Fine-Grained Sentiment Analysis on PeduliLindungi Application Users with Multinomial Naive Bayes-SMOTE

1st Imam Suyuti
Department of Mathematics
Faculty of Mathematics and Natural Sciences
Universitas Sebelas Maret
Surakarta, Central Java, Indonesia
imamsuyuti00@student.uns.ac.id

2nd Dewi Retno Sari S.
Department of Mathematics
Faculty of Mathematics and Natural Sciences
Universitas Sebelas Maret
Surakarta, Central Java, Indonesia
dewi.rss@gmail.com

***Abstract*—The Key Social Disability Policy (PSBB) requires highly mobile people to use the PeduliLindungi application. One application has several reviews from positive and negative users. Review data can be labeled with two types of emotions: negative sentiment and positive sentiment. Fine-grained sentiment analysis is a type of sentiment analysis that can be used to identify user reactions. One method of sentiment analysis is Multinomial Naïve Bayes. In this research, we used Multinomial Naïve Bayes to perform fine-grained sentiment analysis for users of the PeduliLindungi application. The data used is from the Google Play store. The sentiment class labeling results for the PeduliLindungi review data resulted in 9021 reviews, including a total of 6244 negative reviews and 2777 positive reviews. This research uses a data-sharing model that divides 80% of training data and 20% of test data. Many data imbalances for the two sentiment classes can be overcome by using the SMOTE method. SMOTE has been shown to improve classification accuracy more effectively than non-SMOTE, as applying SMOTE has been shown to improve the performance of imbalanced data. The proper classification method used to classify PeduliLindungi's user ratings is Multinomial Naïve Bayes-SMOTE, which has the highest AUC value.**

***Index Terms*—Data Reviews, Fine-Grained Sentiment Analysis, Multinomial Naïve Bayes, SMOTE**

I. INTRODUCTION

The increasing spread of the COVID-19 virus (Coronavirus Disease 2019) in Indonesia has prompted the government to issue a policy to control the spread of the virus by imposing large-scale restrictions which are often known as Large-Scale Social Restrictions (PSBB). This PSBB aims to break the chain of the spread of the corona virus even though many public facilities are closed. However, several vital sectors such as government facilities, health, and markets or minimarkets remain open during the PSBB while still implementing health protocols. The policy is of course based on Law no. 6 of 2018 concerning Health Quarantine.

To prevent the spread of the Covid-19 outbreak, PT Telekomunikasi Indonesia Tbk (Telkom) and the Ministry of Communication and Information (Kominfo) have collaborated to create the PeduliLindungi application. This application was developed by the government to assist in tracking and stopping the spread of the Covid-19 virus. This application can also be used in the implementation of health surveillance in dealing with the spread of Covid-19, by conducting Tracing, namely tracking people who are in contact with people suspected of being infected with Covid-19. In addition, Tracking is tracking the spread of the coronavirus by seeing who has met people with the Covid-19 virus and organizing Warning and Fencing, namely warnings and surveillance by limiting the movement of someone who is in quarantine or isolation [17]. The use of this application is carried out only during the Covid-19 emergency. Community participation is needed by sharing location data while traveling so that contact history can be traced with Covid-19 sufferers. Because the PeduliLindungi application is intended to protect people who are accessing public facilities to ensure that the activities being carried out are safe and can avoid the spread of Covid-19 [11].

In each application, there are also user ratings and reviews regarding the services and features provided. Reviews can be in the form of suggestions, criticisms, or complaints. This becomes very useful and helpful for other users who will use the application. The collection and sorting of review data is not easy because the number of reviews available on the Google Play site in the comments feature is usually very large. According to Moraes et al. [15], a suitable method for collecting the information data is the use of the web scrapping method. However, there is no sentiment analysis feature to classify or filter between positive and negative reviews, given the large number of app reviews and will continue to grow over time.

