JURNAL INFOTEL
VOL. 16, NO. 1, FEBRUARY 2024, PP. 71–81. DOI:10.20895/INFOTEL.V16I1.984

RESEARCH ARTICLE

Imbalance Dataset in Aspect-Based Sentiment Analysis on Game Genshin Impact Review

Prabowo Adi Perwira[1] and Nelly Indriani Widiastuti[2,*]

[1,2]Universitas Komputer Indonesia, Bandung 40132, Indonesia

*Corresponding email: nelly.indriani@email.unikom.ac.id

Received: July 2, 2023; Revised: September 2, 2023; Accepted: October 25, 2023.

Abstract: Sentiment analysis is a common method for determining the polarity of user reviews. However, there is a problem if some user reviews have more than one aspect with different polarities, so the reviews have more than one polarity. This problem also occurred in some user reviews of the game Genshin Impact. Apart from that, there are differences in the amount of data in the sentiment class, which causes imbalanced data. Therefore, this study will handle imbalanced data with random undersampling and random oversampling on aspect-based sentiment analysis of the Genshin Impact review with multinomial naïve Bayes so that the classification prediction does not ignore the minority class due to the dominance of the majority class. The classification process used the K-fold cross-validation (k = 10) validation method and the Laplace smoothing technique on multinomial naïve Bayes. As a result, the conclusion is that random oversampling had better accuracy than random undersampling in handling imbalanced data in aspect-based sentiment analysis of Genshin Impact game review in Indonesian with naïve Bayes multinomial, with the highest accuracy of 85.55 %.

Keywords: imbalance dataset, random undersampling, random oversampling, k-fold cross validation

---

1 Introduction

Reviews are opinions outlined by users or customers who comment on the product. This review is one way for prospective users to determine whether the product is good, for example, the reviews of a game on the Google Play Store [1]. Sentiment analysis for games has been used when game developers want to know what users think about their gaming experience. Reading the reviews of the platforms is the best way to decide on the development direction and further improvements of their games [2]. Analyzing the sentiment review also allows a company to evaluate its products [3]. So, using sentiment analysis, we can generate custom documents that describe user impressions. For game developers, this information is an advantage [4]. In this study, the case used was a review of the game Genshin Impact. This game has been rated by several companies [5,6]. The Genshin Impact game is one of the applications on the Google Play Store that has received some reviews. Based on observations of existing reviews, users comment a lot on, gameplay and storyline, accessibility, graphics and performance, and features.

Sentiment detection performed by machine learning automatically classifies large datasets. In a dataset, a review always has more than one aspect with different polarity [7]. Machine learning requires a balanced number of datasets between the classes. The problem occurred in the dataset used. The dataset contains some sentiments in which these reviews are not always the same as in other user reviews, which will cause an inequality in the amount of data distributed in each class. The amount of data differs between the data classes [8]. Therefore, handling is necessary to overcome this imbalance. Thus, determining the classification does not ignore the minority class because of the dominance of majority classes.

Several sentiment analysis studies have been conducted with mixed results. Research conducted by Rambe *et al.* used naïve Bayes to classify sentiment in beauty product reviews [9]. Even though it produces 90 % accuracy, it decreases as test data increases. It is probably because the model learning results from the training data cannot correctly recognize each class. Similar research was carried out by Salsabila *et al.*, who also used naïve Bayes [10]. This research detected aspects contained in sentiment but chose to ignore the equity of data in the training data because it is considered normal. This assumption is incorrect because the model needs to recognize factual data in a balanced way for better detection performance.

The research conducted by Johnson and Khoshgoftaar [11] handled imbalances in data from the center for medicare and medicaid services (CMS) using random undersampling (RUS), random oversampling (ROS), and random over-under sampling methods. The results obtained are that ROS is slightly superior to RUS and ROUS. ROS can maximize performance and efficiency by speeding up the training process up to four times. Research by Daulay and Asror [12] entitled "Sentiment Analysis on Google Play Store Reviews Using the naïve Bayes Method" has produced a relatively good accuracy of 78.9 % with the MNB method as a sentiment classification model for text-based review classification cases. This result was supported by Rahman and Akter [13]. They have conducted a study entitled "Topic Classification from Text Using Decision Tree, K-NN and Multinomial Naïve Bayes". They chose Multinomial naïve Bayes as the best classification model they used.

