Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 4 (2022) 599 - 605 ISSN Media Electronic: 2580-0760

Word2Vec on Sentiment Analysis with Synthetic Minority Oversampling Technique and Boosting Algorithm

Rayhan Rahmanda[1], Erwin Budi Setiawan[2]
[1,2]Informatics, School of Computing, Telkom University
[1]rayhanr@student.telkomuniversity.ac.id, [2]erwinbudisetiawan@telkomuniversity.ac.id

Abstract

Customer opinion is an important aspect in determining the success of a company or service provider. By determining the sentiment of the existing opinion, the company can use it as an evaluation material to improve the quality of the service or product provided. Sentiment analysis can be used as a measure of opinion sentiment with input data in the form of a corpus which will be classified into positive or negative classes to obtain the level of customer satisfaction with a product or service. Aspect-based sentiment analysis can be used by companies to analyze more specifically and find out what aspects need to be improved. In this research, an aspect-based sentiment analysis was conducted on Telkomsel users on Twitter. The data used is 16,992 tweets from users who discuss several aspects such as Telkomsel's services and signals in Twitter. In this research Word2Vec was used for feature expansion to minimize vocabulary mismatch caused by limited words in tweets. The results showed that Word2Vec, Synthetic Minority Oversampling Technique (SMOTE), and Boosting algorithm combination with Logistic Regression classifier achieve highest accuracy of 95.10% for signal aspect and using hyperparameters makes the service aspect get the highest accuracy of 93.34%.

Keywords: sentiment analysis, logistic regression, smote, boosting algorithm, word2vec, twitter

1. Introduction

The use of social media today can be considered very high, everyone can use it freely according to their wishes. Thus, many people use social media to express opinions, comments, criticisms, and others. Twitter is among of the most largest social media that used by its users to express their opinions. All opinions or comments given can be used as important information for decision making for an organization or person. With large amounts of data, it is not possible to manually read and analyze the data, so to overcome this, Sentiment Analysis is needed. Sentiment Analysis, otherwise called Opinion Mining is a process of retrieval of information by analyzing a set of texts that can be classified based on attitudes, emotions and assessments of an entity such as a product or company [1].

Telkomsel is the largest cellular telecommunications operator in Indonesia which is often the talk of social media, especially Twitter. As a service or product provider company, sentiment analysis is certainly needed as an evaluation material in improving product quality. But this is not enough, companies need to know what aspects or things are commented on by the public, so it is necessary to use an aspect-based sentiment analysis approach.

Up to this point, it is known that not many aspect-based sentiment analysis studies have tried to use feature expansion. Similar research on sentiment analysis was conducted by Joshua Acosta et al [2] on user opinions about United States airlines. In his research, a comparison of classification algorithms was carried out with the aim of finding an effective classification algorithm for sentiment classification. The feature expansion used is Word2Vec using 14,000 tweets data about United States flights. Based on the research that has been done, it is found that the Support Vector Machine and Logistic Regression classification methods have the highest accuracy rate of 72.00% compared to the Naïve Bayes method. The advantage of this research is the number of classification algorithms used by using different training models. However, there is a lack in this research, namely that the dataset used is quite small, resulting in an imbalance between sentiment classes.

Ravinder Ahuja et al [3] compared 6 classification algorithms using TF-IDF and N-Grams feature extraction against SS-TWEET (Sentiment Strength

Accepted: 28-06-2022 | Received in revised: 31-07-2022 | Published: 22-04-2022

Twitter Datasets). The dataset contains a total of 4,242 tweets. The purpose of this research was to determine the effectiveness of various feature extractions, namely TF-IDF and Bag of Words on the performance of sentiment analysis. Of all the classification algorithms used, Logistic Regression has the best performance with the value generated using TF-IDF is 50.00% and using N-Grams produces a value of 54.00%. Thus feature extraction using TF-IDF is the right choice to be used in text classification algorithms compared to N-Gram. The advantage of this research lies in the use of many classification algorithms to compare using different feature extractions, but there is no feature expansion used in this research.

