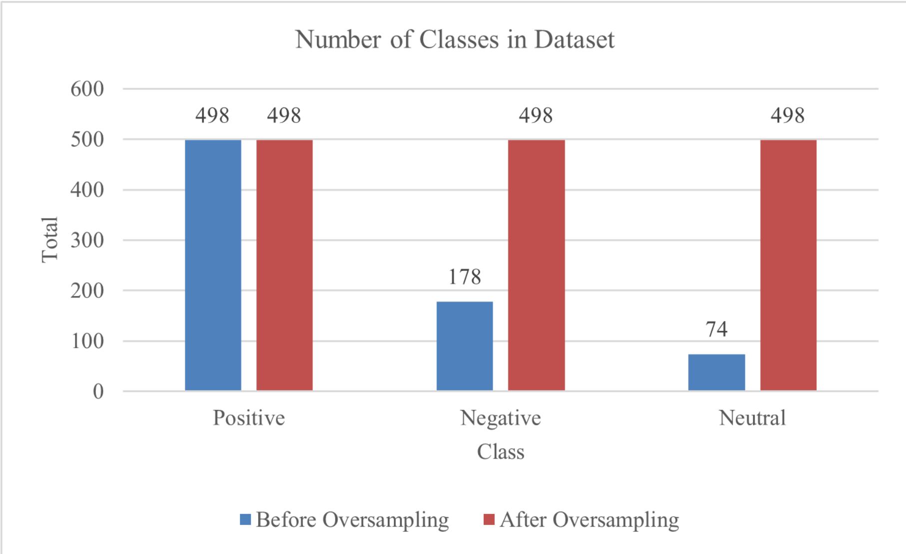

Figure 12. SMOTE-Tomek Links Data Results

The data that has applied SMOTE-Tomek links are then used to train and test the SVM. This test aims to evaluate the performance of the SVM method in classifying data that has been changed through the application of SMOTE-Tomek links. In this test, a performance comparison is made between SVM models that use SMOTE-Tomek Links and those that do not use SMOTE-Tomek Links. This aims to see the difference in performance between the two approaches. Figure 13 shows the performance comparison. In this case, the SVM model without using SMOTE-Tomek Links produces quite good results. However, it should be noted that the relatively low accuracy indicates that the model still has problems in correctly identifying the sentiment. Then, using SMOTE-Tomek Links significantly improves the performance of the SVM model. The accuracy, precision, recall, and F1 Score all reached 98%, indicating that the model can distinguish sentiments very well. These high results show that the SMOTE-Tomek Links technique effectively addresses the minority class imbalance problem that often occurs in sentiment analysis. A balanced dataset makes it easier for the SVM model to find the best hyperplane. This is also caused by SVM obtaining important additional information in the vector space when each class has an equal portion of data.