Based on this background, this research carried out Aspect-Based Sentiment classification using the naïve Bayes Multinomial method to determine the polarity of each aspect. Meanwhile, handling imbalanced data uses RUS and ROS. This research contributed as follows:

1. Classifying aspects in the Genshin Impact game review dataset using the naïve Bayes multinomial algorithm.
2. Shows the effect of balancing RUS and ROS data.

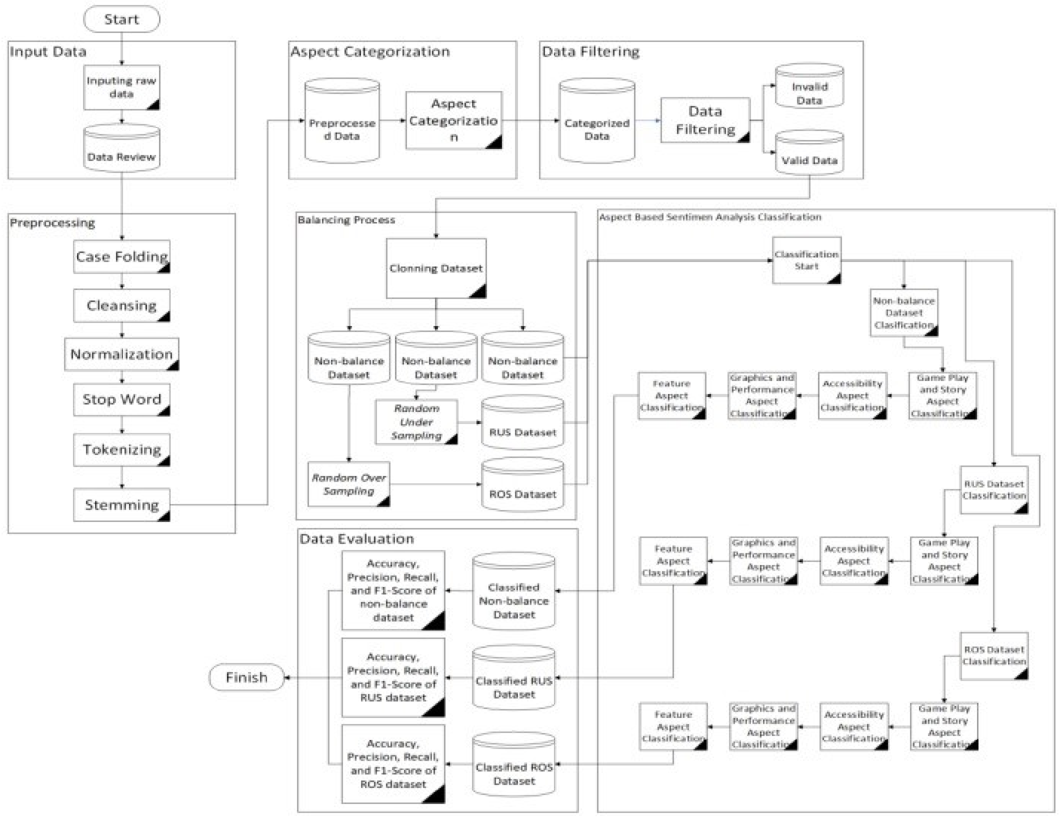

Figure 1: Proposed method.

3. Automatic labelling method for game review datasets. For other datasets, it can adjusted. This method can help with human labelling, which is time-consuming and expensive.

After the introduction, we organized the rest of the paper as follows. In section section 2, we described the proposed method as depicted in Figure 1. This section illustrated the entire process, from handling raw data to testing and evaluation results. Section section 3 explains the test results related to data balance. In this section, we describe the effect of ROS on the classification of aspects in the dataset. In section section 4, namely discussion, we discuss the test results and their relationship with related research. Finally, section section 5 is the section that summarizes the test results into conclusions and further studies based on the findings of this research.

2 Research Method

This section is an overview of the process that raw data goes through. Start from preprocess to evaluation, as shown in Figure 1. This research starts from entering raw data and becomes a data review. This data will enter the pre-processing to be specific, cleaning symbols, numbers, unnecessary words, and word affixes. Furthermore, the review data that has passed the pre-processing phase will be determined based on the terms contained and filtered to avoid reviews that do not carry any aspect from entering the classification process.

