

# Opinion Mining on Chat GPT based on Twitter Users

Muhammad Rikza Nashrulloh<sup>1</sup>, Indri Tri Julianto\*<sup>2</sup>, Rifky Khoerul Muzaky<sup>3</sup>

<sup>1,2,3</sup> Department of Computer Science, Institut Teknologi Garut, Garut

E-mail :rikza@itg.ac.id<sup>1</sup>, indritrijulianto@itg.ac.id\*<sup>2</sup>, 2207013@itg.ac.id<sup>3</sup>

\*Corresponding Author

---

**Abstract** – The presence of Chatbots can assist humans in their everyday lives. Chat GPT is one of the commonly used Chatbots that humans rely on to support their work, serve as an assistant, or even create artistic works or writings. The purpose of this research is to investigate opinions regarding the presence of Chat GPT. This Opinion Mining method is conducted by crawling data from Twitter, which can be categorized into three opinions: Positive, Negative, or Neutral. To calculate the accuracy level of the model created, two algorithms, Naïve Bayes and K-Nearest Neighbour, are compared. The model validation process utilizes K-Fold Cross Validation by varying the value of k (k=2, k=4, k=6, k=8, and k=10) and different sampling methods, namely Linear, Shuffled, and Stratified, to obtain optimal accuracy values. The research results indicate that the K-Nearest Neighbour Algorithm achieves the highest accuracy value of 92.40%. Based on this comparison, the K-Nearest Neighbour Algorithm is deemed suitable for modeling Opinion Mining of Chat GPT. The distribution of Twitter users opinion percentages regarding Chat GPT is as follows: Positive 9.4%, Negative 1.4%, and Neutral 89%. Neutral opinions dominate the results of the conducted Opinion Mining.

**Keywords** – chat GPT, opinion mining, twitter

## 1. INTRODUCTION

---

A Chatbot is a program within the field of Artificial Intelligence that can process inputs and provide information through interactive chat interactions[1]. Chat Generative Pre-trained Transformer (GPT) is a Chatbot developed by the non-profit organization OpenAI, founded in 2015 in collaboration with Microsoft, Reid Hoffman's charitable foundation, and with Khosla Ventures as its primary investor [2]. At the end of 2022, Chat GPT was launched and gained significant attention from the public due to its advanced capabilities in providing information surpassing conventional search engines [3].

Chat GPT is an extensively used computer program that simulates human-like interaction, possessing the ability to comprehend natural language and generate responses that resemble human behavior [4]. It utilizes complex algorithms that have been trained on extensive textual data to deliver coherent and logical replies. This model is the solution with a combination of Reinforcement Learning Algorithm and human input of more than 150 billion parameters [5], [6].

The phenomenon of Chat GPT has sparked debate and controversy within society. Some people support it, while others view the presence of Chat GPT as a threat to humanity. This makes the upcoming research essential, as it is important to understand public opinion regarding Chat GPT. Social media platforms like Twitter provide a platform for individuals to express their opinions about a particular object or ongoing phenomenon [7]. It is estimated that

554.7 million people worldwide actively use Twitter, with the possibility of continued growth and around 58 million collective tweets posted daily [8].

Based on these data and facts, the research aims to conduct Opinion Mining on the existence of Chat GPT. Twitter is used to gather public opinions, which will help determine whether the presence of Chat GPT is acceptable or not. Opinion Mining is the process of detecting, identifying, and classifying opinions within unstructured text [9], [10]. Generally, Opinion Mining is performed by categorizing text into Positive, Negative, or Neutral categories [11].

There have been several previous studies that have discussed Opinion Mining. Firstly, a study examined Opinion Mining on Key Performance Indicator (KPI) functions [12]. The algorithms used in this research were Naïve Bayes Classifier and Support Vector Machine (SVM). The study's results showed that the SVM algorithm demonstrated optimal performance with an accuracy of 72.32%. Secondly, there was a study on Opinion Mining to analyze an application product, specifically Jamsostek Mobile [13]. The algorithm used in this study was Long Short-Term Memory. The research results indicated that the algorithm achieved an accuracy of 87.36%. Thirdly, there was a study on Opinion Mining regarding online learning activities [7]. The algorithm used was Naïve Bayes, and the study showed that when Naïve Bayes was applied in the Opinion Mining process, it resulted in an accuracy of 74.08%. Fourthly, a study focused on Opinion Mining on the activities of the G20 Summit in Indonesia [14]. The algorithm used was K-Nearest Neighbour, and the research revealed that the K-Nearest Neighbour algorithm achieved an accuracy of 99%. Lastly, there was a study on Opinion Mining regarding academic services in a university environment [11]. The algorithm used in this study was also K-Nearest Neighbour, and the research demonstrated an accuracy of 86%. A summarized comparison of the previous studies is presented in Table 1.

