Jurnal Komputer, Informasi dan Teknologi Vol: 5, No 1, 2025, Page: 1-13

Sentiment Analysis of Indonesian Society Toward the Launch of iPhone 16 Using Naive Bayes, Random Forest, and KNN Algorithms

Christopher Ezra Manurung[1], Hendra Mayatopani[2]

[1,2] Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Pradita

DOI: https://doi.org/10.53697/jkomitek.v5i1.2219
*Correspondence: Christopher Ezra Manurung
Email: christopher.ezra@student.pradita.ac.id

Received: 04-04-2025
Accepted: 13-05-2025
Published: 12-06-2025

Abstract: The development of smartphone technology, especially involving global brands like Apple, always attracts the attention of the world, including Indonesia. Every time Apple launches a new product, the public's response, particularly in Indonesia, often appears in the form of tweets on the social media platform Twitter, now known as X, which can be analyzed to reflect public views. This phenomenon presents an opportunity to understand how products are received in today's market. The dataset used in this study was obtained from tweets or comments from the Indonesian public between October and November 2024. The study found that 51.49% of the tweets fell into the positive sentiment category, 28.15% were neutral, and 20.35% were negative. Accuracy evaluation using three algorithms showed that Random Forest had the highest accuracy at 72.4%, followed by KNN with an accuracy of 66.9%, and Naïve Bayes with an accuracy of 66.3%. The results of this study indicate that the majority of the Indonesian public showed a positive sentiment toward the launch of the iPhone 16, reflecting high enthusiasm for the product. Furthermore, the Random Forest algorithm proved to be more effective in sentiment classification compared to KNN and Naïve Bayes, with higher accuracy.

Keywords: iPhone 16, Tweet, Naive Bayes Classifier, Random Forest, KNN

Introduction

With technological advancements, particularly in the fields of communication and mobile devices, the iPhone, which is one of the flagship products of the multinational technology company Apple, has become a symbol of innovation and quality in the smartphone industry. Since its initial launch in 2007, the iPhone has continued to evolve rapidly, with each new model offering advanced features that have impacted the global market (Chen et al., 2021). The iPhone 16, the latest model to be launched in 2024, is expected to continue this innovation trend with updates in design, performance, and AI technology (Farhi et al., 2023). The launch of the iPhone 16 is anticipated to be an important addition to Apple's long list of innovations that consistently capture the attention of consumers, both loyal and new. Apple Inc. is known for its innovative approach to designing and marketing technology products. The company always emphasizes quality and user experience, reflected in its minimalist product design while incorporating advanced features (Simatupang et al., 2023). Every new iPhone model comes with high expectations from consumers, given Apple's reputation for delivering devices with cutting-edge technology. Therefore, it is essential to understand public sentiment toward this new product. Among

the various tech products that often capture the public's attention, the iPhone from Apple is always a focal point, particularly after the release of its latest model. Consumer behavior, especially in responding to information provided before the launch, plays a crucial role in the success of the product's marketing (Bourequat et al., 2021).

Understanding the public's reaction to a particular product can be a determining factor in its success. These responses can be positive, negative, or neutral. Twitter, now known as "X," is one of the social media platforms in Indonesia used by various segments of the population to interact and express their opinions. One form of interaction is providing criticism or comments related to the quality of products. Twitter users have the freedom to express their views or opinions about specific goods and services (Giovani et al., 2020). To measure public sentiment regarding the launch of the iPhone 16, sentiment analysis was conducted to determine whether the opinions are positive, negative, or neutral. Natural Language Processing (NLP) is a branch of computer science that enables machines to understand, analyze, and process human language, allowing computers to recognize meaning, context, and relationships between words in text or speech (Fauzianto et al., 2023). Before testing the algorithms, the data preprocessing stage was conducted to prepare the data for analysis.

Text preprocessing is a crucial step in data classification, aiming to clean and transform raw data, including non-alphabetic characters and irrelevant words. This step ensures that the data becomes more efficient when input into the classification process (Adepu et al., 2023). The text preprocessing stages include case folding, stopword removal, word normalization, stemming, tokenization, and weighting (Kartika et al., 2023). Various classification algorithms can be used to analyze text-based data. In this study, three algorithms were selected for testing: Naïve Bayes Classifier, Random Forest, and K-Nearest Neighbor (KNN). The Naïve Bayes Classifier is a probabilistic algorithm known for its simplicity and efficiency in processing text data, with an accuracy of up to 74% (Wulandari et al., 2023). The primary advantage of Naïve Bayes is its computational speed and ability to handle large datasets efficiently. In contrast, Random Forest is an ensemble method that combines multiple decision trees to improve accuracy, achieving an accuracy of 86.23% (Tantyoko et al., 2023). KNN is an instance-based algorithm that classifies data based on its proximity to other data points, with an accuracy of up to 85% (Dzulhijjah et al., 2021). This approach is particularly useful when the data structure is unclear and can generate flexible models with good accuracy on labeled data.