Sentiment analysis is computational research of textually expressed opinions, sentiments, and emotions [14]. Sentiment analysis is generally used to analyze something in the future. In this case, sentiment analysis can be applied to mobile application reviews. On mobile reviews sometimes there are writing errors that make it difficult to read. This is caused by several factors such as the proximity of the letters on the keyboard, accidentally typing errors in typing, and no re-checking. From this, it is necessary to reinterpret the word to find out the intent of user reviews. The word will later be classified so that it can be seen which sentiment belongs to it. In this case, a classification method is needed to analyze the reviews.

The problem that often arises in most sentiment analysis studies is that most of the review data tend to be unbalanced (imbalanced datasets) in terms of the number of each class, for example, tend to be a positive class or vice versa. In general, machine learning algorithms will produce a model with a low level of sensitivity to minority classes when receiving an unbalanced data set because it causes poor performance of sentiment analysis classification. , i.e. one-class classification [13], re-sampling [7], and cost-sensitive learning [20]. In this research, the Synthetic Minority Oversampling Technique (SMOTE) method was used to handle the case of unbalanced data.

Sentiment analysis research on online transportation services on Twitter uses the Naïve Bayes classifier method [18]. The research concluded that the Naïve Bayes method can analyze online transportation service review data on Twitter and obtain very high accuracy.

Sentiment analysis research using the Naïve Bayes algorithm has been conducted on Tokopedia application review data [2]. The research concludes that the Naïve Bayes method can analyze sentiment on review data and produces very high accuracy.

Sentiment analysis research using the Naïve Bayes classifier method has been carried out on the quick count results of the 2019 Indonesian Presidential election on Twitter social media [1]. The research concludes that the Naïve Bayes method can produce the best accuracy on the quick count results of the 2019 Indonesian Presidential election on Twitter.

Sentiment analysis research using Naïve Bayes and support vector machines has also been conducted on educational curricula for e-sports by [3]. This research shows that the Naïve Bayes method has the highest accuracy in the case of educational curricula for e-sports.

In addition, Daulay and Asror [8] conducted research on sentiment analysis of application review data on the Google Play Store using the Naïve Bayes method. This research shows that the Multinomial Naïve Bayes method can produce very high accuracy for the case of text-based review classification.

Based on this description, in this research, the author will analyze the sentiment of reviews about the PeduliLindungi application using the supervised learning classification method, namely Multinomial Naïve Bayes. Then compare the performance results with the classification method accompanied by the SMOTE technique to get the best results. In addition, this research is useful for getting words that often appear in positive and negative sentiments, so that the information obtained can be used by the PeduliLindungi application developer as an evaluation material and users who will download the application or other parties who want to do further research.

II. Research Method

The type of data used in this research is primary data. Data was collected through a web scrapping process on the Google Play Store website and data was collected from September 14, 2021, to October 14, 2021, with as many as 9021 PeduliLindungi user reviews. The Web Scrapping method is a method used to collect or extract semi-structured information or data from websites, usually in the form of web pages in markup languages, such as HTML or XHTML, then extract certain information or data from these pages for analysis [19]. In addition, the use of the web scrapping method in the process of collecting and sorting reviews can be facilitated by obtaining information from large amounts of data.

User reviews in this research will be classified into positive and negative sentiments. However, these reviews are still in the form of text without any concrete meaning, so analysis is needed that can classify these user reviews into positive and negative sentiments based on the rating on each user review. Then the process of analyzing and exploring unstructured text data becomes structured to identify patterns in the form of new knowledge and meaningful or useful information with the help of a technology or what is called text mining [12]. During the text mining process, the document text or raw data set is prepared first. The process is referred to as the text preprocessing process. The purpose of text preprocessing is to convert text data that was originally unstructured into structured data. In general, the process of text preprocessing stages can be done as follows.

1. Case Folding is a process of standardizing the form of letters so that there is no difference in meaning.
2. Tokenizing is a word-by-word separation process that does not affect each other from the document text.
3. Filtering is the process of filtering or selecting words in a document.
4. Stemming is changing affixed words into base words.

