Jurnal ELTIKOM :
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 9, No. 1, June 2025, page. 91-97
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v9i1.1496
Available online at http://eltikom.poliban.ac.id

IMBALANCED TEXT CLASSIFICATION ON TOURISM REVIEWS USING ADABOOST NAÏVE BAYES

Ika Oktavia Suzanti[1*], Fajrul Ihsan Kamil[1], Eka Mala Sari Rochman[1], Huzain Azis[2], Alfa Faridh Suni[3], Fika Hastarita Rachman[1], Firdaus Solihin[1]
[1)] Informatics Engineering Department, University of Trunojoyo Madura, Bangkalan, Indonesia
[2)] Malaysian Institute of Information Technology (MIIT), Universiti Kuala Lumpur, Kuala Lumpur, Malaysia
[3)] Computer Science Department, Newcastle University, United Kingdom
e-mail: iosuzanti@trunojoyo.ac.id, 200411100172@student.trunojoyo.ac.id, em_sari@trunojoyo.ac.id, huzain.azis@s.unikl.edu.my, a.f.Suni2@newcastle.ac.uk, fika.rachman@trunojoyo.ac.id, fsolihin@trunojoyo.ac.id

Received: 25 March 2025 – Revised: 19 May 2025 – Accepted: 23 May 2025

ABSTRACT

Hidden paradise is a term that aptly describes the island of Madura, which offers diverse tourism potential. Through the Google Maps application, tourists can access sentiment-based information about various attractions in Madura, serving both as a reference before visiting and as evaluation material for the local government. The Multinomial Naïve Bayes method is used for text classification due to its simplicity and effectiveness in handling text mining tasks. The sentiment classification is divided into three categories: positive, negative, and mixed. Initial analysis revealed an imbalance in sentiment data, with most reviews being positive. To address this, sampling techniques—both oversampling and undersampling—were applied to achieve a more balanced data distribution. Additionally, the Adaptive Boosting ensemble method was used to enhance the accuracy of the Multinomial Naïve Bayes model. The dataset was split into training and testing sets using ratios of 60:40, 70:30, and 80:20 to evaluate the model's stability and reliability. The results showed that the highest F1-score, 84.1%, was achieved using the Multinomial Naïve Bayes method with Adaptive Boosting, which outperformed the model without boosting, which had an accuracy of 76%.

Keywords: *Imbalanced Data, Naïve Bayes, Sentiment Analysis, Text Classification, Text Mining.*

I. INTRODUCTION

THE openness of public information is expected to support the advancement of science and technology. In mid-2016, Google Maps introduced a feature that allows users to rate and review places they have visited. Through Google Maps, user-generated reviews of tourist attractions can be collected, offering valuable insights for analyzing the sentiments expressed by visitors [1]. Hidden paradise is a fitting term for the island of Madura, which holds significant tourism potential—including historical, natural, cultural, and religious tourism—spread across four regencies: Sumenep, Pamekasan, Sampang, and Bangkalan [2]. Typically, prospective visitors are influenced by the reviews they read about a destination.

Sentiment analysis is an application of text mining that focuses on identifying and categorizing opinions expressed in textual form [3]. It can be used to analyze opinions, evaluations, or attitudes related to various subjects or social events. Since the data source is often social media—used widely by the public—sentiment analysis is inherently tied to society [4]. As a type of text classification, sentiment analysis can employ supervised learning methods such as Naïve Bayes [5], Support Vector Machines [6], K-Nearest Neighbors [7], and Decision Trees [8].

Naïve Bayes is widely used for processing text data due to its simplicity; it calculates the probability of a word belonging to a particular class [9], assuming that each word is an independent variable [10]. Several studies using Naïve Bayes have demonstrated good accuracy in classifying both short texts—such as reviews [11], sentiment data [12], and social media posts [13]—and longer texts, including web pages [14]. Compared to other classification methods, Naïve Bayes has shown competitive performance.