Table 1. Research Roadmap

| Research | Algorithms         | Dataset  | Outcome        |
|----------|--------------------|--|----------------|
| 1        | Naïve Bayes & SVM  | Crawling From Twitter (KPI Opinion)                    | Opinion Mining |
| 2        | LSTM               | Crawling From Twitter (Jamsostek Mobile App)           | Opinion Mining |
| 3        | Naïve Bayes        | Crawling From Twitter (Online Learning)                | Opinion Mining |
| 4        | K-NN               | Crawling From Twitter (KTT G20 in Indonesia)           | Opinion Mining |
| 5        | K-NN               | Crawling From Twitter (Academic Service at University) | Opinion Mining |
| Present  | Naïve Bayes & K-NN | Crawling From Twitter (Opinion Mining Chat GPT)        | Opinion Mining |

Note:

SVM (Support Vector Machine)

LSTM (Long-Short Term Memory)

K-NN (K-Nearest Neighbour)

KPI (Key Performance Indicator)

This research addresses the gap in previous studies by utilizing a Crawling dataset from Twitter specifically focusing on Chat GPT, which has not been extensively explored before. The algorithms employed in this study are Naïve Bayes and K-Nearest Neighbour (K-NN), as these algorithms have been commonly used in previous research. The validation process for the constructed model utilizes K-Fold Cross Validation and three sampling methods: Linear, Shuffled, and Stratified.

Stratified Sampling is employed when dealing with populations that are divided into groups [15]. 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 [15], [16]. Shuffled Sampling is a random sampling technique that generates a random subset from a portion of the data examples [15], [17].

The Naïve Bayes algorithm was selected for its benefits in the classification process. It stands out because it can work well with limited training data, as it only requires a small amount of it to determine the necessary parameters. Additionally, this algorithm is known for its speed and efficiency [18]. The effectiveness of the Naïve Bayes algorithm has been demonstrated in a study titled "Sentiment Analysis of Brimo Application Reviews on Google Play Using Naive Bayes Algorithm" [19].

The K-Nearest Neighbour algorithm was chosen for its advantageous characteristics in dealing with noisy training data, quick training execution, simplicity of comprehension, and capacity to handle large datasets [20]. Its effectiveness has been validated in a research titled "Application of K-Nearest Neighbor Method in Twitter User Sentiment Analysis on the G20 Summit in Indonesia" [14], where the classification outcomes achieved using the K-Nearest Neighbour algorithm were classified as highly successful in terms of classification quality.

## 2. RESEARCH METHOD

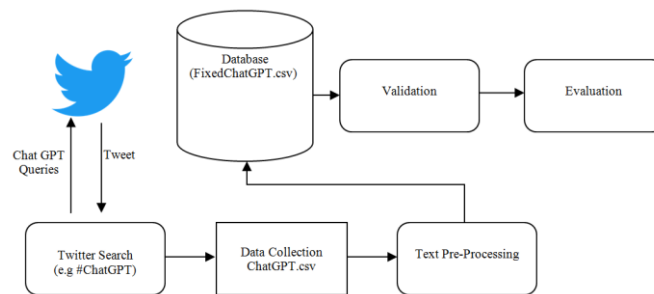


Figure 1. Research Method

The first stage involves collecting the dataset. The dataset is gathered from Twitter using a technique called Crawling. The Crawling process is performed using the Rapidminer application by inputting the keyword "Chat GPT," which retrieves relevant tweets related to Chat GPT. The collected dataset consists of 1,000 data points. Next, the dataset is saved in a .csv file format and unnecessary columns are removed. The results of this process are presented in the form of a table, as shown in Table 2.