After passing the text processing stage, the data needs to be modeled so that the data that is still in the form of words can be processed and calculated by converting the data into vectors, which are then assigned a value and weight for each word. This research uses TF-IDF weighting, TF-IDF weighting is one of the models used to calculate word weights in a document using the term frequency model to count the words that appear in each document [16].

Sentiment analysis is the process of processing and understanding textual data automatically to obtain information about the sentiments contained in an opinion sentence, the analysis is carried out to see reviews or opinion trends on an issue or object by someone, regardless of their tendency to have a negative or positive opinion [5]. There are several types of sentiment analysis, namely Fine-Grained Sentiment Analysis, Sentiment Intent Analysis, and Aspect-Based Sentiment Analysis. In this research, Fine-Grained Sentiment Analysis is used which is often referred to as aspect-based opinion mining, and its basic tasks include aspect extraction, opinion identification, and sentiment classification [9]. This analysis includes a type of sentiment analysis that will group responses or opinions into several categories such as positive and negative. VADER Lexicon is used to automatically label sentiment classes on English review data.

The data that has been labeled sentiment class will then be partitioned into training and testing with an 80:20 data split. In this analysis, the amount of training data is 80% and the test data is 20% for each. However, the problem that often arises in most sentiment analysis studies is that most of the review data tend to be unbalanced (imbalanced datasets) in terms of the number of each class, for example, tend to be a positive class or vice versa. Therefore, the Synthetic Minority Oversampling Technique (SMOTE) method is used to handle the case of unbalanced data. SMOTE is an effective oversampling technique approach that is good for handling imbalanced datasets because it can handle overfitting during the oversampling process for the minority class. The SMOTE approach is carried out by generating synthetic data based on the k-nearest neighbors so it is hoped that this technical approach can have an impact on the classification performance results [7].

The learning process is carried out using the supervised learning classification method, namely Multinomial Naïve Bayes. Multinomial Naïve Bayes is a supervised learning method that uses probability and is more focused on text classification [14]. Multinomial Naïve Bayes also has a unique feature, namely that the results obtained for each class are independent. This means that from one document to the next there is no connection at all so the results obtained are purely from the processed document itself. The calculation of the probability of review d having class c can be seen in the following (1).

$$P(c|d) \propto P(c) \prod_{i=1}^{n_d} P(w_i|c) \tag{1}$$

Description:
$P(c|d)$: probability of a class c on document/text d
$P(c)$: prior probability c
$P(w_i|c)$: probability of a word in class c.
The calculation of the probability of positive sentiment or negative sentiment from a review is

$$P(S|c) = \frac{\textit{The number of occurrences of positive}}{\textit{The number of occurrences of positive}} \cdot \frac{\textit{or negative sentiment words in the text}}{\textit{and negative sentiment words in the text}}. \tag{2}$$

From the formula above, it is explained that to calculate the probability of sentiment from a review, that is by comparing the number of occurrences of positive or negative sentiment words with the total number of positive and negative sentiment words in the processed review text.

The number of each prediction class and actual class must be known when measuring the accuracy of the classification because this is done to see the performance of the classification that has been carried out. The following Table I can be used to calculate classification accuracy.

However, if there is a data imbalance case (balance dataset), then choosing which model is best, it can be done by using the AUC (Area Under Curve) value as the basis for measurement, and the accuracy value is considered inaccurate in the imbalance dataset because it only studies the majority of data so that the results The data obtained do not have important information because there may be bias or overfitting [20]. AUC is an evaluation criterion that uses sensitivity or specificity as the basis for measurement [10]. The AUC value is generally in the interval 0.5 – 1.0. The following table description for each AUC value interval can be seen in Table II below [4].