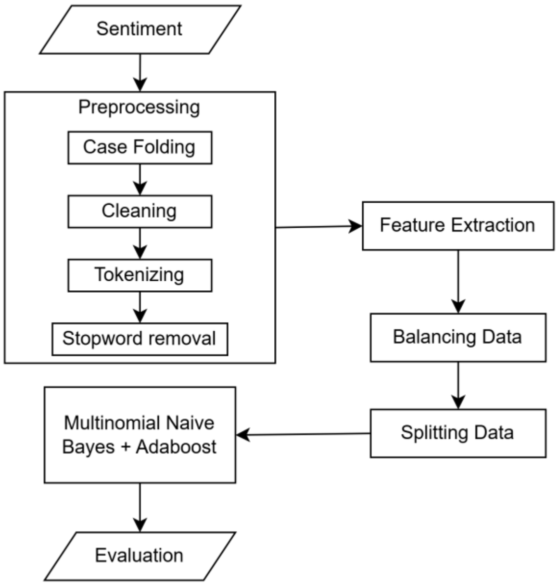

Figure 1. Methodology System

For instance, a study by [15] reported that Naïve Bayes achieved an accuracy of 98.73%, outperforming Decision Tree (97.55%) and SVM (97.73%). Similarly, research by[16] on Chinese language text classification found that Naïve Bayes reached an accuracy of 73.9%, higher than KNN (67.9%), SVM (70.8%), and logistic regression (73%) for long text data.

A large number of features is one of the main challenges in text classification. To address this, many studies have incorporated feature selection processes to reduce the dimensionality of the features being analyzed [17]. Feature selection is expected to enhance the performance of classification methods and yield higher accuracy. The boosting technique can also be used to select features from various samples, based on the outcomes of previous iterations. One of the key goals of boosting is to improve weak learners that suffer from high bias [18]. Adaptive Boosting, an iterative method, identifies optimal features and passes them to the primary classifier to improve classification performance. In a study by [19], Adaptive Boosting increased the accuracy of Naïve Bayes from 88.83% to 99.26%.

An initial analysis of tourist attraction reviews on Madura Island revealed an imbalance in sentiment distribution, with most reviews being positive. Imbalanced data refers to a condition in which some classes contain many samples while others contain significantly fewer [20]. In such cases, the classification process typically favors the majority class [21], which may lower accuracy when the test data includes many samples from the minority class. To address this, resampling techniques are applied to balance the number of samples across all classes [22].

Based on this background, the aim of this study is to propose a hybrid model combining Naïve Bayes and Adaptive Boosting to classify public sentiment toward Madura tourism. The model is further enhanced with resampling techniques to handle imbalanced data and improve overall accuracy.

II. Research Method

A. Methodology

This study applies the Naïve Bayes and Adaptive Boosting methods, with the system flow illustrated in Figure 1. The system flow represents the methodological steps of this research. The process begins by inputting reviews related to selected tourist attractions. These reviews undergo preprocessing, followed by feature extraction using TF-IDF, data balancing, data splitting, and classification. After classification, model evaluation is conducted using a confusion matrix. The data splitting in this study employs three training-to-testing ratios: 60:40, 70:30, and 80:20.

B. Data Preparation

The data used in this study consist of Indonesian-language reviews of tourist attractions on Madura Island, collected from Google Maps using a scraper tool. The dataset includes reviews from the last four

TABLE 1
DATA PREPROCESSING

Raw data	Case Folding	Cleaning	Tokenization	Stopword re-moval
Bagus apalagi kalau cuaca mendukung dan datang di waktu yang tepat sekitar jam 8 atau 9, karena matahari sudah berada diatas tapi suhu ga panas terik dan lebih menaikkan pesona alamnya.	bagus apalagi kalau cuaca mendukung dan datang di waktu yang tepat sekitar jam 8 atau 9, karena matahari sudah berada diatas tapi suhu ga panas terik dan lebih menaikkan pesona alamnya.	bagus apalagi kalau cuaca mendukung dan datang di waktu yang tepat sekitar jam atau karena matahari sudah berada diatas tapi suhu ga panas terik dan lebih menaikkan pesona alamnya	[bagus, apalagi, kalau, cuaca, mendukung, dan, datang, di, waktu, yang, tepat, sekitar, jam, atau, karena, matahari, sudah, berada, diatas, tapi, suhu, ga, panas, terik, dan, lebih, menaikkan, pesona, alamnya]	[bagus, cuaca, mendukung, jam, matahari, diatas, suhu, panas, terik, menaikkan, pesona, alamnya]

