

ENHANCING SENTIMENT ANALYSIS WITH CHATBOTS: A COMPARATIVE STUDY OF TEXT PRE-PROCESSING

Indri Tri Julianto^{*1}, Dede Kurniadi^{*2}, Benedicto B. Balilo Jr³

^{1,2}Department of Computer Science, Institut Teknologi Garut, Indonesia

³CS/IT Department, College of Science, Bicol University, Legazpi City, Philippines

Email: ¹indritrijulianto@itg.ac.id, ²dede.kurniadi@itg.ac.id, ³bjbbalilo@bicol-u.edu.ph

(Article received: October 15, 2023; Revision: November 27, 2023; published: Desember 23, 2023)

Abstract

Text pre-processing plays a crucial role in the Sentiment Analysis process. Machine Learning models like Chat GPT-3.5 by OpenAI and Google Bard serve as alternative methods for text pre-processing. This study aims to evaluate the capabilities of both Chatbots in the text pre-processing stage while assessing their performance using a dataset obtained by crawling from source X. The study involves a comparison of Chat GPT-3.5 and Google Bard using Decision Tree and Naïve Bayes algorithms. The validation process employs K-Fold Cross Validation with a K value of 10. Additionally, three sampling methods, namely Linear, Shuffled, and Stratified Sampling, are utilized. The findings reveal that Chat GPT-3.5 performs best when using the Decision Tree algorithm with a K-Fold Cross value of 10, and employing Stratified Sampling, achieving an Accuracy of 90.68%, Precision of 90.63%, and Recall of 100%. On the other hand, Google Bard's optimal performance is achieved with the Decision Tree algorithm, a K-Fold Cross value of 10, and Shuffled Sampling, resulting in an Accuracy of 74.00%, Precision of 72.73%, and Recall of 98.77%. The study concludes that Chat GPT-3.5 and Google Bard are viable alternatives for text pre-processing in Sentiment Analysis. Performance measurements indicate that Chat GPT-3.5 outperforms Google Bard, achieving an Accuracy of 90.68%, Precision of 90.63%, and Recall of 100%. These results were validated by comparing them to human annotations, which achieved an accuracy score of 85.20%, Precision of 85.71%, and Recall of 99.03% when using the Decision Tree algorithm with a K-Fold Cross value of 10 and employing Stratified Sampling. This suggests that Chat GPT-3.5's text pre-processing performance is on par with human annotations.

Keywords: *altenative, Chat GPT-3.5 Open AI, google bard, sentiment analysis, text pre-processing.*

1. INTRODUCTION

Text Pre-Processing is an important stage and step in Sentiment Analysis, as most text elements such as characters, words, and sentences go through the entire Text Pre-Processing process [1]-[2]. This step involves transforming raw data into a more understandable format [3]. Sentiment Analysis is a process of analyzing a specific event, issue, or problem to determine the responses or opinions regarding that event [3]-[5]. Sentiment Analysis functions in developing systems to analyze, identify, and express opinions or sentiments, including Positive, Negative, or Neutral sentiments [6]. Sentiment Analysis is commonly performed using Natural Language Processing (NLP), where the main goal of NLP is to train machines to understand the meaning of human language and provide appropriate feedback [7].

The Text Pre-Processing stage is commonly performed using tools such as Rapidminer and Python. With the advancement of technology, Chatbots have emerged, which are Artificial Intelligence (AI) products capable of engaging in

automated conversations with humans using natural language, without limitations of space and time [8].

Chat Generative Pre-trained Transformer (Chat GPT-3.5 Open AI) Open AI is a renowned Chatbot with the ability to interact well with humans and provide answers to given questions. It gained popularity in November 2022 as it combines Reinforcement Learning Algorithm with input from humans, leveraging over 150 billion parameters [9]. Chat GPT-3.5 Open AI has now evolved to its fourth version, known as Chat GPT-3.5 Open AI-4, functioning as a personal assistant to humans and capable of performing various tasks and answering inquiries. These include assisting in article creation, generating song lyrics, translating languages, and even creating automated chats in chat apps, among many other capabilities [10].

Google Bard serves as a competitor to Chat GPT-3.5 Open AI, employing the Language Model for Dialogue Applications (LaMDA) [11]. As the name suggests, this Chatbot is a product of Google, aiming to combine their extensive knowledge of the world with the power, intelligence, and creativity of their large-scale language model. Google Bard assists

in information retrieval and can even explain new discoveries from the James Webb Space Telescope to a 9-year-old child [12].

Several previous studies have addressed Text Pre-Processing, emphasizing its importance in Sentiment Analysis [13]–[15]. Further research has shown that techniques such as stopword removal, stemming, and feature selection methods can improve performance by 20.4% [16]. Another study focused on Text Pre-Processing, specifically cleaning, case folding, and stemming techniques without stopword removal, resulting in a 94.24% accuracy increase [17]. Research on Sentiment Analysis utilizing the Decision Tree algorithm has demonstrated good accuracy, such as achieving 99.01% accuracy when applied to an e-commerce application rating dataset [18], and achieving 84.78% accuracy when applied to a Twitter dataset [19]. There are also studies that employ the Naïve Bayes algorithm for Sentiment Analysis, achieving accuracies of 76.39% on a Twitter dataset [20] and 73.57% [10].

Previous research has predominantly utilized Rapidminer applications, as seen in studies [3], [4], [6], [21], [22], while others have used Python, as seen in studies [5], [23]–[26]. The average accuracy value resulting from these three studies using Python is 73.33%. To identify the differences between current research and previous studies, a Research Gap is presented in the form of a table, as shown in Table 1.

Table 1. Research Gap

No	Research	Text Pre-Processing Technique
1	[3]	Rapidminer
2	[4]	Rapidminer
3	[6]	Rapidminer
4	[21]	Rapidminer
5	[22]	Rapidminer
6	[5]	Python
7	[23]	Python
8	[24]	Python
9	[25]	Python
10	[26]	Python
11	Present	Chat GPT-3.5 Open AI and Google Bard

This study bridges the gap with previous research by employing Chat GPT-3.5 Open AI and Google Bard in the Text Pre-Processing stages, namely Cleaning Text, Transforming Cases, and Labeling. Rapidminer Studio is utilized in the Tokenizing, Transforming Cases, Stopwords Filters, Token Filters (by length), and Stemming stages. Performance measurement is then conducted by employing the Decision Tree and Naïve Bayes algorithms, combined with the Linear, Stratified, and Shuffled sampling methods. Model Validation is performed using K-Fold Cross Validation, with values of K (K=10) to determine the best model validation performance.