Another research related to the use of the expansion feature was carried out by M. Ali Fauzi [4] by conducting a sentiment analysis of 772 reviews of Indonesian-language products using Support Vector Machines. The Word2Vec model is used as a feature expansion in this research to be compared with Bag of Words models such as Binary TF, Raw TF, and TF-IDF. The results obtained are that Word2Vec has the lowest accuracy rate of 70.00%, this is because the dataset used is too small while Word2Vec requires large amounts of data to understand word representation and place words that have the same meaning in closer positions.

Based on these studies, it can be concluded that the logistic regression classification algorithm is quite effective for use in sentiment analysis. It is proven by the accuracy results that are produced compared to other classification algorithms. Although in research [4] the use of Word2Vec word embedding got the lowest accuracy, in this research an aspect-based sentiment analysis was carried out by combining similar feature expansion and TF-IDF as feature extraction. As a solution to the problem of imbalanced data, this research used SMOTE as an oversampling technique and Gradient Boosting is also used to optimize the classification process. The purpose of this research is to analyze the effect of the combination SMOTE with Word2Vec as feature expansion and boosting algorithm on aspect-based sentiment analysis.

The rest of the paper is organized as follows. Section 2 describe of the research method of aspect-based sentiment analysis on Twitter. Section 0 provides the experimental results and followed by the conclusion in Section 4.

2. Research Methods

This research was built using a system shown in Figure 1 that begins with data collection by crawling, labeling, pre-processing, feature extraction using TF-IDF, implementing word2vec feature expansion, classification and finally evaluating the performance.

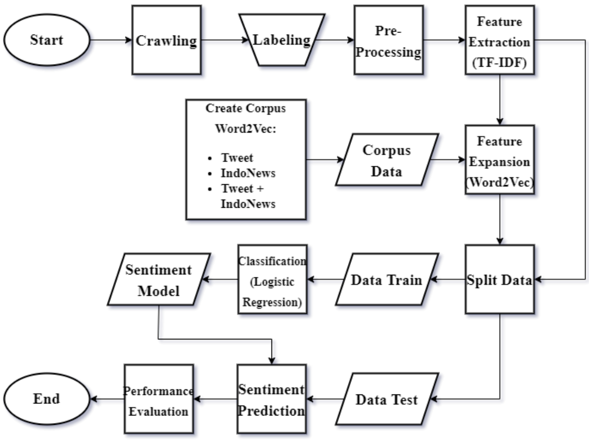

Figure 1. Scenario Overview

2.1 Sentiment Analysis

Sentiment analysis is an analysis of opinions and emotions based on all kinds of text forms [5]. Sentiment analysis can also be seen as a discipline of machine learning, data mining, natural language processing, and computational linguistics which also applies from the sociological and psychological perspectives [6]. With this technique, sentiment can be determined from the opinion of each individual or group of people, in commercial terms, sentiment analysis can be used as evaluation material to improve the services or services provided by a company to customers.

In its implementation, sentiment analysis will classify each opinion in the form of text, reviews, etc. into the positive or negative category. However, there are several things that are complicated in this sentiment analysis, namely if the text contains slang, misspellings, abbreviations, or even the use of emoticons. So it becomes an important task to identify the right sentiment for each word [5].

In sentiment analysis there are many different approaches or techniques, one of which is aspect-based sentiment analysis. Aspect-based sentiment analysis aims to identify the polarity of sentiment towards a more specific aspect within a sentence [7]. The aspect in question is a word that represents an entity. For example, in the following review sentence "The price is expensive but the internet is real fast", in that sentence there are two aspects, namely "price" and "internet" with negative and positive sentiments, respectively.

Thus, the use of aspect-based sentiment analysis can be used to analyze the sentiments of each individual more specifically, namely paying attention to what aspects are commented on by each individual.

2.2 Data Crawling

Data crawl on Twitter is the process of retrieving or downloading data from Twitter server in the form of user data or tweet data by using the Twitter Application Programming Integration (API) [8]. In this study, requires data in the form of tweets that discuss the Telkomsel provider from September 1st, 2021 to February 28th, 2022.

The crawling process is carried out using the Python programming language with the help of the SNScrape library. SNScrape is a library that has many functions to retrieve and collect tweets from Twitter [9]. In collecting this data, several keywords were used as in Table 1 and managed to collect data as many as 16,992 tweets which consists of tweet id, username, date and the tweet itself.