Table 2. Dataset

| No    | Tweet   |
|-------|---|
| 1     | RT @CorkChamber:code  |
| 2     | RT @Hawkeye1745: ????'Lie'  |
| 3     | RT @neuro_rights: OK - now the Chat GPT case, Mata v. Avianca, sanctions against the lawyers who filed a brief with non-existent or hal...; |
| 4     | RT @technology: You told me that Chat GPT supplemented your research - but what was it supplementing...;                                    |
| ....  | ....  |
| 1.000 | RT @LeilaniDowding: I wrote, designed, and published my daughters first book in one weekend with the help of A.I.                           |

Table 2 displays the extracted "Tweet" column during the Opinion Mining process. Each tweet will later be assigned a opinion label of either Positive, Negative, or Neutral.

The second stage is Text Pre-Processing, which is an important stage before entering the modeling process, where data cleaning and attribute selection will be carried out as needed [21]. Text preprocessing is a crucial step aimed at addressing issues that may hinder the accuracy of

data processing outcomes [22]. Its purpose is to remove any obstacles or challenges that could potentially impact the quality and effectiveness of subsequent data analysis [23]. By conducting text preprocessing, potential problems such as noise, irrelevant information, and inconsistencies can be mitigated or eliminated, ensuring that the data is in a suitable format for further processing and analysis. The method in Text Pre-Processing is as follows:

1. The Cleaning Text : Text cleaning is an integral part of the text preprocessing stage, as the data often lacks structure and consistency. The purpose of text cleaning is to address these issues by removing punctuation, standardizing capitalization, eliminating duplicate tweet data, and rectifying spelling errors. These cleaning procedures help to enhance the quality and uniformity of the text data, making it more suitable for subsequent analysis [24]. Text cleaning to remove characters from Twitter such as, (@, RT,#,link);
2. Transform Cases : This stages refers to the process of converting text into either all lowercase or all uppercase [25]. This conversion can be applied to the entire text, changing it to lowercase or uppercase letters uniformly. The purpose of case transformation is to standardize the text and eliminate any inconsistencies that may arise from variations in capitalization [26]. By transforming cases, the text becomes more uniform and easier to process and analyze;
3. Labeling: The labeling process utilizes the Valence-aware dictionary and sentiment reasoner (VADER) method. VADER is a method used to measure the sentiment of text by leveraging a collection of lexical features, such as words, categorized as positive or negative based on their semantic orientation. Through VADER, we can determine the likelihood of a given input statement having a positive (+1), negative (-1), or neutral (0) opinion, along with its associated score [27]. The formula for VADER is as follows:

$$\frac{x}{\sqrt{x^2+a}} \quad (1)$$

In this case, x represents the cumulative sum of sentiment values assigned to the individual words within a sentence, while alpha is a default normalization constant set to 15. Consequently, VADER sentiment analysis achieves optimal performance when applied to shorter documents like tweets and sentences, as opposed to larger documents [28];

4. Tokenize : Tokenization is a procedure that involves splitting the text from documents into sequential tokens [29]–[31]. This process aims to divide the text into smaller units, such as words or subwords, for further analysis or processing [32], [33]. By tokenizing the text, it becomes organized into distinct units, enabling easier manipulation and understanding of the underlying information;
5. Filter Stopword : is a procedure that involves eliminating commonly occurring words such as ("a," "the," "of," "and," and "an.") [34]–[36];
6. Stemming : Stemming involves converting words into their core forms or essential components [37].
7. Filter tokens by length : is refers to the procedure of removing words from the text that have a specific number of characters [38].

The third stage is 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 [38]. It involves evaluating how well the results of a model or analysis can be applied to new and unseen data. KVC 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 [39]

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 [40]. 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 [41], [42]. Accuracy can be classified into different categories based on certain criteria:

- Accuracy 0.90 – 1.00 = Excellent classification;
- Accuracy 0.80 – 0.90 = Good classification;
- Accuracy 0.70 – 0.80 = Fair classification;
- Accuracy 0.60 – 0.70 = Poor classification;
- Accuracy 0.50 – 0.60 = Failure.

### 3. RESULTS AND DISCUSSION

The results of the Dataset and Text Pre-Processing stages are presented in Table 3.