Through this study, it is expected to gain insight into the public's response to the launch of the iPhone 16, whether it tends to be positive, negative, or neutral. Additionally, this study compares the three algorithms Naïve Bayes Classifier, Random Forest, and K-Nearest Neighbor to determine the algorithm with the highest accuracy in assessing sentiment related to the launch of the iPhone 16.

Methodology

This research was conducted through several steps. The first step was data collection or data crawling from the social media platform Twitter (X). Then, in the second step, a data cleaning process was performed. After the data was cleaned, the third step was data

labeling. In the fourth step, the labeled data was further processed through text preprocessing, which included case folding, tokenizing, stopword removal, stemming, and weighting using the TF-IDF (Term Frequency-Inverse Document Frequency) method. The processed data then moved to the fifth step, which involved classification using the Naïve Bayes Classifier, Random Forest, and KNN algorithms. After applying the algorithms, the next step was validation to measure the accuracy and precision of the three algorithms. As the final step, the model was validated using the K-Fold Cross Validation method.

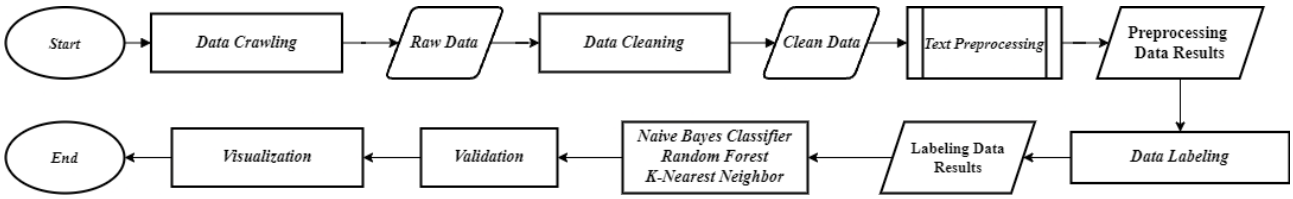

Figure 1. Research Method Flowchart

Crawling or data collection is the first stage in sentiment analysis of a topic. The data collection process from Twitter (X) is carried out using tools available in Google Colab and the Twitter API to obtain tokens that allow tweet data retrieval. Once the data is successfully collected, it will be stored in CSV format (Aliyah et al., 2024). Raw data obtained from the crawling process often contains various irrelevant elements, such as duplicates, mentions, hashtags, emojis, URLs, and other symbols that are not useful for the research. Therefore, a data cleaning process is necessary to ensure that the data used is clean and ready for further analysis. This process aims to remove elements that may interfere with the analysis and ensure that the data used is of good quality (Atmaja et al., 2021).

Text preprocessing is an important step in sentiment analysis that aims to clean and prepare text data to make it more structured and ready for further analysis. With proper preprocessing, data quality can be improved, resulting in more accurate sentiment analysis (Hakim, 2021). The following are the steps in the data preprocessing process:

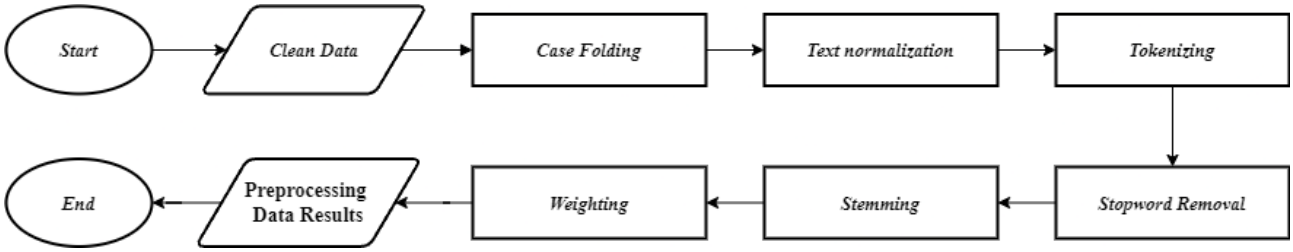

Figure 2. Flowchart of Text Preprocessing

1. **Case Folding**: This process converts all words in the text into lowercase or uppercase letters to ensure consistency and avoid unnecessary differences that may arise from variations in letter case.
2. **Word Normalization**: This process corrects or converts non-standard or abbreviated forms of words into a more proper form according to the rules of the Indonesian language. The goal of this process is to ensure that the information generated can be

clearly understood and easily processed, especially in the context of text analysis and natural language processing (NLP).