Table 1. Crawling Keywords

Keyword
Telkomsel Signal
Telkomsel Network
Telkomsel is slow
Telkomsel Quality
Telkomsel Down
Telkomsel Roaming
Telkomsel trouble
Telkomsel Error
Grapari
Telkomsel Service
MyTelkomsel
Telkomsel SMS Service

2.3 Data Labeling

In this stage, data labeling is carried out to determine sentiment towards the signal and service aspects of the tweet data. The label is in the form of number 1 (positive), number 0 (neutral), and number -1 (negative). Labeling was done manually by 7 people including the author. This mechanism is carried out to facilitate labeling with very large data. If in this process there are doubts or it is difficult to determine the right sentiment label for a tweet, the votes will be counted from all participants and the sentiment label chosen is the most votes. An example of the labeling results can be seen in Table 2. The number of distributions for each label on each aspect can be seen in Table 3 and Table 4.

Table 2. Example of Tweet Class

Tweet	Sinyal	Pelayanan
“dahlah capek gue mana sinyal telkomsel lg jelek bgt :(”	-1	0
“Ah ajg udah niat ambis ngelarin tugas tapi terhalang sinyal telkomsel yg kaya anjing”	-1	0
“@milkeesway @Telkomsel CS Telkomsel emang selalu responsif. Sabar aja nanti pasti dibales”	0	1
“@jessptra pake Telkomsel biar bagus sinyal nya”	1	0

Table 3. Label Distribution for Signal Aspect

Label	Total
Positive	471
Neutral	6,870
Negative	9,649

Table 4. Label Distribution for Service Aspect

Label	Total
Positive	1,216
Neutral	13,546
Negative	2,225

2.4 Pre-processing

Pre-processing is an important step in handling Twitter data. A tweet is a short message that is limited to many uses of irrelevant symbols, emoticons, misspellings, and slang words. Such characteristics can greatly affect the performance of sentiment analysis. So to avoid this, it is necessary to do pre-processing before feature extraction is carried out [10].

The following are the steps involved in this process. Cleaning text and Case folding is a technique to remove URLs, symbols, punctuation marks and change all letters to lowercase. Then after that tokenization is done to change the sentence into a word token. Generalization is also done to change the form of slang or abbreviations into actual words. Stopword Removal is carried out to remove words that have no meaning when they stand alone. Finally, stemming is a technique of removing affixes for each word and changing each word into its basic form. The results of the pre-processing can be seen in Table 5.

Table 5. Example of pre-processing

Before pre-processing	After pre-processing
“@milkeesway @Telkomsel CS Telkomsel emang selalu responsif. Sabar aja nanti pasti dibales”	['customer', ‘service’, 'telkomsel', 'memang', 'responsif', 'sabar', 'saja', 'balas']

2.5 Feature Extraction

Feature extraction is the process of finding and retrieving features from tweets that can explain the properties of the tweet [8]. In this phase, TF-IDF is used as the feature extraction. Term Frequency-Inverse Document Frequency or what is known as TF-IDF is known as a way to assess the importance of words in a document. Term Frequency is used to count the number of times a word appears in a document. Inverse Document Frequency is used to calculate the importance of a word [3].

$$TF(word) = \frac{Count(word)}{\sum_{i=0}^{n} Count(word_i)} \tag{1}$$

$$IDF(word) = \log\left(\frac{Count(docs)}{Count(word, docs)} + 0{,}01\right) \tag{2}$$

$$TF - IDF = TF(word) \times IDF(word) \tag{3}$$

The technique is used for weighting and extracting keywords from the dataset. The word weighting on the word feature is calculated based on equation (3) which calculates the frequency of occurrence of words in a tweet and the frequency of words in a document. The result of this process is a vector that represents the text and each word is given a weight.

2.6 Feature Expansion

Word2Vec is a feature expansion which aims to process non-numeric data into numeric data. Word embedding Word2Vec is often used as a feature expansion for some text classifications [11]. This model works by building vocabulary from existing training data and then studying it and representing each word into a vector [4]. In text classification, this model is used because word embedding can find semantic and contextual relationships between words [12].