Stratified Sampling is employed when dealing with populations that are divided into groups [27]. Within each group, sample selection is conducted randomly and systematically. Linear Sampling is a

sampling method that divides a sample set into partitions while preserving the original order [27], [28]. Shuffled Sampling is a random sampling technique that generates a random subset from a portion of the data examples [27], [29].

The dataset used was the result of crawling from 1000 Twitter users. The keyword used during the data crawling on Twitter was "Artificial Intelligence," which referred to the increasing popularity of Artificial Intelligence in recent years [30]. Therefore, this keyword was chosen to observe Twitter users' opinions based on that keyword.

The Decision Tree algorithm is used in this study due to its advantages in handling attributes with discrete or continuous data types, ability to handle missing values, capability for tree pruning, and attribute selection using gain ratio [31]. The use of this algorithm is supported by a research [32], which showed that the Decision Tree algorithm achieved 91% accuracy.

The Naïve Bayes algorithm is also employed as a comparison to the Decision Tree algorithm. Naïve Bayes is chosen for its simplicity, low computational complexity, and fast and efficient processing [33], [34]. This argument is validated by research [35], which demonstrated an accuracy rate of 96.24% for Naïve Bayes.

This study aims to utilize Chat GPT-3.5 Open AI-3.5 and Google Bard as alternative Text Pre-Processing methods in the Sentiment Analysis process of Twitter datasets. This research is essential as there is limited investigation on Text Pre-Processing using Chat GPT-3.5 Open AI and Google Bard. The performance of the Text Pre-Processing results from both Chatbots will be measured using the Decision Tree and Naïve Bayes algorithms.

2. METHOD

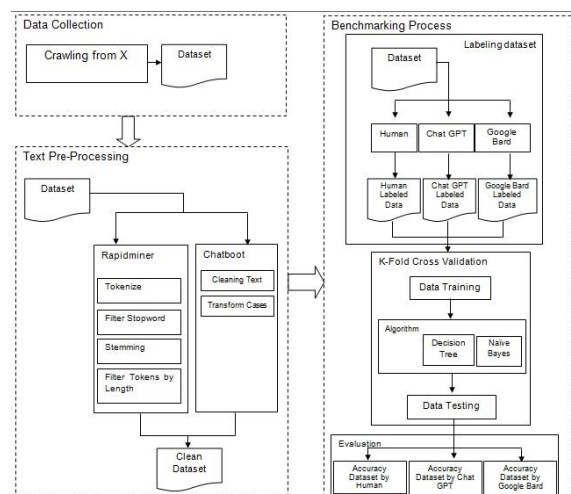


Figure 1. Research Method.

The proposed research framework aims to compare the performance results of Chat GPT-3.5 Open AI and Google Bard in the Text Pre-Processing

stage. Fig. 1 represents the complete workflow of each stage. First, the Dataset used consists of crawled data from Twitter. Second, the Text Pre-Processing stage is conducted using Chat GPT-3.5 Open AI and Google Bard. Third, the performance of the Text Pre-Processing stage is measured using K-Fold Cross Validation. Fourth, the accuracy results are compared between Chat GPT-3.5 Open AI and Google Bard to determine which one performs better.

2.1. Data Collection

The collected dataset is the result of Crawling technique. Crawling is an extraction technique from Twitter that utilizes an application programming interface (API) to access the information contained within [36]. This research gathered the Crawling data from Twitter, consisting of 1,000 records, using the Rapidminer application and entering the keyword "Artificial Intelligence" for data Crawling. The crawling process using RapidMiner is presented as depicted in Figure 2, and The initial data presentation is provided in tabular form, as shown in Table 2.

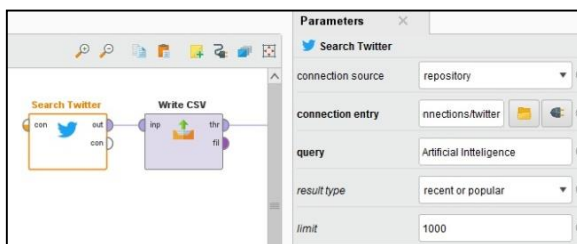


Figure 2. Crawling Dataset Process

Table 2. Preliminary Dataset

Create d-At	Langu age	Source	Posting- Count	Text
5/25/20 23 16:12	english	<ahref="http://twitter.com/download/iphone"rel="nofollow">Twitter for iPhone	6602	Developing talents and technologies in France
5/25/20 23 16:12	english	<ahref="http://twitter.com/download/iphone"rel="nofollow">Twitter for iPhone	0	This statement is very chilling... ?
5/25/20 23 16:12	english	<ahref="http://twitter.com/download/iphone"rel="nofollow">Twitter for ipad	0	Tesla CEO Elon Musk says "there should be a significant third horse" in the artificial intelligence race besides #Microsoft

Create d-At	Langu age	Source	Posting- Count	Text
				and #Google. #AI #dogecoin \$DOGE https://t.co/UJVGkCUmCF
5/25/20 23 16:12	english	<ahref="http://twitter.com/download/android"rel="nofollow">Twitter for Android	1	RT @ejrushing: Can humanitarians harness the power of #AI to track patterns of violence?
5/25/20 23 16:12	english	<ahref="http://twitter.com/download/android"rel="nofollow">Twitter for Android	76	RT @ejrushing: Can humanitarians harness the power of #AI to track patterns of violence?
...
5/24/20 23 3:56	english	Twitter Web App	105	RT @ejrushing: Can humanitarians harness the power of #AI to track patterns of violence?

There are five attributes in Table 2, namely Created-At, Language, Source, Retweet-Count, and Text. This research only utilizes the Text attribute in the Sentiment Analysis process, hence the remaining four attributes will be eliminated from the dataset. The Crawling results from Twitter do not have sentiment attributes for the tweets, thus it is necessary to assign sentiment to all the tweets. The sentiment categories that will be assigned are Positive, Negative, and Neutral. The process of labeling involved assigning labels to the review dataset. The prevalent method of labeling often entailed manual annotation with the aid of linguistic specialists. Manual annotation by human was applied in this study to serve as a comparison to the annotations generated by Chat GPT-3.5 Open AI and Google Bard. Manual annotation by human was performed by the research [37], yielding an accuracy rate of 79.29%.

