Published online on the journal's webpage: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

Journal of Dinda

Data Science, Information Technology, and Data Analytics

Vol. 4 No. 2 (2024) 43 - 49 E-ISSN: 2809-8064

Sentiment Analysis of Handling "Klitih" in Yogyakarta Using Naïve Bayes

Windha MP Dhuhita[1*], Ilham Ferry Pratama[2], Bayu Setiaji[3]

[1*,2,3]Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

[1*]windha@amikom.ac,id, [2]ilham.yu@students.amikom.ac.id, [3]bayu@amikom.ac.id

Abstract

Activities that lead to crimes called "klitih" often occur and disturb the community. The community's response to the handling carried out by the regional government also varied. The public expressed this response using various types of social media, one of which was Twitter. This research analyzes sentiment, or responses given by the public by utilizing the social media Twitter to collect data. Data in the form of tweets that have been taken will go through text processing. After that, the text will be weighted using two methods as a comparison, namely tf-idf and count vector. Then the data will be divided into training data and test data to proceed to the classification stage. Classification is carried out using the Naïve Bayes algorithm. To evaluate the results of Naïve Bayes classification, researchers used the Confusion Matrix, by comparing weighting methods and dividing the training data and test data into several different ratios, to find out the scenario that produces the best level of accuracy. The sentiment obtained was dominated by negative sentiment at 75.8%, while positive sentiment was 24.2%. By using existing data, it was found that weighting with count vector had an accuracy rate of 82%. Meanwhile, weighting using TF-IDF obtained an accuracy of 80%.

Keywords: *Sentiment, Klitih, Naïve Bayes, Weighting, Classification.*

1. Introduction

Klitih, which in Indonesian means doing activities outside the home and wasting time, has become a bad act. With many incidents such as the use of sharp weapons and abuse which are often reported because of klitih. The public is worried about the existence of klitih activities that can be targeted randomly. The government's handling efforts received a response from the public. By utilizing social networking media such as Twitter, people provide responses (tweets) to policies taken by the government. The response given can be in the form of praise, criticism or neither. Public opinions regarding the efforts taken will be collected, then summarized as information. Extracting this information can be done using sentiment analysis. Sentiment analysis or often also called opinion mining is a branch of data mining that focuses on processing opinions. The results of sentiment analysis can be known after going through the pre-processing or text preprocessing stages and then proceeding to the classification process. Classification methods such as the Naive Bayes Classifier are often used in sentiment classification. In this research, sentiment analysis will be carried out on the community's assessment regarding the completion of "klitih" in Yogyakarta to determine the community's response into positive or negative categories. So, it can be used as an assessment of the government's performance in handling conflict issues.

Previous research that raised the theme of sentiment analysis, Ike Verawati produced a high level of accuracy using Naïve Bayes with a multinomial model. Tests carried out with three different dataset ratio scenarios obtained different accuracy values. The accuracy levels obtained were 80%, 80% and 85% [1]. Another research, conducted by Dianati Duei Putri et al, who conducted sentiment analysis on DPR performance using the Naïve Bayes algorithm, obtained an accuracy of 80% with a dataset division of 80:20 [2]. In the two studies above, the data processing stage used TF-IDF, but each had a different number of classes. In the second study, there were 3 classes, namely positive, negative and neutral.

There are many weighting methods that can be used to process data before classification, including TF-IDF, TF-RF and Count Vector. In research conducted by

Received: 20-06-2024 | Accepted: 04-07-2024 | Published: 01-08-2024

Muhammad Zaynurroyhan et al, namely comparing the TF-IDF method with Count Vectorization which was applied to the K-Nearest Neighbor classification algorithm. From this research, it was concluded that the model using Count Vectorization obtained an accuracy of 75%, while TF-IDF produced a lower accuracy, namely 64% [3]. In contrast to the research above, Kristien Margi Suryaningrum compared TF-IDF with Count Vectorization which was applied to the Support Vector Machine algorithm to classify hate speech. Feature extraction using TF-IDF achieves higher performance compared to Count Vectorization in terms of accuracy, precision, recall and F1-score. The SVM method that uses TF-IDF feature extraction gets an accuracy of 88.77%, while Count Vectorization gets an accuracy of 87.14% [4]

Based on previous research, researchers will test several models, to find a model that has the best performance in conducting sentiment analysis on public assessments regarding the completion of "klitih" in Yogyakarta. The method used is to apply the Naive Bayes Classifier to get the most optimal classification results using a Confusion Matrix [5][6], such as accuracy, precision, recall. The Naive Bayes algorithm was chosen to analyze sentiment regarding the resolution of "klitih" in Yogyakarta because of its proven ability in efficient and effective text classification, especially in dealing with relatively small but complex datasets such as evaluations of local social issues [7]. To increase the accuracy of Naïve Bayes, feature extraction will compare the weighting method using TF-IDF with the Count Vectorization method [8].