years (2021–2024), totaling 3,000 reviews across nine tourist destinations, with approximately 400 reviews per destination. The nine attractions are Lombang Beach, Slopeng Beach, Sembilan Beach, Asta Tinggi, Jeddih Hill, Toroan Waterfall, Bebek Sinjay, Api Tak Kunjung Padam, and Syaikhona Kholil Mosque. Of the 3,000 reviews, 2,435 are positive, 250 are negative, and 315 are mixed.

Data preprocessing is performed to transform raw data into a format that is compatible with the applied method, enabling accurate and relevant results [23]. The results of the preprocessing in this study are presented in Table 1, encompassing several essential steps to prepare the sentiment text for classification. The first step is case folding, which involves converting all text to lowercase letters to ensure consistency in analysis. Next is the cleaning process, where unnecessary elements such as links, URLs, @ symbols, hashtags (#), numbers, and punctuation marks are removed to reduce noise and improve classification accuracy. Following this, tokenization is performed by breaking the text into smaller units called tokens—specifically words in this study—based on spaces between them. Lastly, the stopword removal stage eliminates words that do not carry significant meaning, such as "yang", "dan", "di", and "ke". The stopword list used in this stage is derived from the NLTK corpus.

In the text mining process, after preprocessing, the next step is data transformation—converting text data into numerical form. This study employs the TF-IDF (Term Frequency–Inverse Document Frequency) method, which transforms words into numerical values based on their frequency in a review. TF-IDF weighting consists of three stages. The first step is calculating Term Frequency (TF), which measures how often a term appears in a document, indicating its importance within that document. The second step is calculating Inverse Document Frequency (IDF), which assesses how commonly a term appears across all documents—terms that appear less frequently receive higher weights. The final step is multiplying the TF and IDF results to obtain the TF-IDF score [24].

C. Resampling Technique

A common problem when using text data in research is the imbalance in class distribution and the large number of features [25]. Imbalanced data refers to a condition where the number of instances in one class significantly differs from the others. Addressing this issue is important, as balanced data allows machine learning models to make more accurate predictions [26]. Resampling techniques are commonly used to handle imbalanced data. These techniques rebalance the dataset using various sampling algorithms to adjust the number of samples in each class. The rebalanced data is then used for training a classification algorithm [27]. Resampling methods include oversampling, undersampling, and synthetic techniques. In this study, the review data were imbalanced, with a majority of positive reviews. Therefore, after feature extraction, a data balancing process was applied using random oversampling and random undersampling.

Oversampling increases the number of instances in the minority class until it equals the number in the majority class [25]. Random Oversampling (ROS) randomly selects instances from the minority class and duplicates them until the class sizes are equal. However, this method may result in overfitting due to the generation of duplicate instances [21]. In contrast, undersampling reduces the number of instances in the majority class to match the minority class [25]. Random Undersampling (RUS) works by randomly selecting and removing instances from the majority class until both class sizes are equal.

D. Naïve Bayes

Naïve Bayes Classifier is a classification method based on Bayes' Theorem, which predicts the probability of future events using past data [28]. One advantage of Naïve Bayes is that the classification

$$P(Ci) = \frac{Nci}{N} \tag{1}$$

$$P(w|c) = \frac{W_{ct} + 1}{(\sum_{W' \in V} W_{ct'}) + B'} \tag{2}$$

$$V_{MAP} = argmax\ P(x1, x2, x3, \ldots xn)\ P(C) \tag{3}$$

$$e_m = \frac{\sum_{i=1}^{n} w_i^{(m)} l(G_m(x_i \neq y_i)}{\sum_{i=1}^{n} w_i^{(m)}} \tag{4}$$