Word2Vec comes in two models: Skip-Gram and Continuous Bag of Words (CBOW). The Skip-Gram model the efficient one to study a large number of word in a form of vector. The Skip-Gram model attempts to make predictions in the area before and after the current word. The input is from the current word [13]. The illustration of Skip-Gram is shown in Figure 2.

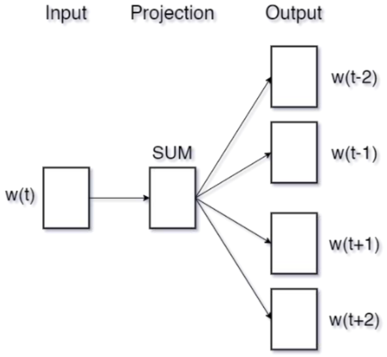

Figure 2. Word2Vec Skip-Gram

Word2Vec output is a list of similar words. Shows the example output of Word2Vec where the target word is “koneksi”.

Table 6. Similar words of koneksi

Rank	Word	Similarity
1	Konektifitas	61.53%
2	Hdspa	58.93%
3	Hspa	58.84%
4	Iptv	58.20%
5	Bandwidth	58.09%
6	Voip	57.90%
7	Tethering	57.77%
8	Lifi	57.68%
9	Wifi	57.59%
10	Internet	57.55%

The Word2Vec model will then be used as a feature expansion. The way it works is to look for a similar word from each word that has a value of zero then if the resulting similar word is match with in the tweet content or sentence, then the zero value will be replaced with weight value of the similar word. As an example, given a tweet “Lagi tethering pakai Telkomsel lambat banget”. Suppose “koneksi” is the word with zero feature value. Word2Vec will find similar words to “koneksi” as shown in Table 6. Since the word “tethering” as one of the similar words appears in the tweet content, then the feature value of “koneksi” will be replaced with “tethering” feature value.

2.7 Classification

In this study, logistic regression was used as a classifier to build a classification model. Logistic Regression (LR) is an algorithm used for classification problems. The algorithm is based on the concept of probability and uses a complex cost function called the Sigmoid function [14]. Logistic Regression can be binomial, ordinal or multinomial. In the case of binomial or what is often called binary logistic regression, it is usually used for data that produces two classes such as "0" and "1". In addition, in the case of multinomial logistic regression, it can produce outputs of more than 2 classes, for example, “positive”, “neutral”, and “negative”.

In the classification process, optimization is carried out by applying SMOTE in handling imbalanced data. The SMOTE algorithm adds an artificial sample to the minority class. Despite that, SMOTE does not oversample based on a simple sample copy. Instead, a small number of new samples are generated beyond the original dataset, which can avoid overfitting [15].

In addition to SMOTE, a boosting algorithm will also be applied during the classification process using Gradient Boosting. Boosting frameworks are a common way to improve the classification performance of learning algorithms that produce highly accurate classifiers by combining coarse and moderately weak classifiers [16].

2.8 Performance Evaluation

Performance evaluation is an important thing that can be done to find out how well the performance of the system that has been built is. The performance can be calculated with the help of the confusion matrix. Confusion matrix is a calculation in the form of a matrix as shown in that states the correct amount of data after being classified and the number of incorrect data after classification [17].

Table 7. Confusion Matrix

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

From Table 7, several values can be calculated, such as Recall, Precision, and Accuracy. Recall (4) is a

DOI: https://doi.org/10.29207/resti.v6i4.4186

calculation to see how much data value has been predicted correctly from all positive classes.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Precision (5) is a calculation of how accurate the model is out of those predicted positive, how many of them are actual positive.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Accuracy (6 which calculates how much of the entire class is predicted correctly. In addition to the calculation of these values, there is also what is called the F1-Score (7) whose value is useful for comparisons between the models used.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{6}$$

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{7}$$

In this study, the metrics that will be used are Accuracy and F1-Score. This is because Accuracy (6 is simply the ratio of the correctly predicted observations to the total observations. However, accuracy is good for use only on balanced datasets, so an F1-Score is required. F1-Score aggregate Recall (4) and Precision (5) measurements under the concept of harmonic mean [18], so it is more useful than Accuracy if there is an uneven class distribution.

3. Result and Discussion

The purpose of this study was to determine the effect of using Word2Vec as an expansion feature combined with SMOTE and a boosting algorithm to optimize the model using Logistic Regression classifier. Evaluation is done by testing several scenarios to find the best combination of methods.