After filtering, proceed to clone this feasible data into three datasets, where the two datasets are processed for balancing while 1 dataset will be non-balanced. These two datasets will each be balanced using the RUS and ROS methods. After that, these datasets were classified using the naïve Bayes multinomial algorithm. The datasets to be classified are non-balanced, undersampled, and oversampled datasets.

The classification process was carried out sequentially, starting from the Non-Balance Dataset, then the Undersampled Dataset, and finally the Oversampled Dataset. The classification was carried out using the K-fold cross-validation, data validation method (k = 10), and the Laplace smoothing technique on naïve Bayes multinomial in each aspect of each dataset.

The classified dataset will enter the calculation process to obtain a confusion matrix table to calculate accuracy, precision, recall, and F1-scores, which will be compared and used as a conclusion drawn from the research results.

2.1 Input Dataset

The dataset was obtained by scraping [14] the user reviews of the Genshin Impact game on the Google Playstore platform using Python programming and was saved in a CSV format file. Table 1 is an example of raw data that was scrapped.

Table 1: Raw data

Username	**Content**	**Review Created Version**	**at**
V1R4_9107	Lebih bagus dibanding game lain walaupun kadang ngeselin karena nge-gacha tapi dapetnya senjata	1.6.0	2021-07-21 06:25:28
Maharani Silvia	Update trus kek ****, update size kecil jga gpp ini gede banget sampah	1.6.0	2021-07-21 06:15:54

The sentiment classes had divided into positive and negative. The aspects contained in the raw data include Gameplay and Story Aspects, Accessibility Aspects, Graphic and Performance Aspects, and Features Aspects.

2.2 Pre-processing

Pre-processing is a way to make raw data into ready data to be processed to get important information. This method can also improve the performance of the classification algorithms. The pre-processing flow in this study starts with Case Folding, Cleansing, Normalization, Stopword, tokenizing, and stemming.

1. Case folding: The process of converting all letters to lowercase [15]

2. Cleaning: The process of removing symbols up to numbers, and only the characters a-z are accepted [16].
3. Normalization: The process of changing words into complete/standard words [17].
4. Stopword: The process of removing words that have no meaning or words that appear frequently [15].
5. Tokenizing: The process of breaking words into broken words.
6. Stemming: The process of removing affixes using the ECS method [18].
7. If there is a word that is the same as the dictionary word accessibility word, then the review is labelled as an aspect of accessibility. Likewise, with other dictionaries, if there are words that are the same as other dictionaries.
8. Reviews can have one to four aspects, depending on the words contained in the review.
9. Reviews also cannot have aspects if the words contained in the review are not the same as the four word-list dictionaries.

2.3 Aspect Categorization

Categorization is a method for grouping sample data into aspect groups based on the words contained therein. These words were stored in a word list dictionary based on their respective aspects. The words in the dictionary determine what 'aspects' are included in the reviews. Table 2 shows an example of a word-list dictionary.

The rules of the categorization process are as follows:

1. Cleansing result data will be read word by word
2. The word read by the system checked whether it is the same as the word in the four word-list dictionaries.
3. The sequential process starts from the accessibility word list dictionary and then proceeds to gameplay and story, Graphics and Performance, and Features.
4. If there are no similar words in the accessibility word list dictionary, then the process will
5. Continue to the gameplay, story, and so on, until the feature word-list dictionary.

2.4 Data Filtering

Filtering is a process in which sample data that is categorized will be filtered based on the number of words from the preprocessing results or the aspects contained. The filtered data must have a minimum of two tokens to produce good data for the classification process. The filtering results will divided into two groups, namely Eligible Data and Ineligible Data. Eligible data is a sample with at least one aspect label to ensure that the data entering the classification process has aspects. Otherwise, it will be considered ineligible data.

2.5 Balancing data

The sample data used has a class imbalance between positive and negative. Therefore, a process is needed to balance the data. There are methods used in this study, namely RUS and ROS.

1. Random undersampling (RUS) is a data balancing technique where data from the majority class is discarded randomly to meet the balance of the sample data. This

technique can cause data shortages and also cause loss of potentially valuable information [19].