Table 3. Dataset

| No    | Tweet  | Positivity Score | Negativity Score | Opinion  |
|-------|--|------------------|------------------|----------|
| 1     | code   | 0.00             | 0.00             | Neutral  |
| 2     | lie  | 0.00             | 0.00             | Neutral  |
| 3     | ok now the chat gpt case mata v avianca sanctions against the lawyers who filed a brief with non existent or hal | 0.00             | 0.31             | Positive |
| 4     | you told me that chat gpt supplemented your research but what was it supplementing                               | 0.00             | 0.00             | Neutral  |
| ....  | ....   |                  |                  |          |
| 1.000 | i wrote designed and published my daughters first book in one weekend with the help of a.i                       | 0.00             | 0.44             | Positive |

The opinion labeling in Table 3 was performed using the VADER method available in the Rapidminer application. There is an example of an error in the labeling results using VADER, so it is necessary to correct the mistake. The specific example is presented as shown in the following. Table 4.

Table 4. Labeling Error

| Tweet  | Positivity Score | Negativity Score | Opinion |
|--|------------------|------------------|---------|
| this won't help my bad habit of reading something then immediately wondering if it was written by chat gpt | 0.64             | 0.44             | Neutral |

In Table 4, the tweet should have been labeled as negative, but it was given scores in both the Positivity and Negativity columns, which led to it being considered as neutral. After assigning the opinion labels, we can determine the percentage distribution for each opinion. The distribution can be visualized in the form of a Pie Chart, as shown in Figure 2.

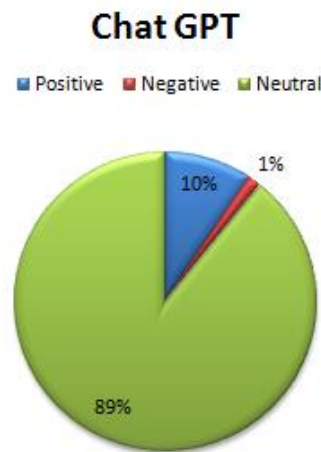


Figure 2. Percentage Rate

The percentage distribution for the opinion labels is as follows: 10% for Positive, 1% for Negative, and 89% for Neutral, out of the total collected data of 1,000 instances. Additionally, the data can be visually represented using a word cloud, as shown in Figure 3.



Figure 3. Word Cloud

A Word Cloud is a graphical representation that depicts the relative frequency of words within a given text. The size of each word in the cloud indicates its frequency of occurrence, with larger font sizes representing higher frequencies. Conversely, smaller font sizes are used for words that occur less frequently. In essence, the Word Cloud provides a visual snapshot of the prominence or rarity of words in the text [27]. The word cloud visualization indicates that the words "Chat" and "GPT" have the highest frequency of occurrence compared to other words in the dataset. The next step is the validation process using K-Fold Cross Validation. The model workflow that has been built is presented in Figure 4.

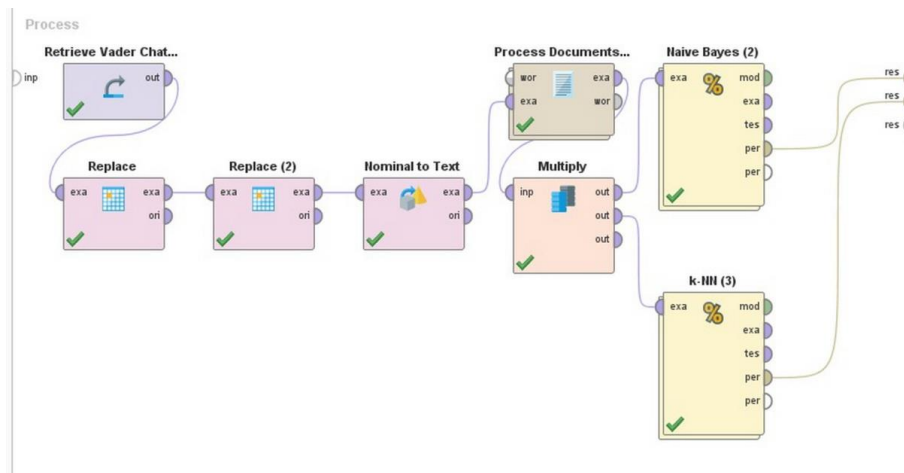


Figure 4. Model Process

The model process was built using the Naïve Bayes and K-NN algorithms. The validation process involved K-Fold Cross Validation with different combinations of k values (k=2, k=4, k=6, k=8, and k=10) and sampling methods (Linear, Shuffled, and Stratified). The selection of the value k to search for the best performance results is aligned and scientifically based on the research that has been conducted regarding “K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining” [43]. The results of the model process are presented in Table 5.