3. **Tokenizing**: This process breaks the text into smaller units such as words, phrases, or sentences. The goal is to separate elements in the text and remove irrelevant characters, such as punctuation, spaces, or numbers.
4. **Stopword Removal**: This process aims to remove common words that do not provide important information, such as "and," "or," and "is." By eliminating these words, the analysis can focus more on the words that are more meaningful and relevant in the context of sentiment analysis.
5. **Stemming**: This process converts words to their root form by removing affixes. For example, the word "berlari" (to run) is simplified to "lari" (run). This process helps to simplify the data and reduce variations of identical words.
6. **Weighting**: This process assigns weights to words in the text to enhance the accuracy of sentiment classification. Each word is given a different weight value based on its frequency in the document or its relevance to a specific sentiment category (Rifaldi et al., 2023).

Once the data is cleaned, the next step is to manually label the data, which will be used as training data. This labeled data is then used in the Naïve Bayes Classifier, Random Forest, and KNN algorithms to build a model that will classify sentiment on the test data. The labeling process consists of three categories: positive, negative, and neutral. This process is crucial to ensure that the model can classify sentiment accurately based on the trained data (Fitriyah et al., 2020). Naïve Bayes Classifier is a classification algorithm based on Bayes' theorem, with the assumption that the features in the data are independent of each other. This algorithm estimates the probability of a class based on the features in the dataset (Prayogo et al., 2023). Below is the formula for Naïve Bayes Classifier:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

Random Forest is a machine learning algorithm that combines multiple decision trees to improve accuracy in classification and regression processes (Hidayah et al., 2024). This method uses random samples from the training data and randomly selected features, which helps reduce the risk of overfitting and enhances the stability of predictions. Random Forest is highly effective in handling large and complex data and is tolerant to outliers.

$$fi_j = \frac{\Sigma\ j{:}node\ j\ splits\ on\ feature\ i\ ni_j}{\Sigma\ k \in all\ nodes\ ni_k} \quad (2)$$

K-Nearest Neighbor is an algorithm that classifies new data by taking the majority vote of the nearest neighbors' classes or calculating the average value for regression. This algorithm is known for its simplicity and its ability to handle unstructured data (Cholil et al., 2021).

$$d_i = \sqrt{\sum_{i=1}^{n} \quad (X_i - Y_i)^2} \quad (3)$$

To ensure optimal performance of the Naïve Bayes Classifier, Random Forest, and K-Nearest Neighbor algorithms, an efficient validation method is required to measure the accuracy of these models. One technique used is K-Fold Cross Validation, which divides the dataset into 'K' equal parts, with each part being used as test data in different iterations. This process is repeated 'K' times, with training and testing alternating on each subset of data, providing a more accurate estimate of the model's performance (Nugroho et al., 2023).

Result and Discussion

This study collected a total of 2,524 tweets between October and November 2024 using the keywords "iPhone 16" or "iPhone 16 Indonesia." The data was obtained through a crawling technique utilizing tools from Google Colab, which used the Python programming language and the pandas library to execute data retrieval commands. Data collection was done randomly on various dates during the period to capture a diverse range of opinions from Twitter users in Indonesia. Only tweets written in Indonesian were included to ensure that the opinion analysis came from the Indonesian public. The raw data contained various irrelevant elements such as URLs, user IDs, usernames, hashtags, @mentions, emojis, IDs, dates, and other symbols. The results of the data retrieval can be seen in Table 1.

Table 1. Data Crawling Result

Number	Teks
1.	1862646212199137391,"Fri Nov 29 23:53:57 +0000 2024","0","Guys ini iphone 15 pro / pro max udh gaada di ibox atau digimap ya????? Kalau mau beli dimana dong huhu org iphone 16 masih belom jelas gini ","1862646212199137391","","","in","Republic of Korea","0","1","0","https://x.com/cherryaraa/status/1862646212199137391" ,"1010479261159809024","cherryaraa"

The raw data collected through the crawling process still contains various elements that need to be removed, such as URLs, user IDs, usernames, hashtags, @mentions, emojis, IDs, dates, symbols, and duplicate data that are irrelevant to this research. To clean the data, RapidMiner tools were used. From the total of 2524 tweets collected, after cleaning, only 1808 tweets remained. The data cleaning process also involved manual checks because some tweets were found to be in languages other than Indonesian. The cleaned data results can be seen below.