$$\alpha_m = \ln \frac{1-e_m}{e_m} \tag{5}$$

$$w_i^{(m+1)} = \frac{w_i^{(m)}}{Z_m}\ exp(\alpha_m l(G_m(x_i)\ \neq yi)) \tag{6}$$

$$T(x) = arg \max_{c} \sum_{m=1}^{M} \alpha_m l(G_m(x_i) = c) \tag{7}$$

$$F1\ score = \frac{2 \text{ x (Recall x Precision)}}{\text{Recall} + \text{Precision}} \tag{8}$$

process can be adapted to the characteristics and needs of individual datasets [29]. Multinomial Naïve Bayes is a development model of the Naïve Bayes algorithm and is effective for text classification tasks [30]. In this model, classes are determined not only by the presence of words but also by their frequency in a document [10]. Multinomial Naïve Bayes is highly suitable for sentiment classification in this study because it offers advantages such as ease of implementation, fast computation, and satisfactory accuracy [29].

The first step is to calculate the class probability using (1). $P(Ci)$ is a probability of class *I*, Nci is the number of instances in class *I,* and *N* is the total number of documents. To estimate the conditional probability, (2) is used. $P(w|c)$ is conditional probability of word *w* given class *c*. Wct is the frequency of term *t* in category *c*. $(\sum_{W' \in V} W_{ct'})$ is the total frequency of all terms in category *c*. B' is the IDF value for the entire document set.

To avoid issues with zero values, Laplace Smoothing is applied by adding one to each Wct frequency used in the conditional probability calculation. The final step is to calculate the maximum a posteriori (MAP) value of class *V* on the test data using (3). $P(x1)$ is the word probability. $P(C)$ is a class probability.

The document is assigned to a class based on the MAP value. If the MAP value for the positive class is higher than that of the negative class, the document is classified as positive, and vice versa. Adaptive Boosting, commonly known as AdaBoost, is an ensemble algorithm designed to improve classifier accuracy. AdaBoost can consistently reduce errors from base learners, resulting in better classification performance than random guessing [31].

AdaBoost and its variants have been successfully applied across various domains due to their strong theoretical foundations, high prediction accuracy, and simplicity. The algorithm operates as follows [27].

1) Initialization

Assign an initial weight to each observation: $w_i = 1/n$ where *n* is the number of training data observations, and i = 1, 2, 3, ..., *n*.

2) Iteration (m = 1, 2, 3, ..., *M*)

- Calculate the classification error on misclassified instances using (4). Here, e_mm is the error at the m-th iteration, $w_i^{(m)}$ is the weight of observation i, $G_m(x_i)$ is the predicted label, y_i is the true label, and $l(G_m(x_i \neq y_i)$ is an indicator function that equals 1 if the predicted class differs from the actual

TABLE 2
CONFUSION MATRIX

Current	Prediction		
	Positive	Negative	*Mixed*
Positive	*True* Positive (TP)	*False* Negative 1 (FN 1)	False Mixed 1 (FM1)
Negative	*False* Positive 1 (FP 1)	*True* Negative (TN)	False Mixed 2 (FM2)
Mixed	*False* Positive 2 (FP2)	*False* Negative 2 (FN2)	*True Mixed* (TM)

class, and 0 otherwise. If $e_m > 1 - \frac{1}{c}$ (where c is the number of classes), the iteration stops. If $e_m \leq 1 - \frac{1}{c}$, the process continues with the calculation of the vote weight.

- Compute the classification vote weight using (5).
- Update the weights using (6). Here, $Z_m = \sum_{i=1}^{n} w_i^{(m)}$, Z_m is the normalization factor to ensure that weights sum to one in each iteration.

3) Determine the final class prediction using the weighted vote of classifiers with (7).

In this function, $T(x)$ represents the final predicted class for input x, c is a class label (either 0 or 1), and $l(G_m(x_i) = c)$ is an indicator function that equals 1 if the predicted class matches c, and 0 otherwise.

E. Evaluation

According to [32], a confusion matrix is a tool used to analyze the predictions of a machine learning model. It provides a summary of the correct and incorrect prediction results generated from the classification process. In this study, the confusion matrix is presented in Table 2.