In the first scenario, Logistic Regression and TF-IDF are used as feature extraction to determine the ratio that will be used in the next scenario and to be used as a baseline. The second scenario would be to combine baseline and use Word2Vec and SMOTE. In the second scenario feature expansion, three corpuses are used, namely Corpus Tweet, Corpus IndoNews, and Corpus IndoNews+Tweet. From each of these corpus feature expansions will be tested using Top 1, Top 5, Top 10, and Top 20 Features. Top N features is the process of taking N number of features or words that have the highest similarity value in the corpus. Then in the third scenario, it will combine the baseline with the feature expansion, SMOTE, and add a boosting algorithm for optimization.

For each scenario, five times running or classifications were carried out, the final result was determined by taking the average value of the results of each classification.

3.1 Results

Table 8. Performance of LR + TF-IDF (Baseline)

Ratio	Aspect	Accuracy	F1-Score
90:10	Signal	93.02%	92.33%
	Service	**85.29%**	**83.52%**
80:20	Signal	**93.09%**	**92.44%**
	Service	85.25%	83.40%
70:30	Signal	93.01%	92.28%
	Service	85.04%	83.06%

The performance of logistic regression combined with TF-IDF can be seen in the Table 8. Because at this stage the data used is still imbalanced, to determine which ratio to use as a baseline, we can see from the resulting F1-Score. F1-score can be used on imbalanced data with high true negative class for the assessment of prediction algorithm [19].

Based on the table, it can be seen that for the signal aspect the best ratio produced is 80:20 with an F1-Score of 92.44% and for the service aspect the ratio is 90:10 with an F1-Score of 83.52%. Thus, the selected ratio will then be used as a baseline and in other scenarios.

Table 9. Performance of Baseline + SMOTE

Aspect	Accuracy	F1-Score
Signal	94.37%	94.35%
Service	89.35%	89.27%

Table 9 shows the performance after trying to add SMOTE to the baseline to overcome the imbalance data. From these results, it was found that there was an increase in accuracy in both aspects. The signal aspect has increased accuracy by 1.37% and the service aspect has increased by 4.75%

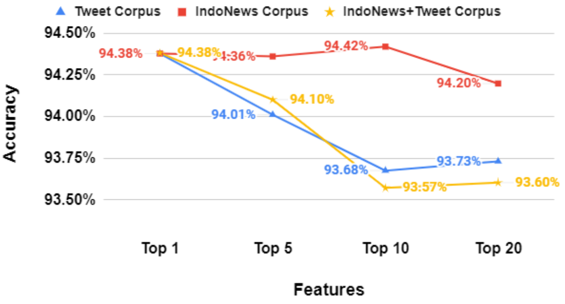

Figure 3. Performance of Baseline + SMOTE + Feature Expansion for Signal Aspect

Figure 3 shows that the highest accuracy for the signal aspect after feature expansion is 94.42% using the IndoNews corpus in Top 10, which means the accuracy has increased by 1.42% compared to the baseline.

DOI: https://doi.org/10.29207/resti.v6i4.4186

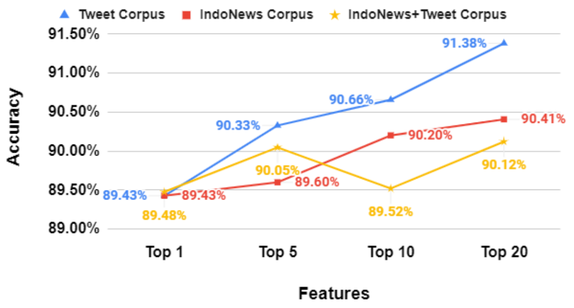

Figure 4. Performance of Baseline + SMOTE + Feature Expansion for Service Aspect

The performance for the service aspect as shown in Figure 4 after the feature expansion got the highest accuracy of 91.38% using a Tweet corpus in Top 20. This indicates that the service aspect has also a significant increase in accuracy, which is 7.14%.