TABLE I
CONFUSION MATRIX

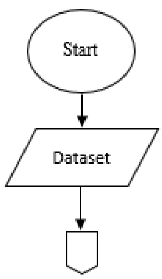
Actual	**Prediction**	
	Positive	**Negative**
Positive	True Positive	False Negative
Negative	False Positive	True Negative

TABLE II
AUC VALUE CATEGORY

AUC Value	**Description**
0,9 - 1,0	Excellent
0,8 - 0,9	Good
0,7 - 0,8	Fair
0,6 - 0,7	Poor
0,5 - 0,6	Failure

This research will use SMOTE and non-SMOTE techniques which will later be used to compare the classification performance when applied to imbalanced datasets so that it can be used to determine the best classification method.

In this research, wordcloud will be used to visualize the results of the classification analysis. Wordcloud is a representation of data that shows a set of important and frequently occurring words in a word. Words that appear frequently are marked with a large number of words and printed in large size in the word cloud [6]. The stages of analysis applied in this research are shown in Fig. 1.



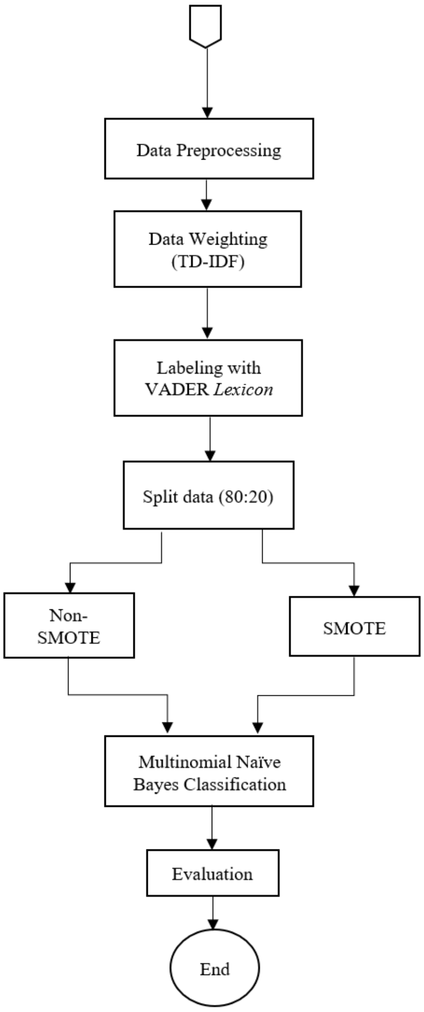


Fig. 1. Research Flow

III. Research Result

A. Text preprocessing

The text preprocessing stage is needed to uniform the spelling of words and letter shapes, reduce vocabulary, and eliminate characters other than letters so that the data text is more structured and informative.

At this stage, the text preprocessing method is used for data cleaning. Several stages that will be carried out include case folding, tokenizing, filtering, and stemming. The results of text preprocessing are in Table III below.

TABLE III
ILLUSTRATION OF TEXT PREPROCESSING

Review	**Text Preprocessing Results**
Worked as intended, however the notification about turning on locations is unnecessary when the app isn't even being used or when you're not checked in on a public space that requires you to scan. Also, one problem I find annoying is you can't pair your phone number with your e-mail account if you had accidentally registered an account for each of them. E-mail and phone number both became separate accounts. I hope there will be a fix for this. I trust the devs the appovertime.	work intend howev notif turn locat unnecessari app isnt even use your check public space requir scan also one problem find annoy cant pair phone number email account accident regist account email phone number becam separ account hope fix trust dev would improv app overtim

B. Fine-Grained Sentiment Analysis

This research uses the VADER lexicon library to carry out the process of labeling sentiment categories on English review data automatically. The review data will only be labeled as two types of sentiment, namely positive sentiment, and negative sentiment. After the category labeling process using sentiment analysis, it can be seen the number of comparisons of reviews labeled positive and negative in Table IV below.