2. Research Methods

The research stages are used as a guide in conducting research, starting from the initial stage to the final stage. The procedures and sequence of research steps that will be carried out in this research can be seen in Figure 1 below:

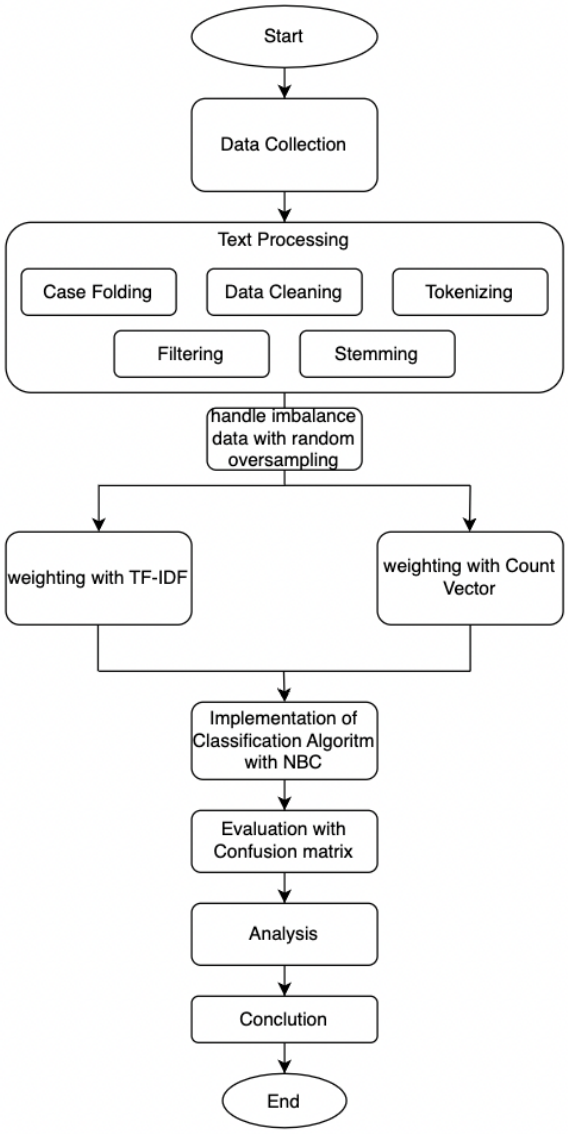


Figure 1. Research Flow

2.1 Data Collection

The data collected are public comments from social media Twitter. This data collection uses a Twitter crawler using the Python programming language. Some of the keywords used to search for tweets are 'klitih action', 'klitih handling', 'klitih resolution' and 'klitih government'. The parameter used in the search is the language used, Indonesian. After the search was carried out and the data was saved into several CSV files, then the data in the form of tweets were combined into one file and had a total of 1326 data. This data has a time range from 2012 to 2023.

2.2 Text Preprocessing

Text Preprocessing is a stage that is carried out after all the text has been collected and is a stage that must be passed before applying the classification algorithm.

a. Case Folding

At the case folding stage, the text will be uniform in either all capital letters or all lower-case letters. In this study, all letters were made into lowercase. This process of changing letters uses the built-in function in Python, namely. lower(). Examples of using case folding can be seen in the following table:

Table 1. Case Folding

Before Case Folding	After Case Folding
Kebijakan jam malam di Jogja ini dilaksanakan untuk menangkal klitih. Namun apakah kebijakan ini bisa jadi solusi terbaik untuk mengatasi klitih? #Kilas https://t.co/wg9F4m5epC	kebijakan jam malam di jogja ini dilaksanakan untuk menangkal klitih. Namun apakah kebijakan ini bisa jadi solusi terbaik untuk mengatasi klitih? #kilas https://t.co/wg9f4m5epc

b. Data Cleaning

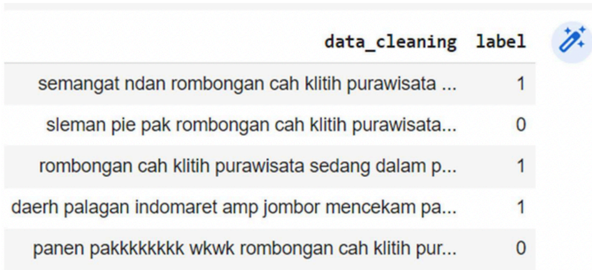
Next, characters other than letters in the text are removed, such as numbers, dots or other characters. Data cleaning is carried out by utilizing the regular expression (regex) feature in Python to look for certain patterns, which are then replaced with blank characters or spaces. By carrying out the data cleaning process, the tweets that have been collected will only be text without any other characters. An example of implementing data cleaning can be seen in the following table:

Table 2. Data Cleaning

Before Cleaning	After Cleaning
kebijakan jam malam di jogja ini dilaksanakan untuk menangkal klitih. Namun apakah kebijakan ini bisa jadi solusi terbaik untuk mengatasi klitih? #kilas https://t.co/wg9f4m5epc	kebijakan jam malam jogja ini dilaksanakan untuk menangkal klitih namun apakah kebijakan ini bisa jadi solusi terbaik untuk mengatasi klitih

c. Labeling

The labeling process is used to provide labels in the form of positive and negative to text data that has been cleaned. This stage is carried out automatically using the help of the transformer's library. Transformers can only do English labeling, therefore data in the form of Indonesian text needs to be translated first into English, and after labeling the text needs to be converted again into Indonesian. The deep-translator library can be used to translate words. The labels provided are still in the form of positive and negative text, so that they can be processed by the classification algorithm they need to be converted into numbers or integer data types. Positive labels will be represented as the number 1 and negative labels will be represented as the number 0. Figure 3 shows tweet data that has been labeled.



data_cleaning	label
semangat ndan rombongan cah klitih purawisata ...	1
sleman pie pak rombongan cah klitih purawisata...	0
rombongan cah klitih purawisata sedang dalam p...	1
daerh palagan indomaret amp jombor mencekam pa...	1
panen pakkkkkkkk wkwk rombongan cah klitih pur...	0

Figure 3. Data Labeling

d. Tokenizing

Data that has been cleaned will be cut or chopped into several parts in the tokenizing process. An example of the application of tokenizing can be seen in table 3 below:

Tabel 3. Data Tokenizing

Before Tokenizing	After Tokenizing
kebijakan jam malam jogja ini dilaksanakan untuk menangkal klitih namun apakah kebijakan ini bisa jadi solusi terbaik untuk mengatasi klitih	'kebijakan', 'jam', 'malam', 'jogja', 'ini', 'dilaksanakan', 'untuk', 'menangkal', 'klitih', 'namun', 'apakah', 'kebijakan', 'ini', 'bisa', 'jadi', 'solusi', 'terbaik', 'untuk', 'mengatasi', 'klitih'

e. Filtering

Filtering or often known as stopword removal is the stage of removing words that are considered unimportant. Connecting words like "dan", "atau" and "di", or other words that have no meaning will be deleted. The reason for removing stop words is that their use is too common, so users can focus on other more important words [9]. Examples of filtering implementation can be seen in the following table 4.1:

Table 4.1 Data Filtering in Indonesian Language as we used

Before Filtering	After Filtering
'kebijakan', 'jam', 'malam', 'jogja', 'ini', 'dilaksanakan', 'untuk', 'menangkal', 'klitih', 'namun', 'apakah', 'kebijakan', 'ini', 'bisa', 'jadi', 'solusi', 'terbaik', 'untuk', 'mengatasi', 'klitih'	'kebijakan', 'jam', 'malam', 'jogja', 'dilaksanakan', 'menangkal', 'klitih', 'kebijakan', 'solusi', 'terbaik', 'mengatasi', 'klitih'