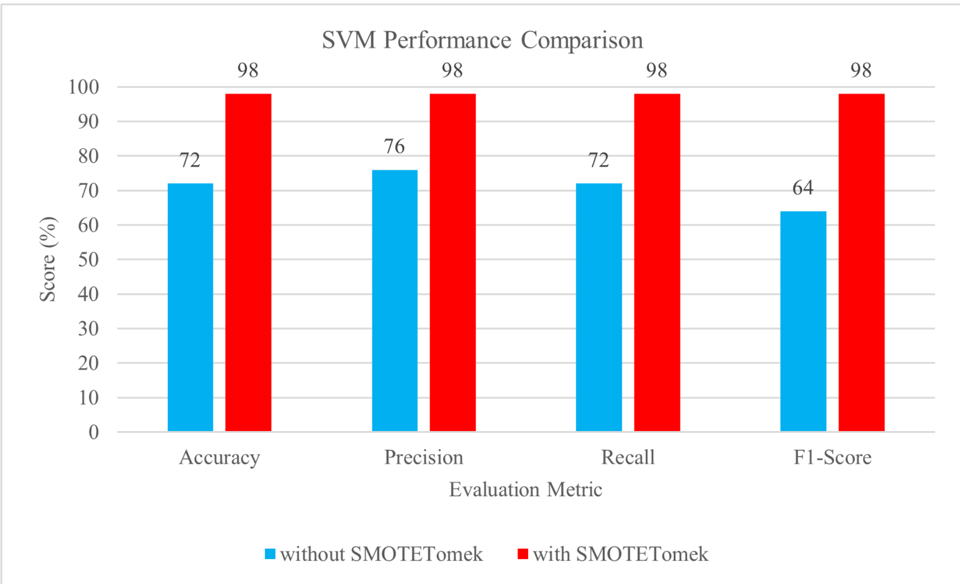

Figure 13. Comparison of SVM performance without/with Smote-Tomek Link

The confusion matrix can show the difference in classification results in more detail. Figure 14 and 15 shows a comparison of the confusion matrix visualization of the SVM model that uses SMOTE-Tomek Links with the one that does not use SMOTE-Tomek Links. Label 0 is negative sentiment, 1 is neutral sentiment, and 2 is positive sentiment. The SVM model without SMOTE-Tomek Links still makes many wrong predictions in the negative and neutral classes. These classes are minority classes that are then classified into the majority class. This error is most likely experienced because the model cannot distinguish between negative and neutral classes due to a lack of data during training. Compared to the SVM model with SMOTE-Tomek Links, the model can predict the sentiment of each class overall. Adding negative and neutral sentiment data causes the model to be better able to distinguish sentiment in the dataset, especially for minority classes. This result is in line with research [14] [15] [16], which states that the use of the Smote-Tomek Link method can improve the performance of the classification method used. Table 2 shows that the proposed method has better performance than previous research in the state of the art section

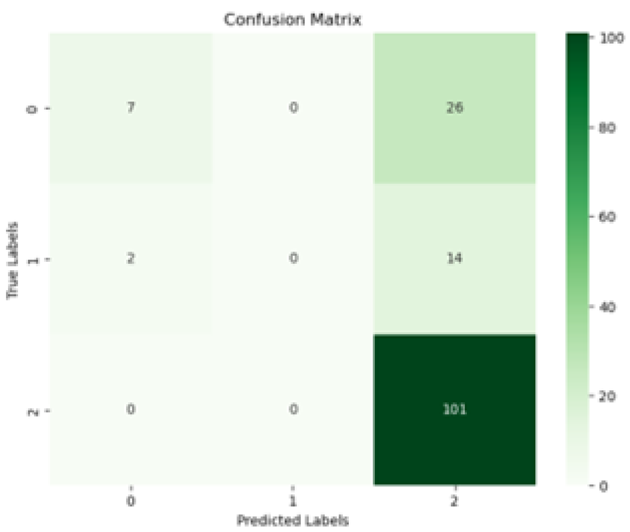

Figure 14. Without SMOTE-Tomek Links

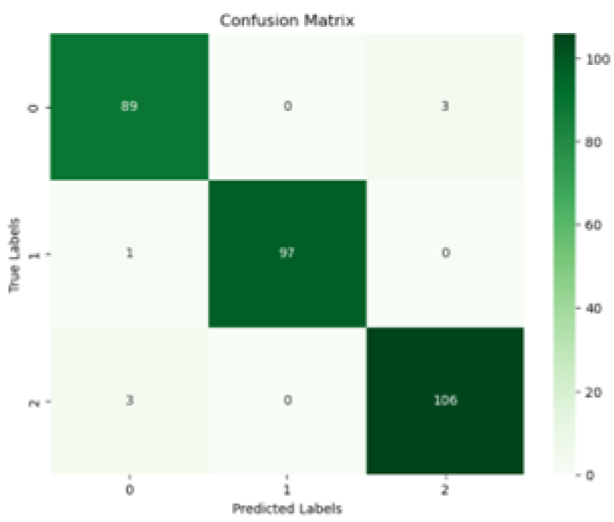

Figure 15. With SMOTE-Tomek Links

Table 2. Comparison Results of this Research with Previous Research

No	Researcher(s)	Methods	Case Study	Accuracy
1	Putra, et.al [3]	Random Forest with Smote	PeduliLindungi Application Reviews	96.3%
2	Amy, et.al [17]	Text Preprocessing	Indonesian Government Policy in the Environmental Sector	74.4%
3	Dany, et.al [18]	Nave Bayes Shopee App Reviews	96,6%	
4	Ala, F [19]	SVM with Kernel Linear	MyPertamina Application Reviews	96%
5	**Proposed Method**	**SVM with Smote-Tomek Link**	**Kitabisa Application Reviews**	**98%**

4. CONCLUSION

The Kitabisa app is an effective and easy-to-use platform for donating through balance transfers. The app provides convenient access and allows people to assist those in need easily. Kitabisa can also be used to raise funds effectively and benefit the community significantly. Thus, Kitabisa is a useful and efficient donation platform for helping needy people. So that the sentiment in the reviews shows more positive sentiment, from the modeling results, SVM can distinguish sentiment from reviews by achieving high performance even in a highly imbalanced dataset. In this case, the SMOTETomek Links technique used to oversample the minority data significantly improves SVM performance. However, future research should also investigate other oversampling techniques, such as KMeansSMOTE, ADASYN, etc. It can also be combined with deep learning models such as Recurrent Neural Network (RNN) to achieve higher performance.