Table 5. Modelling Result

| Algorithm   | k  | Sampling   | Accuracy | Algorithm | k  | Sampling   | Accuracy |
|-------------|----|------------|----------|-----------|----|------------|----------|
| Naïve Bayes | 2  | Linear     | 62.32%   | K-NN      | 2  | Linear     | 90.78%   |
|             | 4  |            | 62.09%   |           | 4  |            | 90.77%   |
|             | 6  |            | 63.32%   |           | 6  |            | 90.75%   |
|             | 8  |            | 63.85%   |           | 8  |            | 90.79%   |
|             | 10 |            | 63.76%   |           | 10 |            | 90.77%   |
|             | 2  | Shuffled   | 60.52%   |           | 2  | Shuffled   | 89.78%   |
|             | 4  |            | 63.92%   |           | 4  |            | 91.59%   |
|             | 6  |            | 64.11%   |           | 6  |            | 91.37%   |
|             | 8  |            | 63.74%   |           | 8  |            | 92.40%   |
|             | 10 |            | 63.92%   |           | 10 |            | 91.99%   |
|             | 2  | Stratified | 63.72%   |           | 2  | Stratified | 91.18%   |
|             | 4  |            | 64.72%   |           | 4  |            | 91.79%   |
|             | 6  |            | 63.74%   |           | 6  |            | 92.40%   |
|             | 8  |            | 64.76%   |           | 8  |            | 91.39%   |
|             | 10 |            | 65.13%   |           | 10 |            | 92.39%   |

Table 5 shows that the best performance is achieved by the K-NN algorithm with an accuracy value of 92.40%. This was obtained through the combination of k=8 using the Shuffled Sampling method and k=6 using the Stratified Sampling method. On the other hand, the Naïve Bayes algorithm achieved an optimal accuracy value of 65.13% through the combination of k=10 using the Stratified Sampling method. The accuracy value of 92.40% obtained by the K-NN algorithm falls under the category of "Excellent Classification," while the accuracy value of 65.13% obtained by the Naïve Bayes algorithm falls under the category of "Poor Classification".

#### 4. CONCLUSION

The results of this study indicate that Opinion Mining on Chat GPT primarily reflects the Neutral o, accounting for 89% of the data. The remaining opinion include 10% Positive and 1%

Negative. In comparing the performance of the Naïve Bayes and K-NN algorithms, the K-NN algorithm achieved the highest accuracy with a value of 92.40%. This demonstrates the successful implementation of the Opinion Mining process, as the 92.40% accuracy of the K-NN algorithm falls within the "Excellent Classification" category, while the accuracy value of 65.13% obtained by the Naïve Bayes algorithm falls under the category of "Poor Classification".

However, there are several limitations in this study. Firstly, the focus was on determining the best algorithm between Naïve Bayes and K-NN, without exploring other algorithms. Secondly, opinion assignment in this study was based on the VADER method, without comparing it to other opinion assignment techniques.

Future research could incorporate other comparative algorithms to identify if there are alternatives with higher accuracy than K-NN for Opinion Mining. Additionally, exploring other opinion assignment methods such as Text Blob or utilizing Chatbot technology could be considered.

## REFERENCES

- [1] X. Deng and Z. Yu, "A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education," *Sustainability*, vol. 15, no. 4, pp. 1–19, Feb. 2023, doi: 10.3390/su15042940.
- [2] V. Taecharungroj, "'What Can ChatGPT Do?' Analyzing Early Reactions to the Innovative AI Chatbot on Twitter," *Big Data Cogn. Comput.*, vol. 7, no. 1, pp. 1–10, 2023, doi: 10.3390/bdcc7010035.
- [3] J. Y. Lee, "Can an artificial intelligence chatbot be the author of a scholarly article?," *J. Educ. Eval. Health Prof.*, vol. 20, no. 1, p. 6, 2023, doi: 10.3352/jeehp.2022.20.6.
- [4] S. Singh, S. Tiwari, and P. K. Yadav, "Chat GPT : Exploring The Capabilities And Limitations Of A Large Language," *Int. J. Eng. Appl. Sci. Technol.*, vol. 7, no. 12, pp. 111–115, 2023.
- [5] 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.
- [6] I. T. Julianto, D. Kurniadi, Y. Septiana, and A. Sutedi, "Alternative Text Pre-Processing using Chat GPT 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>.
- [7] 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.
- [8] 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>.
- [9] C. Messaoudi, Z. Guessoum, and L. ben Romdhane, "A Deep Learning Model for Opinion mining in Twitter Combining Text and Emojis," *Procedia Comput. Sci.*, vol. 207, no. Kes, pp. 2628–2637, 2022, doi: 10.1016/j.procs.2022.09.321.
- [10] 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.
- [11] L. R. Dharmawan, I. Arwani, and D. E. Ratnawati, "Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K- Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 959–965, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j->