TABLE IV
COMPARISON OF THE NUMBER OF REVIEWS DATA ON SENTIMENT CLASS

Sentiment	**PeduliLindungi**
Positive	2777
Negative	6244
Total	9021

Based on the review text that will be analyzed and has passed the preprocessing stage, a simulation of the sentiment score calculation can be carried out. The formula used to calculate the sentiment score is as follows [4]:

$$Score = (positive\ word\ count) - (negative\ words\ count) \quad (3)$$

If the final score of the calculation simulation $>$ 0, then the result of the review classification is a positive sentiment. The simulation results of the sentiment score calculation can be seen as follows.

TABLE V
SENTIMENT SCORE CALCULATION SIMULATION

Review Text	**Score**	**Sentiment Label**
The plot was **good**, but the characters are **uncompelling** and the dialog is **not great**.	$1 - 2 = -1$	Negative

Words in green indicate that the word has a positive meaning in the lexicon, while words in red have a negative meaning. If the final score of the calculation simulation is negative, then the results of the review classification include a negative sentiment. Based on the results of labeling the sentiment class category, it can be seen that the number of negative reviews is more than the number of positive reviews. The imbalance of a lot of data in the positive and negative classes is an indication that this research has a case of imbalance. Therefore, synthetic data (oversampling) in the minority (positive) class was made using the SMOTE method on the classification method used.

C. Classification

This research will use SMOTE and non-SMOTE techniques which will later be used to compare the classification performance when applied to imbalanced datasets. The SMOTE technique is carried out by generating artificial (synthetic) data on minority class data based on the nearest neighbor (k-nearest neighbors) to balance with the majority class, the SMOTE technique only works or is carried out on the training data used.

Before the classification prediction is made in this research, the data is first partitioned into two types, namely split data. The dataset that has passed the preprocessing stage will then continue with the learning process using the supervised learning classification method, namely Multinomial Naïve Bayes.

Furthermore, classification predictions are made with a supervised learning algorithm. From the results of the classification method used, the AUC, recall, precision, and accuracy values are obtained. In this research, there was a case of data imbalance (balance dataset) so choosing which model is best, it can be done by using the AUC (Area Under Curve) value as the basis for measurement [14] because the accuracy value is considered less precise in the imbalance dataset and can only be measured. only research the majority of data so that the results obtained do not have important information because there may be bias or overfitting.

This analysis will partition the dataset with a ratio of 80% as training data and 20% as test data. The following is a comparison table for the results of split data if not using the SMOTE technique and using the SMOTE technique.

TABLE VI
COMPARISON OF ALGORITHM RESULTS WITHOUT SMOTE AND WITH SMOTE ON PEDULILINDUNGI

Multinomial Naïve Bayes Method	**AUC**	**Recall**	**Precision**	**Accuracy**
Non-SMOTE	0,8727	0,9890	0,8209	0,8404
SMOTE	0,9061	0,8890	0,9025	0,8543

Based on Table VI, the results obtained will be compared to the classification performance using the AUC (Area Under Curve) value. The results show that SMOTE can improve the accuracy of the model on imbalanced data, as evidenced by the AUC value of 0.9061 which indicates that the Multinomial Naïve Bayes method with SMOTE has a very good classification performance on imbalanced data from PeduliLindungi user reviews.

D. Visualization

In this research, wordcloud will be used to visualize the results of the classification analysis. The purpose of the visualization is to extract information in the form of topics that are often discussed by PeduliLindungi users so that information that is considered important can be extracted from the many existing review texts.

Based on Fig. 2(a) information about the words that appear most often, namely care, protect, good, help, better, and thank. That is, these words become words that are often spoken by users because they are printed larger on wordcloud than other words. It was obtained information that positive

(a) (b)

Fig. 2. (a) PeduliLindungi Positive Sentiment Wordcloud; (b) PeduliLindungi Negative Sentiment Wordcloud

ratings from PeduliLindungi users related to the application helped users, protected users, was good, helped users, and became a better application. Meanwhile, based on Fig. 2(b) it can be seen that the words difficult, bad, error, fix, slow, time, and problems contained in negative sentiment are words that are often discussed so that negative reviews of users often complain about these things.