The evaluation results include several metrics such as F1-score, accuracy, recall, and precision. In this study, the main metric used is the F1-score. This choice is based on the fact that accuracy tends to be high when the dataset is balanced, whereas the F1-score is more effective for evaluating models trained on imbalanced data. The F1-score is calculated using (8).

III. RESULTS AND DISCUSSION

In this study, classification was performed on tourist attraction reviews from Madura Island using a combination of the Multinomial Naïve Bayes and Adaptive Boosting methods. Due to the presence of imbalanced data, oversampling and undersampling techniques were also applied to balance the dataset across different test scenarios. The data were split into training and testing sets using three ratios: 80:20, 70:30, and 60:40.

A. Multinomial Naïve Bayes

In the first three scenarios, classification was conducted using the Multinomial Naïve Bayes method with three different approaches: without resampling, with undersampling, and with oversampling. The F1-score results from applying these methods are presented in Table 3.

As shown in Table 3, the evaluation results using the Multinomial Naïve Bayes method yield the highest F1-score under the 80:20 data split. Specifically, Multinomial Naïve Bayes with oversampling achieved an F1-score of 76.4%, with undersampling 70.3%, and without any resampling 60.0%. The improvement in F1-score as the proportion of training data increases suggests that more training data may enhance model accuracy, as the machine is exposed to a larger dataset during learning.

However, this trend does not guarantee continuous improvement. At a certain point, increasing the amount of training data may not lead to better performance and could even result in decreased performance. Such a decrease may occur due to limited variation in the testing set or because the undersampling technique significantly reduces the volume of data, thereby affecting classification accuracy when compared to the oversampling technique.

B. Adaptive Boosting

In the next three scenarios, the classification process was carried out using the Multinomial Naïve Bayes method with the addition of Adaptive Boosting in the feature selection stage. Resampling techniques—both undersampling and oversampling—were also applied. In these scenarios, the value of the n_estimators parameter for Adaptive Boosting was set to 300, selected based on 10% of the total dataset. The n_estimators parameter typically ranges from 1 to 1000, where a higher value generally results in improved classification accuracy. Table 4 presents the evaluation results, showing that the best performance was achieved in the 80:20 data split using Multinomial Naïve Bayes + Adaptive Boosting with

TABLE 3
F1-SCORE VALUES OF MULTINOMIAL NAÏVE BAYES

Data Split	MNB + Undersampling	MNB + Oversampling	MNB (No Resampling)
80:20	68.3%	76.4%	59.3%
70:30	70.3%	76.1%	60%
60:40	70.1%	75.6%	57.5%

TABLE 4
ADAPTIVE BOOSTING F1-SCORE VALUES

Data Split	MNB + AdaBoost + Undersampling	MNB + AdaBoost + Oversampling	MNB + AdaBoost
80:20	66 %	84.1%	58.1%
70:30	68.5%	83.8%	59.5%
60:40	68%	81.6%	58.8%

oversampling, yielding an F1-score of 84.1%. The application of Adaptive Boosting across the test scenarios significantly improved performance, with gains of up to 8%.

Based on the results in Tables 3 and 4, a comparative analysis across the six scenarios indicates that the best F1-score was achieved using the combination of Multinomial Naïve Bayes and Adaptive Boosting with Random Oversampling. This scenario yielded the highest score of 84.1%. In all scenarios, the 80:20 data split consistently provided the best performance. This finding suggests that a larger proportion of training data allows the model to learn more effectively and produce more accurate predictions. Furthermore, Adaptive Boosting significantly enhances classification accuracy, especially when the volume of processed data is increased.

IV. CONCLUSION

Based on the series of tests conducted using the Multinomial Naïve Bayes method combined with Adaptive Boosting on tourism reviews from Madura, the following conclusions can be drawn: the best classification performance was achieved with the 80:20 data split, resulting in an accuracy of 84.1%. The application of Adaptive Boosting as an ensemble method successfully improved accuracy by up to 7% across several scenarios. However, Adaptive Boosting did not perform well when combined with the Random Undersampling technique, as it led to a decrease in accuracy of up to 16% across all data splits.