Table 10. Performance of Baseline + Feature Expansion + SMOTE + Boost

Conditions	Aspect	Accuracy	F1-Score
Baseline + Feature Expansion + SMOTE + Boost	Signal	95.10%	95.08%
	Service	93.08%	93.02%

Based on Table 10, the performance results from the combination of Feature Expansion, SMOTE, and boost are shown. At this stage the signal aspect using the IndoNews corpus with Top 10 features and for the service aspect using the Tweet corpus with Top 20. With this method, both aspects have improved accuracy which is quite good. The increase in accuracy for the signal aspect is 2.16% and for the service aspect it is 9.13%.

With the condition as shown in the Table 10, re-experimentation is carried out by adding hyperparameters such as solver, regularization, and C parameters that can control the penalty strength in the classifier to see whether the hyperparameters have an effect or not.

Table 11. Performance of Baseline + Feature Expansion + SMOTE + Boost + Hyperparameters

Conditions	Aspect	Accuracy	F1-Score
Baseline + Feature Expansion + SMOTE + Boost + Hyperparameter	Signal	94.86%	94.83%
	Service	93.34%	93.22%

The results from the Table 11 show that after using hyperparameters, the signal aspect gets an accuracy of 94.86% which indicates that there is a 0.25% decrease in accuracy compared to without using hyperparameters, while for the service aspect the accuracy obtained is 93.34%, which means an increase of 0.28%.

3.2 Discussion

Table 12. Comparison Results

Conditions	Accuracy	
	Sinyal	Pelayanan
LR + TF-IDF (Baseline)	93.09%	85.29%
Baseline + SMOTE	94.37% (+1.37%)	89.35% (+4.75%)
Baseline + SMOTE + Feature Expansion	94.42% (+1.42%)	91.38% (+7.14%)
Baseline + SMOTE + Feature Expansion + Boost	**95.10% (+2.16%)**	93.08% (+9.13%)
Baseline + SMOTE + Feature Expansion + Boost + Hyperparameter	94.86% (+1.90%)	**93.34% (+9.34%)**

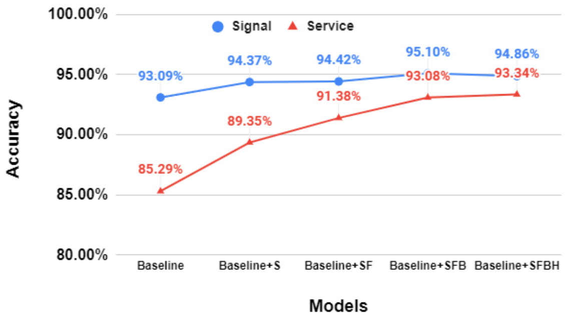

Figure 5. Comparison Results

S = SMOTE, F = Feature Expansion, B = Boosting Algorithm, H = Hyperparameter.

Based on these experiments results on Table 12 and Figure 5, SMOTE can work well in dealing with data imbalances, this is proven by increasing accuracy for both aspects from the previous accuracy for the signal aspect 93.09% increased to 94.37% and for the service aspect from 85.29% to 89.35%.

In addition, Word2Vec as feature expansion is important in this research because it can improve accuracy as well by identifying missing words in tweets and replacing them with words that are related semantically or meaning [20] and gets accuracy of 94.42% for signal aspect and 91.38% for service aspect.

The combination of Word2Vec, SMOTE, and Boosting Algorithm is effectively in increasing accuracy, the signal aspect gets the highest accuracy of 95.10%. However, when using hyperparameters the signal aspect decreased by 0.25% while the service aspect gets the highest accuracy of 93.34%.

4. Conclusion

In this research, combination of Word2Vec as a feature expansion, SMOTE as imbalance data handler and boosting algorithm for classification optimization has been implemented in aspect-based sentiment analysis. The dataset used in this study is 16,992 tweets that discuss several aspects such as signal and service

DOI: https://doi.org/10.29207/resti.v6i4.4186

aspects. Data in the form of news articles from Indonesian media are also used during Feature Expansion. This study applies Logistic Regression as a Classifier and Gradient Boosting for optimization. Based on the final results, it was found that Word2Vec, SMOTE, and Boosting Algorithm combination with Logistic Regression classifier has high accuracy results for signal aspect with the result of 96.10%. Meanwhile, the use of hyperparameters for logistic regression only succeeded in increasing accuracy for the service aspect with the result of 93.34%. Thus, it is hoped that future research can try to use different hyperparameters and also different classification methods with a larger amount of data.