Table 4.2 Data Filtering That We Translate in English Language

Before Filtering	After Filtering
'policy', 'curfew', 'jogja', 'this', 'held', 'for', 'ward off', 'klitih', 'but', 'is', 'policy', 'this', 'can', 'so', 'solution', 'best', 'for', 'overcome', 'klitih'	'policy', 'curfew', 'jogja', 'this', 'held', 'for', 'ward off', 'klitih', 'but', 'is', 'policy', 'this', 'can', 'so', 'solution', 'best', 'for', 'overcome', 'klitih'

f. Stemming

The stemming process is carried out to remove affixes from a word, so that the word used becomes the base word. The Nazief and Adriani algorithm in the literary library is used to carry out the stemming process[10]. An example of the application of stemming can be seen in the following table 5.1:

Table 5.1 Data Stemming in Indonesian Language as we used

Before Stemming	After Stemming
'kebijakan', 'jam', 'malam', 'jogja', 'dilaksanakan', 'menangkal', 'klitih', 'kebijakan', 'solusi', 'terbaik', 'mengatasi', 'klitih'	'bijak', 'jam', 'malam', 'jogja', 'laksana', 'tangkal', 'klitih', 'bijak', 'solusi', 'baik', 'atas', 'klitih'

Table 5.2 Data Stemming That We Translate in English Language

Before Stemming	After Stemming
'policy', 'curfew', 'jogja', 'held', 'ward off', 'klitih', 'policy', 'can', 'solution', 'best', 'overcome', 'klitih'	'policy', 'curfew', 'jogja', 'held', 'ward off', 'klitih', 'policy', 'can', 'solution', 'best', 'overcome', 'klitih'

g. Word Weighting

Word weighting uses TF-IDF and count vector to represent data in the form of text into numbers. In this research, word weighting was carried out using the Sciket-Learn library, TF-IDF then displays the numbers sequentially document index, word index, and word weight [11].

h. Algorithm Implementation

At this stage the classification algorithm will be implemented on data that has been given weights. However, before that, oversampling of the data was carried out to handle unbalanced data after labeling [12][13]. The algorithm that will be used to carry out classification is Naïve Bayes. The Naïve Bayes algorithm is used to classify text whether it includes positive or negative labels. The text will be divided into training data and test data. This research compares weighting methods as well divide the training and test data into several different ratios to find out which scenario produces the best level of evaluation. The scenario for dividing training data and test data can be seen in the following ratio table 6.

Table 6 Data Sharing Ratio

	Training Data	Testing Data
1	85%	15%
2	75%	25%
3	65%	35%

3. Results and Discussion

The data collection process was carried out by scraping the Twitter search page and obtaining data for 1027 tweets. The data that has been taken from this source is presented in Excel format, which will then go through a preprocessing stage before being weighted and finally classified [14]. To get more detailed information on the dataset, exploratory data analysis (EDA) was carried out. From the data obtained, it is known that 75.8% of people responded negatively to the completion of klitih in Yogyakarta. Meanwhile, 24.2% gave a positive response. This shows that the public is still dissatisfied with the existing settlement of Klitih in Yogyakarta.

The sentence labels that have been produced need to be further validated to obtain groups of labels and sentences that are appropriate to the context. This can happen because sentence labeling is done automatically. Therefore, experts are needed so that the sentences and labels given are appropriate.

Apart from knowing people's sentiments regarding the completion of klitih, from the dataset in this research you can also find out words that often appear. From the visualization results, the words that often appear are 'klitih', 'solution', 'government' with the word being shown the largest compared to the others, the information that can be obtained is the role of the government in resolving klitih cases.

Apart from visualizing using word clouds, this research carried out information searches to find out when people tweet. Each tweet is grouped by year to make it easier to analyze the dataset. Based on data collected from 2012 to 2023 regarding the completion of klitih in Yogyakarta, it is known that every year problems regarding klitih still occur.

After cleaning the data, it was found that the dataset used contained unbalanced positive and negative category data. For this reason, handling needs to be done because it will affect the classification model used [15]. One method used to overcome the problem of unbalanced data is using the random oversampling method. This method was chosen because it suits the number of data sets that are owned. Table 7 below shows the results of the oversampling process.