- ptiik/article/view/7099.
- [12] E. D. Nurhazizah and I. Puspitasari, "Opinion Mining Fungsi KPI (Key Performance Indikator) Dengan Algoritma Naïve Bayes Clasifier Dan Support Vector Machine (SVM)," *J. Cahaya Mandalika*, vol. 3, no. 2, pp. 290–302, 2023.
- [13] Z. Azhari, L. Efrizoni, W. Agustin, and R. Yanti, "Opinion Mining menggunakan Algoritma Deep Learning untuk Menganalisis Penggunaan Aplikasi Jamsostek Mobile," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 666–678, 2023.
- [14] H. Andriana, S. S. Hilabi, and A. Hananto, "Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter Terhadap KTT G20 di Indonesia," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 1, pp. 60–67, 2023, doi: 10.30865/jurikom.v10i1.5427.
- [15] A. Fauzy, *Metode Sampling*. Tangerang Selatan: Universitas Terbuka, 2019.
- [16] A. A. Abdillah, A. Azwardi, and I. Wahyudi, "Optimasi Linear Sampling dan Information Gain pada Algoritma Decision Tree untuk Diagnosis Penyakit Diabetes," *Multinetics*, vol. 7, no. 1, pp. 21–29, 2021.
- [17] 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.
- [18] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima PIP," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 452–458, 2023.
- [19] M. K. Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023.
- [20] 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.
- [21] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2957–2966, 2018.
- [22] 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.
- [23] 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.
- [24] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.
- [25] 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.
- [26] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMAFIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, vol. 1, no. 1, pp. 299–305, 2021.
- [27] O. Abiola, A. Abayomi-Alli, O. A. Tale, S. Misra, and O. Abayomi-Alli, "Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, 2023, doi: 10.1186/s43067-023-00070-9.

- [28] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 423–431, 2022, doi: 10.32672/jnkti.v5i3.4382.
- [29] M. Murali, B. Duraisamy, and J. Vankara, "Measurement : Sensors Independent component support vector regressive deep learning for sentiment classification," *Meas. Sensors*, vol. 26, no. December 2022, pp. 1–8, 2023, doi: 10.1016/j.measen.2023.100678.
- [30] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Twitter Social Media Sentiment Analysis Against Bitcoin Cryptocurrency Trends Using Rapidminer," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1183–1187, 2022.
- [31] I. T. Julianto, "Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut," *J. Algoritm.*, vol. 19, no. 1, pp. 449–456, 2022, doi: 10.33364/algoritma/v.19-1.1112.
- [32] N. Alnuaimi, A. Almemari, 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.
- [33] 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.
- [34] J. Sangeetha and U. Kumaran, "A hybrid optimization algorithm using BiLSTM structure for sentiment analysis," *Meas. Sensors*, vol. 25, no. December 2022, pp. 1–7, 2023, doi: 10.1016/j.measen.2022.100619.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] Y. Septiana, Y. H. Agustin, M. Nashir Mudzakir, A. Mulyani, D. D. Siti Fatimah, and I. Tri Julianto, "Implementation of Classification Algorithm C4.5 in Determining the Emergency Patient in the Maternity Hospital Queue System," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Feb. 2023, pp. 790–794, doi: 10.1109/ICCoSITE57641.2023.10127842.
- [43] D. Normawati and D. P. Ismi, "K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining," *Signal Image Process. Lett.*, vol. 1, no. 2, pp. 23–35, 2019, doi: 10.31763/simple.v1i2.3.