Reference

[1] Y. Cahyono, “Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequency,” *JURNAL INFORMATIKA UNIVERSITAS PAMULANG*, vol. 14, no. 1, 2017.

[2] J. Acosta, N. Lamaute, M. Luo, E. Finkelstein, and A. Cotoranu, “Sentiment Analysis of Twitter Messages Using Word2Vec,” 2017.

[3] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The impact of features extraction on the sentiment analysis,” in *Procedia Computer Science*, 2019, vol. 152, pp. 341–348. doi: 10.1016/j.procs.2019.05.008.

[4] M. A. Fauzi, “Word2Vec model for sentiment analysis of product reviews in Indonesian language,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, p. 525, Feb. 2019, doi: 10.11591/ijece.v9i1.pp525-530.

[5] P. Baid, A. Gupta, and N. Chaplot, “Sentiment Analysis of Movie Reviews using Machine Learning Techniques,” *International Journal of Computer Applications*, vol. 179, no. 7, pp. 45–49, Dec. 2017, doi: 10.5120/ijca2017916005.

[6] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,” *Knowledge and Information Systems*, vol. 60, no. 2, pp. 617–663, Aug. 2019, doi: 10.1007/s10115-018-1236-4.

[7] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, “A Challenge Dataset and Effective Models for Aspect-Based Sentiment Analysis,” 2019. [Online]. Available: https://github.com/siat-

[8] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, “Feature expansion for sentiment analysis in twitter,” in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Oct. 2018, vol. 2018-October, pp. 509–513. doi: 10.1109/EECSI.2018.8752851.

[9] T. Sarkar and N. Rajadhyaksha, “TLA: Twitter Linguistic Analysis,” Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.09710

[10] S. Elbagir and J. Yang, “Sentiment analysis of twitter data using machine learning techniques and scikit-learn,” Dec. 2018. doi: 10.1145/3302425.3302492.

[11] Naufal Adi Nugroho and Erwin Budi Setiawan, “Implementation Word2Vec for Feature Expansion in Twitter Sentiment Analysis,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 837–842, Oct. 2021, doi: 10.29207/resti.v5i5.3325.

[12] R. I. Kurnia, Y. D. Tangkuman, and A. S. Girsang, “Classification of user comment using word2vec and SVM classifier,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1, pp. 643–648, 2020, doi: 10.30534/ijatcse/2020/90912020.

[13] R. P. Nawangsari, R. Kusumaningrum, and A. Wibowo, “Word2vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study,” in *Procedia Computer Science*, 2019, vol. 157, pp. 360–366. doi: 10.1016/j.procs.2019.08.178.

[14] A. Poornima and K. Sathiya Priya, *A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques*. 2020.

[15] S. Guo, Y. Liu, R. Chen, X. Sun, and X. Wang, “Improved SMOTE Algorithm to Deal with Imbalanced Activity Classes in Smart Homes,” *Neural Processing Letters*, vol. 50, no. 2, pp. 1503–1526, Oct. 2019, doi: 10.1007/s11063-018-9940-3.

[16] J. Tanha, “A multiclass boosting algorithm to labeled and unlabeled data,” *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 12, pp. 3647–3665, Dec. 2019, doi: 10.1007/s13042-019-00951-4.

[17] A. Filcha and M. Hayaty, “Implementasi Algoritma Rabin-Karp untuk Pendeteksi Plagiarisme pada Dokumen Tugas Mahasiswa (Rabin-Karp Algorithm Implementation to Detect Plagiarism on Student’s Assignment Document),” 2019.

[18] M. Grandini, E. Bagli, and G. Visani, “Metrics for Multi-Class Classification: an Overview,” Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05756

[19] Z. DeVries *et al*., “Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability,” *Spine Journal*, vol. 21, no. 7, pp. 1135–1142, Jul. 2021, doi: 10.1016/j.spinee.2021.02.007.

[20] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, “Feature expansion using word embedding for tweet topic classification,” Mar. 2017. doi: 10.1109/TSSA.2016.7871085.

DOI: https://doi.org/10.29207/resti.v6i4.4186