2.2. Text Pre-Processing

Text Pre-Processing in Sentiment Analysis is a stage aimed at eliminating inconsistent data, duplicate data, and data that does not influence the polarity of the existing documents [38]. The Text Pre-Processing stage in this research is divided into two activities. The first activity is Text Pre-Processing using Chat GPT-3.5 Open AI and Google Bard. The tasks to be

performed by Chat GPT-3.5 Open AI and Google Bard are as follows:

2.2.1. Text Cleaning

Text Cleaning aims to remove punctuation, generalize the usage of uppercase or lowercase letters, eliminate duplicate tweets, and correct spelling [39]. In the Text attribute, the following elements are removed during this process: Twitter-specific attributes such as the characters RT, @, #, and links.

2.2.2. Transform Cases

This stage aims to convert all text in the Text attribute to either lowercase or uppercase [40]. In this research, lowercase is used for all text.

2.2.3. Labeling

The objective of this stage is to assign sentiment to the collected dataset. A new attribute, Sentiment, will be added to the dataset. The labeling categories used are Positive, Negative, and Neutral. The labeling process could be done manually by humans or automatically using algorithms. Manual labeling involved reading and human interpretation of the text, while automatic labeling used natural language processing and machine learning techniques.

The labeling process was not only carried out using a Chatbot but was also annotated by humans by examining the context of each dataset. This annotation was performed by three researchers to avoid misinterpretation of each sentence in the dataset. The purpose of human annotation is to serve as a comparison to the labeling results generated by the Chatbot.

All these stages are performed using Chat GPT-3.5 Open AI and Google Bard with a relatively fast estimated time, depending on the commands provided in the chat column of both Chatbots. This process takes approximately 5 minutes for each stage, using the following device specifications:

1. Processor Intel Pentium P6000 (1.86 GHz, 3 MB L3 Cache);
2. Intel HD Graphics;
3. 3 GB DDR 3 Memory;
4. 320 GB HDD. must be embedded into the text and not supplied separately.

With such specifications, it can work in approximately 5 minutes, and especially if the specifications exceed that, the job is likely to be much faster. The second activity is Text Pre-Processing using Rapidminer. The processes performed are as follows:

2.2.4. Tokenize

Tokenization is the process of splitting text into small pieces called tokens, aiming to facilitate computer processing of textual data [41]–[43].

2.2.5. Filter Stopwords

Stopword filtering is the process of eliminating frequently occurring words such as "a," "the," "of," "and," and "an," which do not contribute to the analysis process [44], [45]. This process also aims to reduce the dimensionality of the data, thereby speeding up computation time.

2.2.6. Stemming

Stemming aims to transform words into their base form by removing prefixes and suffixes from the words [46], [47].

2.2.7. Filter Token By Length

Filtering tokens by length refers to the procedure of removing words from the text that have a specific number of characters [48]. This process utilizes the Rapidminer application and involves setting the minimum character limit to 4 and the maximum character limit to 25. The goal is to restrict the length of words within the text, ensuring they are at least 4 characters long but no longer than 25 characters.

2.3. Validation

The Validation stage will display information on the accuracy performance of the constructed model. The algorithms used in this Validation process are the Decision Tree Algorithm and Naïve Bayes Algorithm, employing the K-Fold Cross-Validation (KCV) validation operator. Cross-validation is a technique used to assess the generalizability of statistical analysis outcomes to an independent dataset [49]. It involves evaluating how well the results of a model or analysis can be applied to new and unseen data. KCV involves dividing the dataset into k parts and performing k iterations. In each iteration, one part of the dataset is chosen as the testing data, while the remaining k - 1 parts are used for training. This process is repeated k times, and the average deviation (error) value is calculated based on the k different test results [50].

2.4. Evaluation

The final stage is Evaluation. The evaluation stage involves assessing the outcomes of applying the model to determine if it has achieved the research objectives. Based on this evaluation, a decision is made regarding the utilization of the modeling results [51]. During this stage, the accuracy value generated by the constructed model is evaluated. Accuracy refers to the percentage of data records that are correctly classified after testing the classification results [52].

3. RESULT AND DISCUSSION

3.1. Data Collection and Text Pre-Processing

The dataset obtained from Twitter Crawling, as shown in Table 1, is then subjected to Text Pre-Processing using Chat GPT-3.5 Open AI and Google

Bard. The Text Pre-Processing stages performed are Text Cleaning, Transforming Cases, and Labeling. As for the results of the Text Cleaning stage, they are presented in the form of images, as shown in Figure. 3 and Figure. 4.

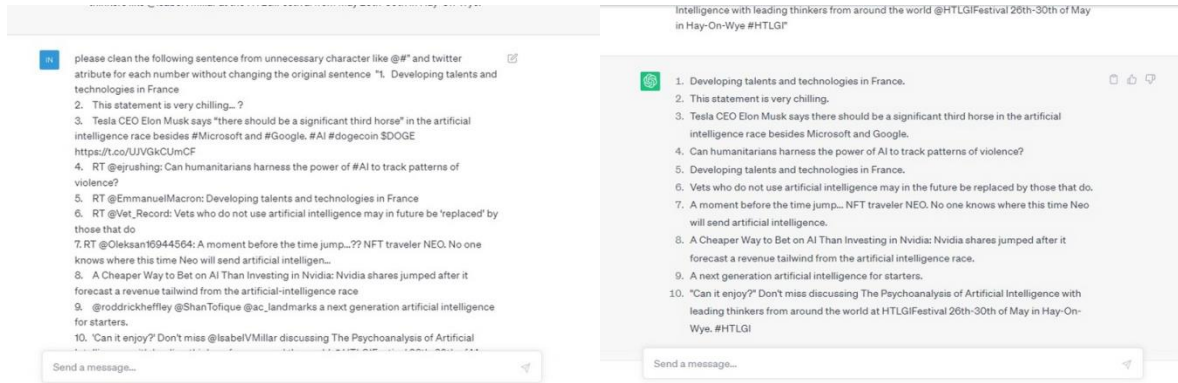


Figure. 3 (a) instructions to clean text by Chat GPT-3.5 Open AI (b) Cleaning Result