Future research can explore comparisons with other classification methods or further enhancements to the Multinomial Naïve Bayes and Adaptive Boosting combination. This could involve parameter tuning for Adaptive Boosting or alternative approaches such as data augmentation to address imbalanced datasets more effectively.

REFERENCES

[1] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Syst Appl*, vol. 167, p. 114324, 2021. https://doi.org/10.1016/j.eswa.2020.114324

[2] S. Arifin, "Digitalisasi pariwisata madura," *Jurnal Komunikasi*, vol. 11, no. 1, pp. 53–60, 2017. https://doi.org/10.21107/ilkom.v11i1.2835

[3] E. D. Madyatmadja, B. N. Yahya, and C. Wijaya, "Contextual text analytics framework for citizen report classification: A case study using the Indonesian language," *IEEE Access*, vol. 10, pp. 31432–31444, 2022. https://doi.org/10.1109/ACCESS.2022.3158940

[4] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Syst Appl*, vol. 110, pp. 298–310, 2018. https://doi.org/10.1016/j.eswa.2018.06.022

[5] H. Hartono, A. Hajjah, and Y. N. Marlim, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Judul Berita," *Jurnal Simantec*, vol. 12, no. 1, pp. 37–46, 2023. https://doi.org/10.21107/simantec.v12i1.19398

[6] F. Fitriyani and T. Arifin, "Penerapan Word N-Gram Untuk Sentiment Analysis Review Menggunakan Metode Support Vector Machine (Studi Kasus: Aplikasi Sambara)," *Sistemasi: Jurnal Sistem Informasi*, vol. 9, no. 3, pp. 610–621, 2020. https://doi.org/10.32520/stmsi.v9i3.95

[7] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," in *2015 international conference on data and software engineering (ICoDSE)*, 2015, pp. 170–174. https://doi.org/10.1109/ICODSE.2015.7436992

[8] F. Gürcan, "Multi-class classification of turkish texts with machine learning algorithms," in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2018, pp. 1–5. https://doi.org/10.1109/ISMSIT.2018.8567307

[9] W. Zhang and F. Gao, "An improvement to naive bayes for text classification," *Procedia Eng*, vol. 15, pp. 2160–2164, 2011. https://doi.org/10.1016/j.proeng.2011.08.404

[10] V. Vijay and P. Verma, "Variants of Naive Bayes algorithm for hate speech detection in text documents," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, 2023, pp. 18–21. https://doi.org/10.1109/AISC56616.2023.10085511

[11] J. B. Awotunde, S. Misra, V. Katta, and O. C. Adebayo, "An ensemble-based hotel reviews system using naive bayes classifier," 2023. https://doi.org/10.32604/cmes.2023.026812

[12] H. Candra, E. D. Madyatmadja, J. Nathaniel, and M. R. Jonathan, "Sentiment Analysis on Indonesian Telegram Reviews Using Naive Bayes, SVM, Random Forest, and Boosting Models," in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2024, pp. 493–498. https://doi.org/10.1109/ICITISEE63424.2024.10730718