2. Random oversampling (ROS) is a data balancing technique where the data from the minority class is duplicated or increased randomly so that the data from each class can be equal or balanced. This technique can cause data overload and duplication of information that is the same as minority class data [20]

Table 2: Word list dictionary

Gameplay and story	Accessibility	Graphics and performance	Feature
gameplay	mouse	low	senjata
stori	keyboard	medium	artifak
story	joystick	high	artifact
storyline	controller	lag	artefak
aether	server	graphic	gacha
lumine	login	grafik	karakter
lore	connection	grafis	banner
konsep	koneksi	snapdragon	map
cerita	internet	primogem	mora
elemental	download		

2.6 Aspect Based Classification

The next stage is the Classification process. The data will be trained using the K-fold Cross Validation validation technique [21] [22], with the number of folds to be determined as 10-fold [23] [24], this allows for feature extraction of all data other than the data to have tested.

The classification process in this study uses the multinomial naïve Bayes method [12,25] with the Laplace smoothing technique [26] so that the probability results are not zero. The following is the naïve Bayes multinomial formula using the Laplace smoothing technique.

This classification was carried out on each aspect of the dataset. In this research, there are three datasets, namely non-balanced data, undersampling data, and oversampling data.

2.7 Evaluation

The confusion matrix is commonly used to measure the algorithm's performance. Classified data was evaluated to obtain a Confusion matrix table which consists of variables such as True Positive (TN), False Positive (FP), True Negative (TN), and False Negative (FN) to calculate accuracy in (1), precision in (2), recall in (3), and F1-scores in (4).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\text{Precission} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

3 Results

The total review data used in this test was 1,000 data, with 636 review data on Gameplay and Story aspects, which consisted of 495 positive review data and 141 negatives review data, 332 review data on the Accessibility aspect, which consisted of 97 positive review data and 235 negatives review data, 430 graphical and performance review data have contained 182 positive review data and 248 negatives review data, and 329 feature review data consisting of 119 positive review data and 210 negatives review data. From the results of tests that had carried out on review data both from non-balanced, undersampled, and oversampled data, the results and discussion of conclusions had obtained as follows.

The highest accuracy was found in oversampled data in 'Gameplay and Story Aspects', with an accuracy value of 85.55 %, precision of 85.72 %, recall of 85.18 %, and F1-score of 85.45 %. Likewise, in the Aspect of Accessibility, the data were oversampled, reaching 80.33 %, precision of 76.87 %, recall of 87.38 %, and F1-score of 81.78%. In the Graphic and Performance Aspects, the data have oversampled and gained 76.79 %, precision of 74.11 %, recall 83.11 %, and F1-score 78.35 %, and in 'the Features aspects, the data have oversampled and gained 77.14 %, precision 75.08 %, recall 81.88 %, and F1-score 78.34 %. For more details, see Figure 2.

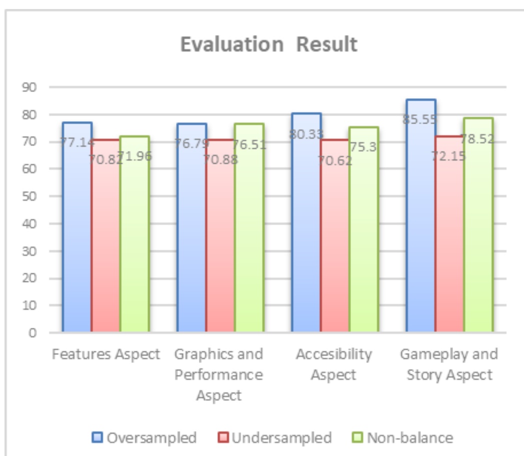

Figure 2: Accuracy.

The highest F1-score value was found in non-balanced data with Gameplay and Story Aspects, and then in Accessibility Aspects and Graphic and Performance Aspects, the oversampled data has the highest F1-score; meanwhile, the Features Aspect has the same F1-score on undersampled data and oversampled. For more details, see Figure 3.

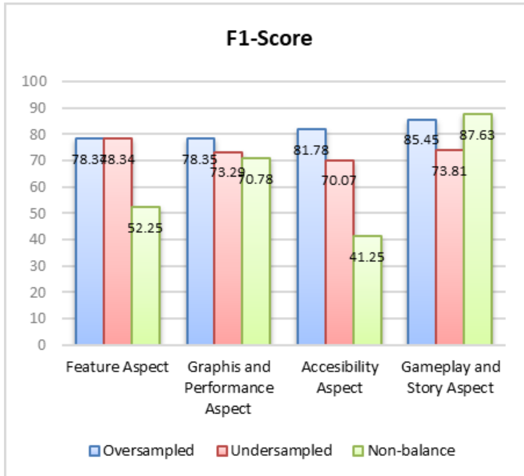

Figure 3: Level of F1-score.