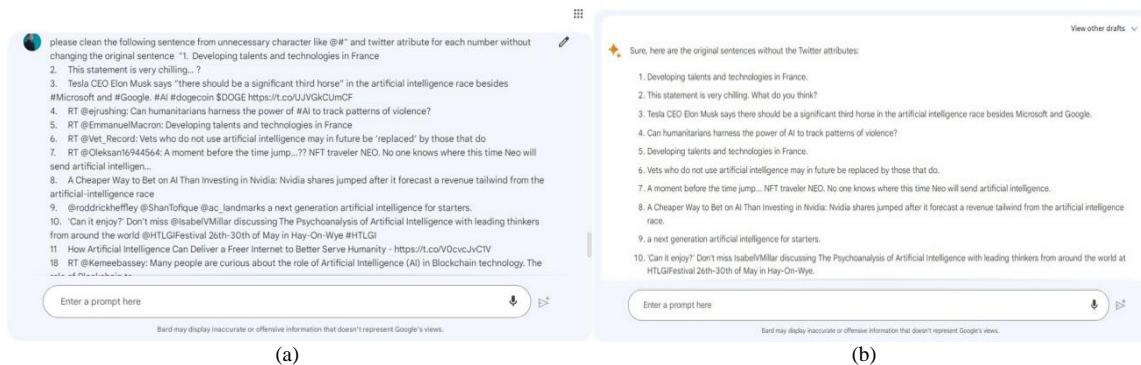


Figure. 4 (a) Instructions to clean text by Google Bard (b) Cleaning result

In Figure. 3 and Figure. 4, it can be observed that both Chat GPT-3.5 Open AI and Google Bard are capable of cleaning text from unnecessary characters, resulting in clear and meaningful sentences. The given commands must be specific, aiming for the desired outcome. In this study, the command provided for Cleaning Text is "Please remove the

following sentences from unnecessary characters like @# and preserve the Twitter attributes for each number without altering the original sentence." The subsequent step performed by Chat GPT-3.5 Open AI and Google Bard is the transformation of cases. The outcomes of this step are presented in Figure. 5 and Figure. 6, as depicted.

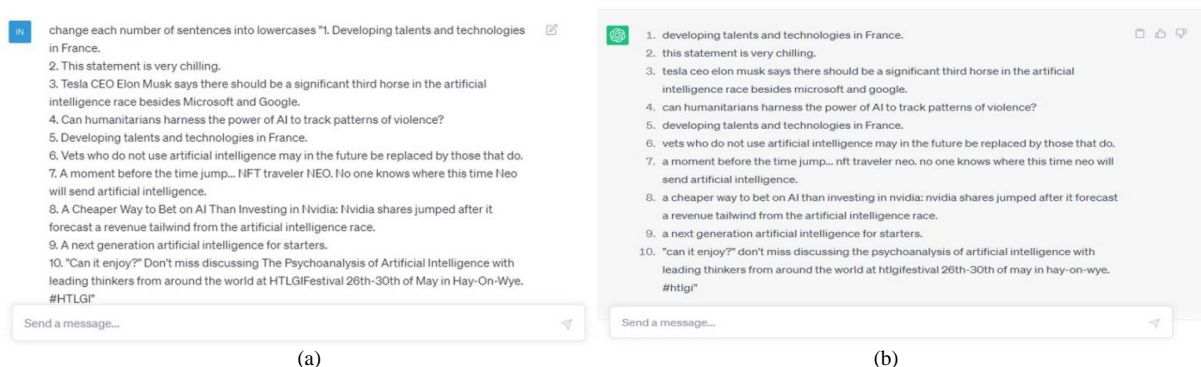


Figure. 5 (a) Transform Cases by Chat GPT-3.5 Open AI (b) Transform Cases Result

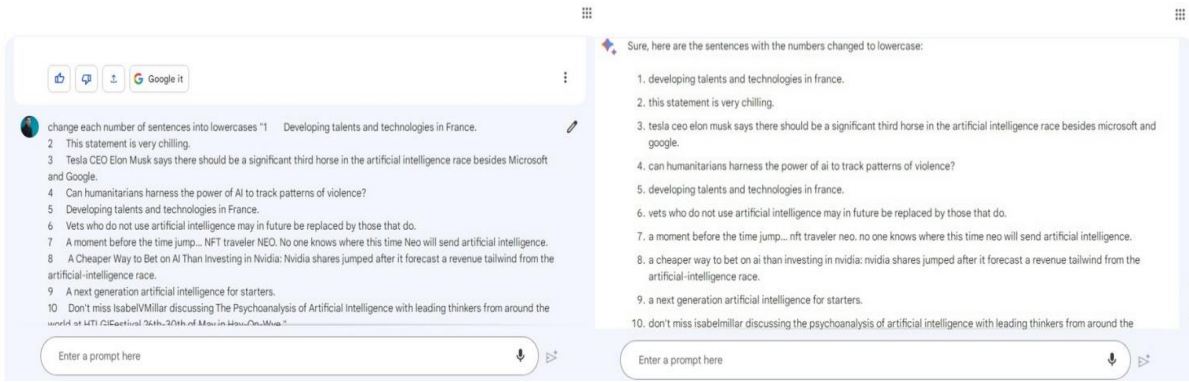


Figure. 6 (a) Transform Cases by Google Bard (b) Transform Cases Result

After the Cleaning Text process, each sentence is then transformed in the Transform Cases step into lowercase, ensuring uniformity in the text format across all datasets. Both Chat GPT-3.5 Open AI and Google Bard were given the same instruction, which is to "Change each number of sentences into

lowercase." As depicted in Fig. 4 and Fig. 5, both chatbots successfully carried out this instruction. The final step performed by Chat GPT-3.5 Open AI and Google Bard is Labeling. The outcomes of this step are presented in the form of images, as shown in Fig. 7 and Fig. 8.