IV. Conclusion

The use of SMOTE is proven to increase the accuracy of the model on unbalanced data (imbalance) seen from the AUC value in the model with SMOTE higher than the AUC value generated by the non-SMOTE model so that the application of SMOTE is more effective in increasing the accuracy of classification accuracy. The classification method used to classify PeduliLindungi user reviews is Multinomial Naïve Bayes with SMOTE. The resulting AUC value is 0.9061, so the method is classified as having a very good classification performance. In general, user reviews have more negative sentiments related to the error application system, while positive reviews relating to the benefits of using the application.

References

[1] L. A. Andika, P., A., N. Azizah, and Respatiwulan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier," *Indonesian Journal of Applied Statistics*, vol. 2, no. 1, 2019.

[2] R. Apriani and D. Gustian, "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia", *Jurnal Rekayasa Teknologi Nusa Putra*, vol. 6, no.1, pp. 54-62, 2019.

[3] R. Ardianto et al., "Sentiment Analysis On E-Sports For Education Curriculum Using Naïve Bayes And Support Vector Machine", *Journal of Computer Science and Information*, vol. 13, no. 2, 2020.

[4] M. Bekkar, H. K. Djemma, and T. A. Alitouche, "Evaluation Measures for Models Assessment over Imbalanced Data Sets", *Journal of Information Engineering and Applications*, vol. 3, no. 10, 2013.

[5] G. Berliana, Shaufiah, and S. Sa'adah, "Klasifikasi Posting Tweet mengenai Kebijakan Pemerintah Menggunakan Naïve Bayesian Classification", *E-Proceeding of Engineering*, vol. 5, no. 1, pp. 1562–1569, 2018.

[6] Q. Castellà and C. Sutton, "Word storms: Multiples of Word Clouds for Visual Comparison of Documents", *WWW 2014 - Proceedings of the 23rd International Conference on World Wide Web*, pp. 665–675, 2014.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Oversampling Technique", *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[8] E. D. P. Daulay and I. Asror, "Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes". *E-Proceeding of Engineering*, vol. 7, no. 2, 2020.

[9] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews", *Information Sciences*, vol. 488, pp. 190–204, 2019.

[10] H. He and Y. Ma, "Imbalanced Learning: Foundations, Algorithms, and Applications", *Wiley-IEEE Press*, 2013.

[11] D. Herdiana, "Aplikasi Peduli Lindungi: Perlindungan Masyarakat Dalam Mengakses Fasilitas Publik Di Masa Pemberlakuan Kebijakan PPKM", *Jurnal Inovasi Penelitian*, vol. 2, no. 6, 2021.

[12] J. -L. Hung and K. Zhang, "Examining Mobile Learning Trends 2003-2008: A Categorical Meta-Trend Analysis Using Text Mining", *Journal of Computing in Higher Education*, vol. 24, pp. 1-17, 2012.

[13] P. Juszczak and R. P. W. Duin, "Combining one-class classifiers to classify missing data", *In Proceedings of the 5th International Workshop MCS, Roli, F., Kittler, J. and Windeatt, T. (eds)*, Springer-Verlag, vol. 3077, pp. 92–101, 2004.

[14] L. Liu and M. T. Ozsu, "Encyclopedia of Database Systems", *In Encyclopedia of Database Systems*. Springer, 2009.

[15] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN" *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013.

[16] W. Nurjanah, R. Perdana, and M. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 12, pp. 1750-1757, 2017.

[17] Nurhidayati et al., "Pengaturan Perlindungan Data Pribadi dalam Penggunaan Aplikasi PeduliLindungi", *Widya Cipta: Jurnal Sekretari dan Manajemen*, vol. 5, no. 1, 2021.

[18] B. M. Pintoko and K., L. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier", *E-Proceeding of Engineering*, vol. 5, no.3, 2018.

[19] M. Turland, "php—architect's Guide to Web Scraping with PHP", 2010.

[20] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem", *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.