4 Discussion

Based on the test results, it can conclude that the multinomial naïve Bayes used as a classification method in aspect-based sentiment analysis in the Genshin Impact Game review has a pretty good level of accuracy with an average resulting accuracy of more than 70 % as a classification model in classifying reviews based on text and also balancing RUS and ROS techniques are proven to be able to influence the naïve Bayes multinomial algorithm as a classification method which can increase the accuracy value of naïve Bayes multinomial when using ROS and can make the F1-score value more stable in both techniques. Then, for the accurate results that look different in each aspect, this is due to differences in the amount of data and the quality of the training data in each 'aspect'. The F1-score on 'Gameplay and Story aspects' is higher because the recall average was high. But overall, the undersampled and oversampled data have an average F1-score that is higher and more stable than non-balanced data.

The ROS data balancing technique has the effect of being able to increase the accuracy of the naïve Bayes Multinomial on the Genshin Impact Game review data. ROS makes information from the minority class reproduced so that when calculating the probability of $p(x)$, it produces a high enough value, and not only that, ROS has a higher level of precision and recall than RUS or non-balanced data.

Meanwhile, the RUS data balancing technique has the effect of making the accuracy value decrease. This is probably, due to the loss of information due to the undersampled balancing process during the probability calculation process for $p(x)$, and the naïve Bayes algorithm is very dependent on this information. The second possibility is the selection of random data due to the random function of RUS so that this allows good data or undersample quality also applies to ROS because of the random method; moreover, the multinomial naïve Bayes algorithm utilizes information from this training data to be able to predict its classification. this multinomial naïve Bayes algorithm. Even so, the level of accuracy produced by RUS is still quite good at 70 %.

ROS is proven to have a better effect on accuracy than RUS on Genshin impact game review data using the naïve Bayes multinomial classification method. RUS data also have lower results than Non-Balance data. This method was also supported by the research of Sabani *et al.* [27], who used K-fold as their validation method. The classification method uses the Supervised learning algorithm, namely SVM and RUS. With the imbalance data handling method, the result is that RUS data has lower accuracy than unbalanced data but more stable precision, recall, and F1-score.

5 Conclusion

The conclusion was drawn that naïve Bayes multinomial can be used in classifying aspect-based sentiments in Genshin Impact Game reviews based on text in Indonesian. The ROS and RUS data balancing techniques are proven to influence aspect-based sentiment analysis in Indonesian language Genshin Impact game text reviews, where ROS data has a better effect in terms of accuracy than RUS data or non-balanced data, with the highest accuracy rate reaching 85.55 % on gameplay and story aspects.

Future research can focus on improvements in data selection during undersampling and oversampling. The method may combine with data selection that hopefully could have a good impact on performance. We can try data selection methods such as random forest or information gain.

References

[1] S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning," *Expert Systems with Applications*, vol. 181, p. 115111, Nov. 2021.

[2] A. Polhemus, S. Simblett, E. Dawe Lane, B. Elliott, S. Jilka, E. Negbenose, P. Burke, J. Weyer, J. Novak, M. F. Dockendorf, G. Temesi, and T. Wykes, "Experiences of health tracking in mobile apps for multiple sclerosis: A qualitative content analysis of user reviews," *Multiple Sclerosis and Related Disorders*, vol. 69, p. 104435, Jan. 2023.

[3] P. Mukherjee, Y. Badr, S. Doppalapudi, S. M. Srinivasan, R. S. Sangwan, and R. Sharma, "Effect of negation in sentences on sentiment analysis and polarity detection," *Procedia Computer Science*, vol. 185, pp. 370–379, 2021.

[4] D. Blazquez and J. Domenech, "Big Data sources and methods for social and economic analyses," *Technological Forecasting and Social Change*, vol. 130, pp. 99–113, May 2018.

[5] "The phenomenon of genshin impact - https://weplayholding.com/blog/the-phenomenon-of-genshin-impact/."