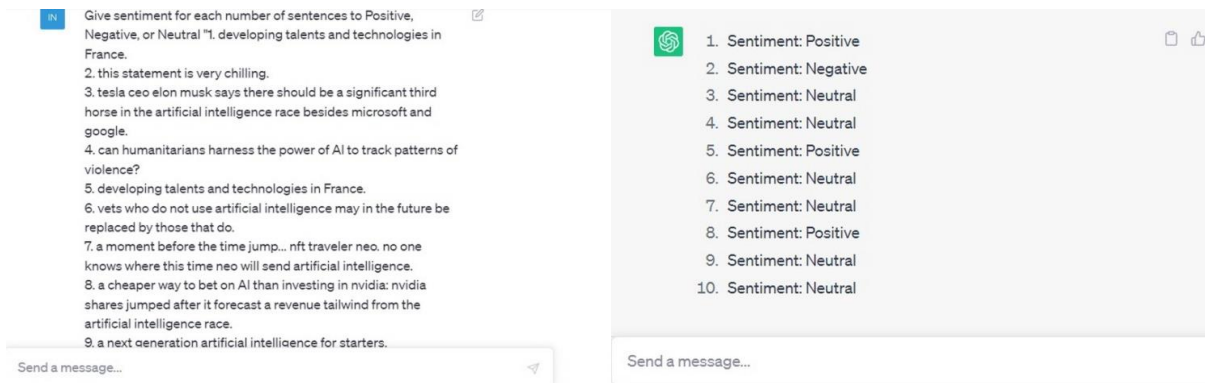


Figure. 7 (a) Labeling by Chat GPT-3.5 Open AI (b) Labeling Result

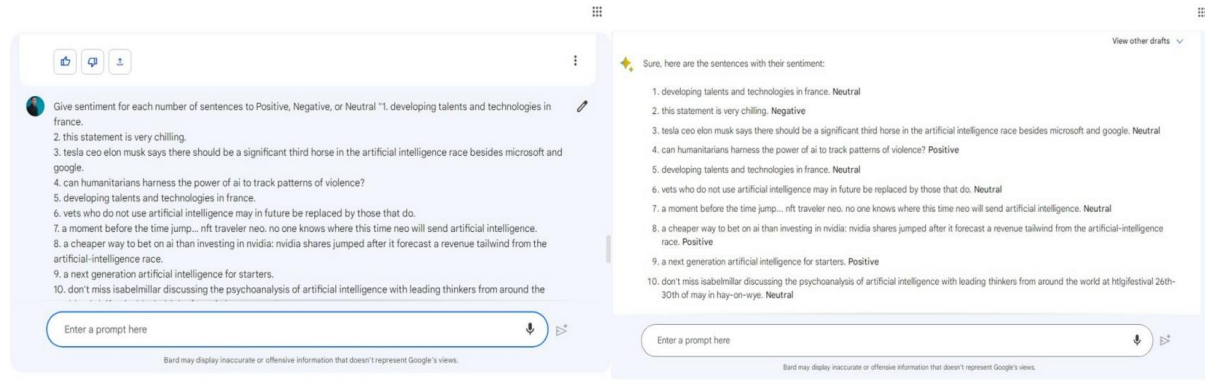


Figure. 8 (a) Labeling by Google Bard (b) Labeling Result

The Labeling step was successfully performed by both Chat GPT-3.5 Open AI and Google Bard, where by inputting the same command, "Assign sentiment for each sentence as Positive, Negative, or Neutral," they provided sentiment results for each sentence in the dataset. The Text Pre-Processing step using Chat GPT-3.5 Open AI and Google Bard is now completed, and the output from this step is then

inputted into Ms. Excel for later retrieval using Rapidminer. The Text Pre-Processing conducted using Rapidminer includes Tokenization, Stopword Filtering, Stemming, and Token Filtering (by length). The outcomes of this step are presented in Table 3, as depicted.

Table 3. Text Pre-Processing Result

Text Pre-Processing	Result
Preliminary Dataset	Tesla CEO Elon Musk says “there should be a significant third horse” in the artificial intelligence race besides #Microsoft and #Google. #AI #dogecoin \$DOGE https://t.co/UJVGkCUMCF
Cleaning Text	Tesla CEO Elon Musk says there should be a significant third horse in the artificial intelligence race besides
Tranform Cases	tesla ceo elon musk says there should be a significant third horse in the artificial intelligence race besides
Tokenize	["tesla", "ceo", "elon", "musk", "says", "ther e", "should", "be", "a", "significant", "third", "horse", "in", "the", "artificial", "intelligence", "race", "besides"]
Filter Stopword	["tesla", "ceo", "elon", "musk", "significant", "third", "horse", "artificial", "intelligence", "race"]
Stemming	["tesla", "ceo", "elon", "musk", "signific", "t hird", "hors", "artifici", "intellig", "race"]
Result	tesla ceo elon musk signific third hors artific intellig race

The Text Pre-Processing step has been completed. Next, based on the sentiment labels provided by Chat GPT-3.5 Open AI and Google Bard, the distribution can be visualized in the form of a chart, as depicted in Figure. 9.

Figure. 9 demonstrates the differences in label distribution between Chat GPT-3.5 Open AI and Google Bard. Google Bard assigns a higher

proportion of Positive sentiment, specifically 46%, in the discussion of Artificial Intelligence, compared to Chat GPT-3.5 Open AI, which assigns 40%. On the other hand, Chat GPT-3.5 Open AI assigns a lower proportion of Negative sentiment, specifically 19%, compared to Google Bard, which assigns 34%. As for the Neutral sentiment, Chat GPT-3.5 Open AI assigns a larger proportion, specifically 41%, while Google Bard assigns 20%. Moving forward, this data will be visualized using a Word Cloud.

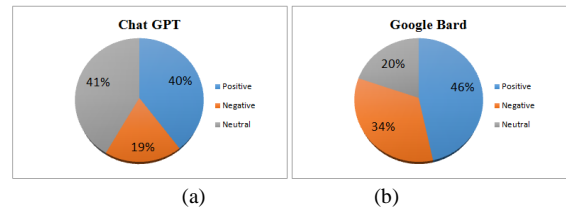


Figure. 9 (a) Percentage Rate Chat GPT-3.5 Open AI (b) Percentage Rate Google Bard

3.2. Validation

The Validation stage utilizes the Rapidminer application to measure the performance resulting from the Text Pre-Processing steps of Chat GPT-3.5 Open AI and Google Bard. The algorithms employed are Decision Tree and Naïve Bayes, which are then connected to the validation operator called K-Fold Cross Validation. In this stage, various values of K are utilized, including k=2, k=4, k=6, k=8, and k=10. These values are combined with three sampling methods: Linear, Stratified, and Shuffled sampling. The process of this stage is presented in a visual format, as shown in Figure. 10.

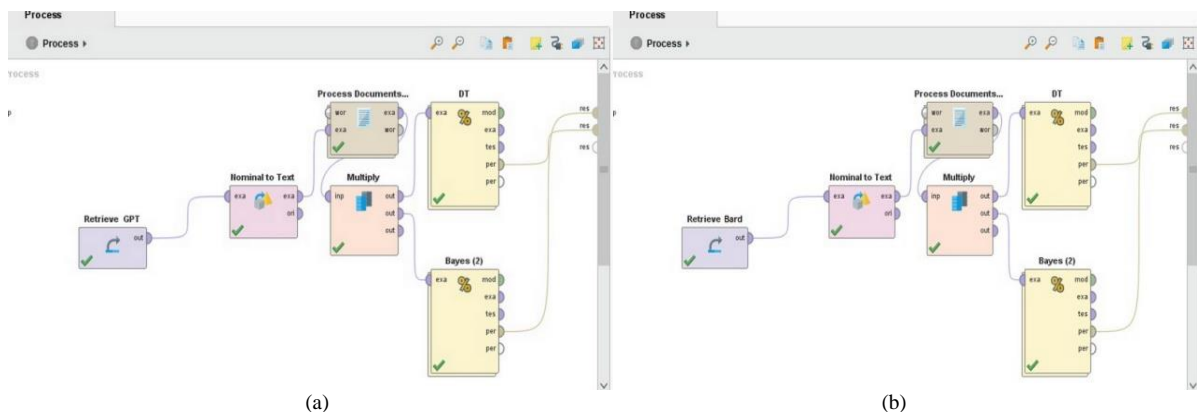


Figure. 10 (a) Model Process Chat GPT-3.5 Open AI (b) Model Process Google Bard

The validation process has been conducted for the Model Process Chat GPT-3.5 Open AI and the Model Process Google Bard. The next stage is to examine the results of the validation process, which will be discussed in the subsequent subsection, namely Evaluation.