[13] U. K. Agrawal, B. V Ramana, D. Singh, and N. Panda, "Leveraging LSTM and Multinomial Naive Bayes for Nuanced Textual-Based Sentiment Analysis," *SN Comput Sci*, vol. 5, no. 8, p. 1090, 2024. https://doi.org/10.1007/s42979-024-03463-3
[14] R. Cekik, "A New Filter Feature Selection Method for Text Classification," *IEEE Access*, 2024. https://doi.org/10.1109/ACCESS.2024.3468001
[15] Y. A. Singgalen, "Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci*, vol. 4, no. 3, pp. 1634–1646, 2022. https://doi.org/10.47065/bits.v4i3.2766
[16] P. Wang *et al.*, "Classification of proactive personality: Text mining based on weibo text and short-answer questions text," *Ieee Access*, vol. 8, pp. 97370–97382, 2020. https://doi.org/10.1109/ACCESS.2020.2995905
[17] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature selection for text classification with Na\"\ive Bayes," *Expert Syst Appl*, vol. 36, no. 3, pp. 5432–5435, 2009. https://doi.org/10.1016/j.eswa.2008.06.054
[18] M. R. H. Suryanto and D. W. Utomo, "Pembelajaran Ensemble Untuk Klasifikasi Ulasan Pelanggan E-commerce Menggunakan Teknik Boosting," *Infotekmesin*, vol. 15, no. 2, pp. 238–244, 2024. https://doi.org/10.35970/infotekmesin.v15i2.2314
[19] D. Setiyawati and N. Cahyono, "Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia," *The Indonesian Journal of Computer Science*, vol. 12, no. 1, 2023. https://doi.org/10.33022/ijcs.v12i1.3154
[20] M. R. Indrahimawan, P. I. Santosa, and T. B. Adji, "Handling Data Imbalance Using Text Augmentation For Classifying Public Complaints," in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 2023, pp. 284–289. https://doi.org/10.1109/IC3INA60834.2023.10285813
[21] J. Song, X. Huang, S. Qin, and Q. Song, "A bi-directional sampling based on K-means method for imbalance text classification," in *2016 IEEE/ACIS 15th international conference on computer and information science (ICIS)*, 2016, pp. 1–5. https://doi.org/10.1109/ICIS.2016.7550920
[22] X. Chen, W. Zhang, S. Pan, and J. Chen, "Solving data imbalance in text classification with constructing contrastive samples," *IEEE Access*, vol. 11, pp. 90554–90562, 2023. https://doi.org/10.1109/ACCESS.2023.3306805
[23] J. Laksana and A. Purwarianti, "Indonesian Twitter text authority classification for government in Bandung," in *2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2014, pp. 129–134. https://doi.org/10.1109/ICAICTA.2014.7005928
[24] I. O. Suzanti and A. Jauhari, "Comparison of Stemming and Similarity Algorithms in Indonesian Translated Al-Qur'an Text Search," *Jurnal Ilmiah KURSOR*, vol. 11, no. 2, p. 91, 2021. https://doi.org/10.21107/kursor.v11i2.280
[25] A. Khurana and O. P. Verma, "Optimal feature selection for imbalanced text classification," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 1, pp. 135–147, 2022. https://doi.org/10.1109/TAI.2022.3144651
[26] C. Magnolia, A. Nurhopipah, and B. A. Kusuma, "Penanganan imbalanced dataset untuk klasifikasi komentar program kampus merdeka pada aplikasi twitter," *Edu Komputika Journal*, vol. 9, no. 2, pp. 105–113, 2022. https://doi.org/10.15294/edukomputika.v9i2.61854
[27] G. I. Winata and M. L. Khodra, "Handling imbalanced dataset in multi-label text categorization using Bagging and Adaptive Boosting," in *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*, 2015, pp. 500–505. https://doi.org/10.1109/ICEEI.2015.7352552
[28] S. Liu, H. Tao, and S. Feng, "Text classification research based on bert model and bayesian network," in *2019 Chinese automation congress (CAC)*, 2019, pp. 5842–5846. https://doi.org/10.1109/CAC48633.2019.8996183
[29] R. A. Pane, M. S. Mubarok, N. S. Huda, and others, "A multi-lable classification on topics of quranic verses in english translation using multinomial naive bayes," in *2018 6th International Conference on Information and Communication Technology (ICoICT)*, 2018, pp. 481–484. https://doi.org/10.1109/ICoICT.2018.8528777
[30] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between multinomial and Bernoulli na\"\ive Bayes for text classification," in *2019 International conference on automation, computational and technology management (ICACTM)*, 2019, pp. 593–596. https://doi.org/10.1109/ICACTM.2019.8776800
[31] Y. Lin, Z. Yu, K. Yang, Z. Fan, and C. L. P. Chen, "Boosting Adaptive Weighted Broad Learning System for Multi-Label Learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 11, pp. 2204–2219, 2024. https://doi.org/10.1109/JAS.2024.124557
[32] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning and Data Mining*, Boston, MA: Springer US, 2017, pp. 260–260. https://doi.org/10.1007/978-1-4899-7687-1_50