Tabel 7. Oversampling

Label	**Before Oversampling**	**After Oversampling**
Positive	316	964
Negative	964	964

Sentiment analysis on the topic of conflict resolution that occurred in Yogyakarta was implemented using Naive Bayes as the classification algorithm [16], [17]. The process of creating a sentiment model can be done after preprocessing and word weighting. In this research, there are several parameters used to obtain optimal results as in table 8 below:

Tabel 8. Parameter

Category	Value
Ratio of Training and Testing Data	[0,85 : 0.15, 0,75 : 0.25, 0,65 : 0.35]
Word Weighting	TF-IDF, Count Vector
Clasifier Algoritm	Naive Bayes Multinomial
Evaluation Parameter	Accuracy, precision, recall

Testing was carried out in two scenarios to obtain the best evaluation parameter results. The first scenario, using weighting of the word count vector with changes in the test data ratio value, with a test data amount of 15% of the dataset, an accuracy value of 0.82 or 82% is obtained. A more detailed explanation is displayed in the form of a classification report as seen in Figure 4

```
              precision    recall  f1-score   support

           0       0.84      0.77      0.80       139
           1       0.79      0.86      0.83       144

    accuracy                           0.82       283
   macro avg       0.82      0.82      0.82       283
weighted avg       0.82      0.82      0.82       283
```

Figure 4. Results of Scenario 1 with 15% test data

Furthermore, increasing the comparison ratio between test data to 25%, from the results obtained in this experiment the accuracy value was 82%. A more detailed explanation regarding the results of this classification can be seen in Figure 5:

```
              precision    recall  f1-score   support

           0       0.85      0.77      0.81       233
           1       0.80      0.87      0.83       238

    accuracy                           0.82       471
   macro avg       0.82      0.82      0.82       471
weighted avg       0.82      0.82      0.82       471
```

Figure 5. Results of Scenario 1 with 25% test data

The final test in this scenario was to increase the training data value to 35%. From the results obtained in this test, the accuracy value was 81%, a decrease of 2% from the previous two experiments. The results of the classification report can be seen more clearly in Figure 6:

	precision	recall	f1-score	support
0	0.82	0.78	0.80	318
1	0.80	0.84	0.82	341
accuracy			0.81	659
macro avg	0.81	0.81	0.81	659
weighted avg	0.81	0.81	0.81	659

Figure 6. Results of Scenario 1 with 35% test data

The second scenario in this research is by replacing the word weighting method using TF-IDF, the process of transforming text data into numbers is carried out before entering the sentiment model creation. The first test in the second scenario is by setting the test data ratio to 15% of the total dataset. The results that have been carried out show an accuracy of 80%. Figure 7 is a classification report for this experiment.

	precision	recall	f1-score	support
0	0.83	0.73	0.78	139
1	0.77	0.85	0.81	144
accuracy			0.80	283
macro avg	0.80	0.79	0.79	283
weighted avg	0.80	0.80	0.79	283

Figure 7. Result of Scenario 2 with 15% test data

The next step in the second scenario is to increase the test data ratio to 25%, the results obtained in this test are the same as the previous test of 80% accuracy as shown in Figure 8.

	precision	recall	f1-score	support
0	0.81	0.77	0.79	233
1	0.79	0.83	0.81	238
accuracy			0.80	471
macro avg	0.80	0.80	0.80	471
weighted avg	0.80	0.80	0.80	471

Figure 8. Results of Scenario 2 with 25% test data

The final test in the second scenario was to increase the test data value to 35%, in this test the accuracy results obtained were 79% which can be seen in Figure 9.

	precision	recall	f1-score	support
0	0.80	0.75	0.78	318
1	0.78	0.82	0.80	341
accuracy			0.79	659
macro avg	0.79	0.79	0.79	659
weighted avg	0.79	0.79	0.79	659

Figure 9. Results of Scenario 2 with 35% test data

The results of the two scenarios were then visualized into a bar chart to make comparisons easier as can be seen in Figure 10. From the bar chart it can be seen that the first scenario using count vector weighting got a higher accuracy value compared to weighting using TF-IDF. The highest accuracy and recall are in the scenario with 25% training data, accuracy value 82%, recall 87%, precision 80%. The second scenario has the highest accuracy value of 80% with a recall value of 85% and precision of 77%, in the scenario with 15% training data.