3.3. Evaluation

The experimental results, combining various values of k, two algorithms, and three sampling methods, show the performance of the process model as presented in Table 4.

Table 4. Modelling Result

Chat GPT-3.5 Open AI					Google Bard				
Algorithms	Sampling	Accuracy	Precision	Recall	Algorithms	Sampling	Accuracy	Precision	Recall
Decision Tree	Linear	90.14%	90.14%	100%	Decision Tree	Linear	68.18%	77.78%	99.18%
	Stratified	90.68%	90.63%	100%		Stratified	68.24%	72.73%	98.77%
	Shuffled	90.41%	90.38%	100%		Shuffled	68.27%	72.73%	98.77%
Naïve Bayes	Linear	71.44%	93.61%	75.68%	Naïve Bayes	Linear	48.75%	75.48%	56.16%
	Stratified	75.62%	94.27%	79.94%		Stratified	52.05%	76.54%	60.27%
	Shuffled	74.00%	93.82%	78.42%		Shuffled	52.60%	77.78%	61.64%

Table 4 presents the measurement results of the Text-Pre Processing that were carried out using Chat GPT-3.5 Open AI and Google Bard with the implementation of Decision Tree and Naïve Bayes algorithms. For Chat GPT-3.5 Open AI, the highest accuracy was achieved through the combination of the Decision Tree algorithm using the Stratified Sampling method, with an accuracy score of 90.68%, precision of 90.14%, and recall of 100%. As for Google Bard, the best performance was observed with the combination of the Decision Tree algorithm using the Shuffled Sampling method, resulting in an accuracy score of 68.27%, precision of 72.72%, and recall of 98.77%.

When the results of Text Pre-Processing between Chat GPT-3.5 Open AI and Google Bard were compared, it was Chat GPT-3.5 Open AI that exhibited the best performance with an accuracy score of 90.68%, precision of 90.14%, and recall of 100%. However, these data were not considered valid until there was a point of comparison. Therefore, the next step involved comparing the results of Text Pre-Processing by Chat GPT-3.5 Open AI with the results of Text Pre-Processing using Annotation by Human. The comparative results between Text Pre-Processing Annotation by human with Chat GPT-3.5 Open AI are presented in Table format, as seen in Table 5.

Table 5. Benchmark Modelling Result

Annotation By Human					Chat GPT-3.5 Open AI				
Algorithms	Sampling	Accuracy	Precision	Recall	Algorithms	Sampling	Accuracy	Precision	Recall
Decision Tree	Linear	84.65%	84.85%	99.68%	Decision Tree	Linear	90.14%	90.14%	100%
	Stratified	85.20%	85.71%	99.03%		Stratified	90.68%	90.63%	100%
	Shuffled	84.91%	85.57%	99.03%		Shuffled	90.41%	90.38%	100%
Naïve Bayes	Linear	61.03%	89.24%	64.40%	Naïve Bayes	Linear	71.44%	93.61%	75.68%
	Stratified	64.65%	92.34%	66.34%		Stratified	75.62%	94.27%	79.94%
	Shuffled	63.00%	90.27%	66.02%		Shuffled	74.00%	93.82%	78.42%

Based on Table 5, the best performance values obtained from Text Pre-Processing Annotation by Human used as a benchmark against Chat GPT-3.5 Open AI, were achieved using the Decision Tree algorithm and the Stratified Sampling method, with an Accuracy of 85.20%, Precision of 85.71%, and Recall of 99.03%. This indicates that the Text Pre-Processing generated by Chat GPT-3.5 Open AI outperformed the performance results of Annotation by Human, where the Accuracy achieved was 90.68%, Precision was 90.63%, and Recall was 100%. This study successfully found that chatbots like ChatGPT and Google Bard can perform the Text Pre-Processing phase in Sentiment Analysis. The performance measurement generated by the chatbots was conducted using classification, ultimately resulting in performance metrics such as accuracy, precision, and recall. Subsequently, validation was carried out by comparing the performance results to those obtained through human annotation. The comparative results are presented in the form of a graph, as shown in Fig. 11.

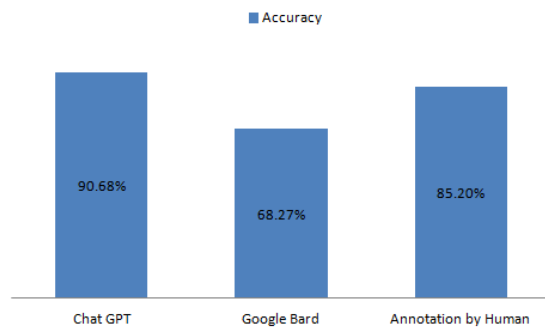


Figure. 11 Performance Result

The best Text Pre-Processing results are demonstrated by Chat GPT-3.5 Open AI, which achieves an Accuracy score of 90.68%, Precision 90.63%, and Recall 100% using a combination of the Decision Tree algorithm, 10-Fold Cross Validation, and Stratified Sampling as the sampling method. On the other hand, Google Bard obtains a lower performance score, with an Accuracy of 68.27%, Precision 72.73%, and Recall 98.77%. This result is obtained using the Decision Tree algorithm, 10-Fold Cross Validation, and Shuffled Sampling as the

sampling method. The results of Text Pre-Processing using Chat GPT-3.5 Open AI surpassed the values generated by Google Bard and even the benchmark models Annotation by Human with an Accuracy of 85.20%, Precision of 85.71%, and Recall of 99.03%. The possible reason for Chat GPT-3.5 Open AI outperforming Google Bard is that Google Bard was still in the testing phase and had not been launched yet. In contrast, Chat GPT-3.5 Open AI had already been launched in 2022.