5. ACKNOWLEDGEMENTS

We would like to thank those who have supported this research.

6. DECLARATIONS

AUTHOR CONTIBUTION

All authors contributed to the writing of this article.

FUNDING STATEMENT

-

COMPETING INTEREST

The authors declare no conflict of interest in this article.

REFERENCES

[1] I. Irwansyah and N. Hutami, "Pemanfaatan Aplikasi Mobile Kitabisa Dalam Pelaksanaan Crowdfunding Di Indonesia." *Jurnal Komunikasi*, vol. 13, no. 2, pp. 183–194, 2019.

[2] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 19–25, jul 2021. [Online]. Available: https://journal.unimma.ac.id/index.php/komtika/article/view/5185

[3] P. T. Putra, A. Anggrawan, and H. Hairani, "Comparison of Machine Learning Methods for Classifying User Satisfaction Opinions of the PeduliLindungi Application," *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 22, no. 3, pp. 431–440, 2023.

[4] 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 and Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, 2021, pp. 312–315.

[5] J. W. Fernanda, "Pemodelan Persepsi Pembelajaran Online Menggunakan Latent Dirichlet Allocation," *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 9, no. 2, pp. 79–85, 2021.

[6] M. F. Madjid, D. E. Ratnawati, and B. Rahayudi, "Sentiment Analysis on App Reviews Using Support Vector Machine and Naïve Bayes Classification," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 8, no. 1, pp. 556–562, 2023. [Online]. Available: https://jurnal.polgan.ac.id/index.php/sinkron/article/view/12161

[7] G. K. Locarso, "No Title," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 62, pp. 353–361, 2022.

[8] S. Rahayu, Y. MZ, J. E. Bororing, and R. Hadiyat, "Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 98–106, jun 2022. [Online]. Available: http://e-journal.hamzanwadi.ac.id/index.php/edumatic/article/view/5433

[9] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *Journal of Information Technology*, vol. 2, no. 1, pp. 36–40, mar 2022. [Online]. Available: https://journal.shantibhuana.ac.id/index.php/jifotech/article/view/330

[10] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123–131, feb 2021. [Online]. Available: http://jurnal.iaii.or.id/index.php/RESTI/article/view/2587

[11] M. I. Putri and I. Kharisudin, "Analisis Sentimen Pengguna Aplikasi Marketplace Tokopedia Pada Situs Google Play Menggunakan Metode Support Vector Machine (SVM), Naïve Bayes, dan Logistic Regression," in *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, 2022, pp. 759–766.

[12] Y. A. Sir and A. H. H. Soepranoto, "Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas," *Jurnal Komputer dan Informatika*, vol. 10, no. 1, pp. 31–38, mar 2022. [Online]. Available: https://ejurnal.undana.ac.id/index.php/jicon/article/view/6554

[13] R. Khalida and S. Setiawati, "Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain," *Journal of Informatic and Information Security*, vol. 1, no. 1, pp. 19–26, may 2020. [Online]. Available: http://ejurnal.ubharajaya.ac.id/index.php/jiforty/article/view/137

[14] 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.

[15] H. Hairani and D. Priyanto, "A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data," *International Journal of Advanced Computer Science and Application*, vol. 14, no. 8, pp. 585–590, 2023.

[16] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset," *Sensors*, vol. 22, no. 9, pp. 1–21, apr 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/9/3246

[17] A. Fontanella, H. Ihsan, and Z.-h. Wang, "Review of Accounting , Finance and Governance Sentiment Analysis of Indonesian Government Policy in the Environmental Sector Using Machine Learning Method," *Jurnal Rafgo*, vol. 1, no. 2, pp. 1–5, 2021.

[18] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," in *Journal of Physics: Conference Series*, vol. 1641, no. 1, 2020, pp. 1–7.

[19] F. Y. A'la, "Indonesian Sentiment Analysis towards MyPertamina Application Reviews by Utilizing Machine Learning Algorithms," *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 5, no. 1, pp. 80–91, 2022.