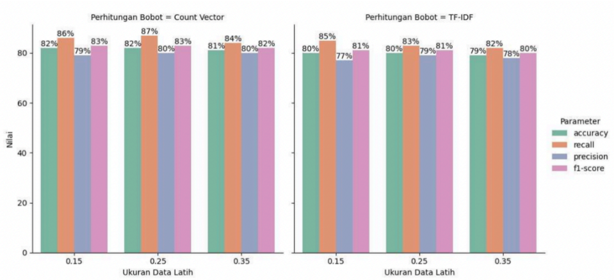


Figure 10. Comparative visualization of scenarios 1 and 2

4. Conclusion

Based on the results of the research that has been carried out, it can be concluded as follows: Sentiment analysis obtained with the topic of community assessment related to completion of klitih in Yogyakarta as much as 75.8% of the total data had a negative sentiment, while 24.2% had a positive response. classification using the Naive Bayes algorithm has value optimally using count vector weighting with the highest accuracy, recall and precision values found in the scenario test data sharing 25%, with accuracy values 82%, recall 87% and precision 80%. The data weighting method and changes in the test data ratio influence performance of the Naïve Bayes classification algorithm.

References

[1] I. Verawati and B. S. Audit, “Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1411, Jul. 2022.

[2] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022.

[3] Zaynurroyhan, Muhammad, and Agung Pambudi. "Perbandingan TF-IDF dengan Count Vectorization Dalam Content-Based Filtering Rekomendasi Mobil Listrik." *Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika,* vol. 15, no.1, pp. 8-15, 2023.

[4] K. M. Suryaningrum, "Comparison of the TF-IDF Method with the Count Vectorizer to Classify Hate Speech," *Engineering, MAthematics and Computer Science (EMACS) Journal*, vol. 5, no. 2, pp. 79–83, May 2023.

[5] Normawati, Dwi, and Surya Allit Prayogi. "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter." *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697-711, 2021.

[6] Alamsyah, Rizki, et al. "Sentiment Analysis Destinasi Wisata Berdasarkan Opini Masyarakat Menggunakan Naive Bayes." *Journal of Dinda: Data Science, Information Technology, and Data Analytics,* vol. 2, no. 2, pp. 64-74, 2022.

[7] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022.

[8] Aninditya, Annisa, Muhammad Azani Hasibuan, and Edi Sutoyo. "Text mining approach using TF-IDF and naive Bayes for classification of exam questions based on cognitive level of bloom's taxonomy." *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*. IEEE, 2019.

[9] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, vol. 2, no. 2, pp. 1–9, 2021.

[10] Dipti P. Rana and Rupa G. Mehta, *Data Preprocessing, Active Learning, and Cost Perceptive Approaches for Resolving Data Imbalance*. United States of America: IGI Global, 2021.

[11] Suryani, Luluk, and Kasmi Edy. "PENGEMBANGAN APLIKASI" LOST & FOUND" BERBASIS ANDROID DENGAN MENGGUNAKAN METODE TERM FREQUENCY–INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY." *Electro Luceat,* vol. 6, no. 2 pp. 190-204, 2020.

[12] H. Ali, M. Najib, M. Salleh, R. Saedudin, K. Hussain, and M. Faheem Mushtaq, "Imbalance class problems in data mining: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1560–1571, 2019.

[13] Ridwan, Achmad, and Sandi Mundakir. "Penerapan Teknik Bagging Pada Algoritma Naive Bayes Dan Algoritma C4. 5 Untuk Mengatasi Ketidakseimbangan Kelas." *JURNAL ILMU KOMPUTER DAN MATEMATIKA,* vol. 1, no. 1, pp. 63-70, 2020.

[14] I. N. Husada, E. H. Fernando, H. Sagala, A. E. Budiman, and H. Toba, "Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 3, Jan. 2020.

[15] Mutmainah, Siti. "Penanganan imbalance data pada klasifikasi kemungkinan penyakit stroke." *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 1, no.1, pp. 10-16, 2021.

[16] Sari, Dea Ropija. "Klasifikasi Sentimen Masyarakat di Twitter Terhadap Ancaman Resesi Ekonomi 2023 dengan Metode Naïve Bayes Classifier." *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 4, pp. 577-585, 2023.

[17] Prakoso, Cahyo, and Arief Hermawan. "Perbandingan Model Machine Learning dalam Analisis Sentimen Ulasan Pengunjung Keraton Yogyakarta pada Google Maps." *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1292-1302, 2023.