This research proves that Chatbots such as Chat GPT-3.5 Open AI and Google Bard can be considered as alternative text pre-processing methods in Sentiment Analysis, offering an option alongside Python-based text pre-processing as presented in studies [5], [24]–[26], or solely relying on Rapidminer as shown in studies [3], [4], [6], [21], [22]. This research stands out as unique compared to previous studies because it combines widely used Text Pre-Processing tools like Rapidminer with Artificial Intelligence products, namely Chat GPT-3.5 Open AI and Google Bard. Furthermore, this study also aims to compare the performance achieved by these two Chatbots. The objective is to obtain more optimal results in Sentiment Analysis. Moreover, the performance achieved is also considered to be in the category of Fair Classification, indicating a satisfactory level of accuracy.

4. CONCLUSION

This research has successfully contributed to Text Pre-Processing in Sentiment Analysis by utilizing Chatbot Chat GPT-3.5 Open AI and Google Bard for Text Cleaning, Transform Cases, and Labeling. The best performance results were achieved by Chat GPT-3.5 Open AI, using a combination of the Decision Tree Algorithm, 10-Fold Cross Validation and the Stratified Sampling method, resulting in an accuracy score of 90.68%, precision 90.63%, and recall 100%. These findings demonstrate that Chatbots such as Chat GPT-3.5 Open AI and Google Bard are suitable as Alternative Text Pre-Processing methods in the Sentiment Analysis process. There are some limitations in this study. Firstly, the research focuses on testing the capabilities of Chatbots, namely Chat GPT-3.5 Open AI and Google Bard, in the Text Pre-Processing process, without making improvements to the constructed process model. Secondly, the algorithms used, Decision Tree and Naïve Bayes, are employed solely for the purpose of comparing the performance of Text Pre-Processing generated by the two Chatbots. In future research, improvements to the performance of the constructed model can be applied to Alternative Text Pre-Processing using Chat GPT-3.5 Open AI and Google Bard, such as the use of Feature Selection or Backward Elimination. Additionally, the inclusion of other comparative algorithms can be considered to determine if there are alternative algorithms besides Decision Tree that exhibit more optimal performance.

ACKNOWLEDGMENTS

Authors wishing to acknowledge Institut Teknologi Garut that support and funds this research publication.

REFERENCES

- [1] L. Hermawan and M. B. Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *Transformatika*, vol. 17, no. 2, pp. 188–199, 2020.
- [2] S. K. Assayed, K. Shaalan, M. Alkhatib, and S. Maghaydah, "Machine Learning Chatbot for Sentiment Analysis of Covid-19 Tweets," in *Computer Networks & Communications*, Feb. 2023, pp. 41–55, doi: 10.5121/csit.2023.130404.
- [3] B. Kurniawan Rachmat, A. Suwarisman, I. Afriyanti, A. Wahyudi, and D. D. Saputra, "Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTE dan Naïve Bayes," *J. Teknol. Inf. dan Komunikasi*, vol. 7, no. 1, pp. 107–113, 2023, [Online]. Available: <https://doi.org/10.35870/jti>.
- [4] M. Fahmi, Y. Yuningsih, and A. Puspita, "Sentiment Analysis Of Online Gojek Transportation Services On Twitter Using The Naïve Bayes Method," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 8, no. 2, pp. 84–90, 2023, doi: 10.33480/jitk.v8i2.4004.
- [5] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [6] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 7, no. 1, pp. 11–18, 2022.
- [7] R. Parlika, S. I. Pradika, A. M. Hakim, and K. R. N. M., "Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob," *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 33–37, 2020.
- [8] M. Sarosa, M. Kusumawardani, A. Suyono, and Z. Sari, "Implementasi Chatbot Pembelajaran Bahasa Inggris menggunakan Media Sosial," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 317, 2020, doi: 10.26418/jp.v6i3.43191.
- [9] M. Dowling and B. Lucey, "ChatGPT for (Finance) research: The Bananarama Conjecture," *Financ. Res. Lett.*, no. 103662, pp. 1–20, 2023, doi:

- 10.1016/j.frl.2023.103662.
- [10] I. T. Julianto, D. Kurniadi, Y. Septiana, and A. Sutedi, "Alternative Text Pre-Processing using Chat GPT-3.5 Open AI-3.5 Open AI," *Janapati*, vol. 12, no. 1, pp. 67–77, 2023, [Online]. Available: <https://wjaets.com/content/artificial-intelligence-ai-based-chatbot-study-chatgpt-google-ai-bard-and-baidu-ai>.
- [11] M. S. Rahaman, M. M. T. Ahsan, N. Anjum, M. M. Rahman, and M. N. Rahman, "The AI Race is on! Google's Bard and Openai's Chatgpt Head to Head: An Opinion Article," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4351785.
- [12] Google, "An important next step on our AI journey," *google.blog*, 2023. <https://blog.google/technology/ai/bard-google-ai-search-updates/> (accessed May 27, 2023).
- [13] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, "Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification," *Appl. Sci.*, vol. 10, no. 23, pp. 1–26, 2020, doi: 10.3390/app10238631.
- [14] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Appl. Sci.*, vol. 12, no. 17, pp. 1–21, 2022, doi: 10.3390/app12178765.
- [15] N. Garg and K. Sharma, "Text pre-processing of multilingual for sentiment analysis based on social network data," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 1, pp. 776–784, 2022, doi: 10.11591/ijece.v12i1.pp776-784.
- [16] V. V. Nhlabano and P. E. N. Lutu, "Impact of Text Pre-processing on the Performance of Sentiment Analysis Models for Social Media Data," *2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst.*, pp. 1–6, 2018.
- [17] L. G. Irham, A. Adiwijaya, and U. N. Wisesty, "Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine," *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 284, 2019, doi: 10.30865/mib.v3i4.1410.
- [18] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *J. Teknol. Inf.*, vol. 18, no. 2, pp. 173–184, 2021.
- [19] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap PSBB Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 2, no. 2, pp. 91–97, 2021, doi: 10.37859/coscitech.v2i2.2851.
- [20] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Sentimen Analisis Belajar Online Di Twitter Menggunakan Naïve Bayes," *JANAPATI*, vol. 11, no. 3, pp. 310–320, 2022.
- [21] A. H. Anshor and A. Safuwani, "Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization," *JINTEKS (Jurnal Inform. Teknol. dan Sains)*, vol. 5, no. 1, pp. 170–178, 2023.
- [22] F. Syah, H. Fajrin, A. N. Afif, M. R. Saeputra, D. Mirranty, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *urnal JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 54–58, 2023.
- [23] M. R. Qisthiano, I. Ruswita, and P. Armilia, "Implementasi Metode SVM dalam Analisis Sentimen Mengenai Vaksin dengan Menggunakan Python 3," *J. Ilm. Sist. Inf.*, vol. 13, no. 1, pp. 1–7, 2023.
- [24] D. Setiyawati and N. Cahyono, "Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 262–272, 2023.
- [25] M. T. Anwar, D. Riandhita, A. Permana, P. Sistem, I. Industri, and J. Pusat, "Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 783–792, 2023.
- [26] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," *J. TEKNO KOMPAK*, vol. 17, no. 1, pp. 25–38, 2022.
- [27] A. Fauzy, *Metode Sampling*. Tangerang Selatan: Universitas Terbuka, 2019.
- [28] A. A. Abdillah, A. Azwardi, and I. Wahyudi, "Optimasi Linear Sampling dan Informasi Gain pada Algoritma Decision Tree untuk Diagnosis Penyakit Diabetes," *Multinetics*, vol. 7, no. 1, pp. 21–29, 2021.
- [29] A. Bisri and M. Man, "Machine Learning Algorithms Based on Sampling Techniques for Raisin Grains Classification," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 7–14, 2023, doi: 10.30630/joiv.7.1.970.
- [30] L. Vincent, H. A. Indahsi, and D. Suryawinata, "Klasifikasi Jenis-Jenis Anjing Menggunakan GoogleNet," *J. Mach. Learn. Comput. Intell.*, vol. 2, no. 1, pp. 5–8, 2023.