[6] "Genshin impact - lantern rite - aplikasi di google play - https://play.google.com/store/apps/details?id=com.miHoYo.GenshinImpact&hl=id."

[7] J. Wei, J. Liao, Z. Yang, S. Wang, and Q. Zhao, "Bilstm with multi-polarity orthogonal attention for implicit sentiment analysis," *Neurocomputing*, vol. 383, pp. 165–173, Mar. 2020.

[8] A. A. Arifiyanti and E. D. Wahyuni, "Smote: metode penyeimbang kelas pada klasifikasi data mining," *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, vol. 15, pp. 34–39, Feb. 2020.

[9] T. S. Rambe, M. N. S. Hasibuan, and M. H. Dar, "Sentiment analysis of beauty product applications using the naïve bayes method," *SinkrOn*, vol. 8, pp. 980–989, Apr. 2023.

[10] S. Salsabila, S. M. P. Tyas, Y. Romadhona, and D. Purwitasari, "Aspect-based sentiment and correlation-based emotion detection on tweets for understanding public opinion of covid-19," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, pp. 84–94, Apr. 2023.

[11] J. M. Johnson and T. M. Khoshgoftaar, "Deep learning and data sampling with imbalanced big data," in *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, (Los Angeles, CA, USA), pp. 175–183, IEEE, July 2019.

[12] E. D. P. Daulay and I. Asror, *Sentimen analisis pada ulasan google play store menggunakan metode naïve bayes*. Universitas Telkom, S1 Informatika, Oct. 2020.

[13] M. A. Rahman and Y. A. Akter, "Topic classification from text using decision tree, k-nn and multinomial naïve bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, (Dhaka, Bangladesh), pp. 1–4, IEEE, May 2019.

[14] A. Josi, L. A. Abdillah, and Suryayusra, "Penerapan teknik web scraping pada mesin pencari artikel ilmiah," Oct. 2014. arXiv:1410.5777 [cs].

[15] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving text preprocessing for student complaint document classification using sastrawi," *IOP Conference Series: Materials Science and Engineering*, vol. 874, p. 012017, June 2020.

[16] A. H. Yunas, *Penerapan metode SVM untuk klasifikasi tweet transaksi E-commerce di twitter*. Universitas Islam Negeri Sultan Syarif Kasim Riau, 2018.

[17] F. F. Rachman and S. Pramana, "Analisis sentimen pro dan kontra masyarakat indonesia tentang vaksin covid-19 pada media sosial twitter," *Indonesian of Health Information Management Journal (INOHIM)*, vol. 8, pp. 100–109, Dec. 2020.

[18] A. Arifin, P. Mahendra, and H. T. Ciptaningtyas, "Enhanced confix stripping stemmer and ants algorithm for classifying news document in indonesian language," 2009.

[19] B. Liu and G. Tsoumakas, "Dealing with class imbalance in classifier chains via random undersampling," *Knowledge-Based Systems*, vol. 192, p. 105292, Mar. 2020.

[20] M. Hayaty, S. Muthmainah, and S. M. Ghufran, "Random and synthetic over-sampling approach to resolve data imbalance in classification," *International Journal of Artificial Intelligence Research*, vol. 4, p. 86, Jan. 2021.

[21] S. Geisser, "The predictive sample reuse method with applications," *Journal of the American Statistical Association*, vol. 70, pp. 320–328, June 1975.

[22] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, pp. 111–133, Jan. 1974.

[23] Y. Jung, "Multiple predicting *K* -fold cross-validation for model selection," *Journal of Nonparametric Statistics*, vol. 30, pp. 197–215, Jan. 2018.

[24] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, "Peringkasan dan support vector machine pada klasifikasi dokumen," *JURNAL INFOTEL*, vol. 9, p. 416, Nov. 2017.

[25] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between multinomial and bernoulli naïve bayes for text classification," in *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, (London, United Kingdom), pp. 593–596, IEEE, Apr. 2019.

[26] V. Cherian, "Heart disease prediction using naive bayes algorithm and laplace smoothing technique," 2017.

[27] A. F. Sabani, A. Adiwijaya, and W. Astuti, "Analisis sentimen review film pada website rotten tomatoes menggunakan metode svm dengan mengimplementasikan fitur extraction word2vec," *eProceedings of Engineering*, vol. 9, June 2022.