- [31] J. Amalia, N. Yosevin Nababan, K. G. Tambunan, and I. S. Sinaga, "Decision Tree Dengan Binary Bat Algorithm Optimization Pada Heart Catheterization Prediction," *Hexag. J. Tek. dan Sains*, vol. 3, no. 2, pp. 46–51, 2022, doi: 10.36761/hexagon.v3i2.1640.
- [32] A. A. Arifiyanti, M. F. Pandji, and B. Utomo, "Analisis Sentimen Ulasan Pengunjung Objek Wisata Gunung Bromo pada Situs Tripadvisor," *Explor. J. Sist. Inf. dan Telemat.*, vol. 13, no. 1, p. 32, 2022, doi: 10.36448/jisit.v13i1.2539.
- [33] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *Prisma*, vol. 4, pp. 628–634, 2021.
- [34] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021.
- [35] M. Siddik, H. Hendri, R. N. Putri, Y. Desnelita, and G. Gustientiedina, "Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 162–166, 2020, doi: 10.31539/intecom.v3i2.1654.
- [36] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, vol. 1, no. 1, pp. 299–305, 2021.
- [37] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [38] M. Dennis, F. Zoromi, and M. K. Anam, "Penerapan Algoritma Naïve Bayes Untuk Pengelompokan Predikat Peserta Uji Kemahiran Berbahasa Indonesia," *J. Media Inform. Budidarma*, vol. 6, no. 2, pp. 1183–1190, 2022, doi: 10.30865/mib.v6i2.3956.
- [39] Junadhi, Agustin, M. Rifqi, and M. K. Anam, "Sentiment Analysis Of Online Lectures Using K-Nearest Neighbors Based On Feature Selection," *Janapati*, vol. 11, no. 3, pp. 216–225, 2022.
- [40] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Comparison Of Classification Algorithm And Feature Selection in Bitcoin Sentiment Analysis," *JUTIF*, vol. 3, no. 3, pp. 739–744, 2022.
- [41] M. Madine, K. Salah, R. Jayaraman, A. Battah, H. Hasan, and I. Yaqoob, "Blockchain and NFTs for Time-Bound Access and Monetization of Private Data," *IEEE Access*, vol. 10, pp. 94186–94202, 2022, doi: 10.1109/ACCESS.2022.3204274.
- [42] A. K. Fauziyyah and D. H. Gautama, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *J. Ilm. SINUS*, vol. 18, no. 2, pp. 31–42, 2020, doi: 10.30646/sinus.v18i2.491.
- [43] N. Alnuaimi, A. Almamari, M. Madine, K. Salah, H. Al Breiki, and R. Jayaraman, "NFT Certificates and Proof of Delivery for Fine Jewelry and Gemstones," *IEEE Access*, vol. 10, pp. 101263–101275, 2022, doi: 10.1109/ACCESS.2022.3208698.
- [44] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 3, pp. 475–488, 2020, doi: 10.28932/jutisi.v6i3.2892.
- [45] L. K. Harsono, Y. Alkhalifi, Nurajijah, and W. Gata, "Analisis Sentimen Stakeholder atas Layanan haiDJPb pada Media Sosial Twitter Dengan Menggunakan Metode Support Vector Machine dan Naïve Bayes," *J. Ilmu-ilmu Inform. dan Manaj.*, vol. 14, no. 1, pp. 36–44, 2020.
- [46] A. Ahmad and W. Gata, "Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 548–555, 2022, doi: 10.35870/jtik.v6i4.569.
- [47] A. C. Herlingga, I. P. E. Prisma, D. R. Prehanto, and D. A. Dermawan, "Algoritma Stemming Nazief & Adriani dengan Metode Cosine Similarity untuk Chatbot Telegram Terintegrasi dengan E-layanan," *J. Informatics Comput. Sci.*, vol. 2, no. 01, pp. 19–26, 2020, doi: 10.26740/jinacs.v2n01.p19-26.
- [48] M. K. Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023.
- [49] G. Feng, M. Fan, and Y. Chen, "Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining," *IEEE Access*, vol. 10, pp. 19558–19571, 2022, doi: 10.1109/ACCESS.2022.3151652.

- [50] I. T. Julianto, D. Kurniadi, and F. M. Khoiriyyah, "Price Prediction of Non-Fungible Tokens (NFTs) using Data Mining Prediction Algorithm," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Feb. 2023, pp. 633–637, doi: 10.1109/ICCoSITE57641.2023.10127679.
- [51] Yunitasari, H. S. Hopipah, and R. Mayasari, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes," *Technomedia J.*, vol. 6, no. 1, pp. 99–110, 2021, doi: 10.33050/tmj.v6i1.1531.
- [52] D. Nurlaela, "Penerapan Adaboost untuk Meningkatkan Akurasi Naive Bayes Pada Prediksi Pendapatan Penjualan Film," *Inti Nusa Mandiri*, vol. 14, no. 2, pp. 181–188, 2020.