Table 2. Data Cleaning Results

Number	Teks
1.	Guys ini iphone 15 pro / pro max udh gaada di ibox atau digimap ya Kalau mau beli dimana dong huhu org iphone 16 masih belom jelas gini
2.	Imbas Larangan Jual iPhone 16 di Indonesia Harga Baru iPhone Turun Drastis

After the data cleaning is completed, the next step is to preprocess the data. In this stage, the tweets are transformed into a more standardized, simple, and common language, and words that do not have clear meanings are removed. The data preprocessing process involves several important stages, namely Case Folding, Word Normalization, Tokenizing, Stopword Removal, Stemming, and Weighting. These processes can be seen in the following tables.

Table 3. Text Preprocessing

Teks	Preprocessing
Imbas Larangan Jual iPhone 16 di Indonesia Harga Baru iPhone Turun Drastis	Before Preprocessing
imbas larangan jual iphone di indonesia harga baru iphone turun drastis	*Case Folding*
imbas larangan jual iphone di indonesia harga baru iphone turun drastis	*Word Normalization*
['imbas', 'larangan', 'jual', 'iphone', 'di', 'indonesia', 'harga', 'baru', 'iphone', 'turun', 'drastis']	*Tokenizing*
['imbas', 'larangan', 'jual', 'iphone', 'indonesia', 'harga', 'iphone', 'turun', 'drastis']	*Stopword Removal*
['imbas', 'larang', 'jual', 'iphone', 'indonesia', 'harga', 'iphone', 'turun', 'drastis']	*Stemming*

The next step is to apply the TF-IDF (Term Frequency-Inverse Document Frequency) method to convert the text into a numerical form. This process is crucial because it enables machine learning algorithms to analyze and understand the text data more effectively. The dataset that has undergone preprocessing does not yet have classification labels, so labeling is required. The data is manually labeled and used for testing with the Naïve Bayes, Random Forest, and K-Nearest Neighbor classification algorithms. This dataset consists of 1808 tweets. The labeling process is carried out based on three categories to determine the sentiment of the Indonesian public toward the iPhone 16. Positive sentiment indicates interest in purchasing the iPhone 16, while negative sentiment reflects rejection, criticism, or disagreement with the iPhone 16. Neutral sentiment reflects an impartial attitude or ambiguity toward the policy. Examples of labeling and the classification result percentages can be seen in the following table and figure:

Table 4. Data Labeling

Number	Teks	Sentimen
1.	iPhone masuk Indonesia	Positive
2.	saing iPhone China bos Huawei pamer Mate Pro	Neutral
3.	musnah iPhone bro	Negative

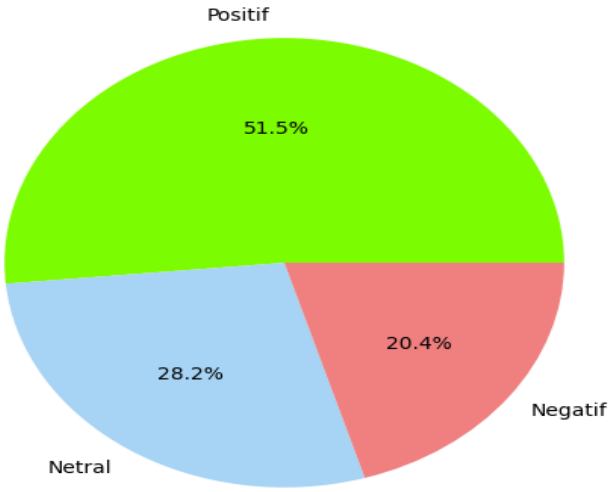

Figure 3. Percentage of Sentiment Classification Results

The results of the tweet data labeling show that 931 tweets fall into the positive sentiment category, accounting for 51.49%, 509 tweets are classified as neutral sentiment with a percentage of 28.15%, and 368 tweets are categorized as negative sentiment, making up 20.35% of the total data analyzed.

To ensure the accuracy of the Naïve Bayes algorithm in classifying the data, validation was carried out using the K-Fold Cross Validation method. The data was divided into several balanced folds, with each fold alternately used as test data, while the remaining folds were used as training data. This approach allows all the data to be used for both training and testing, making the performance evaluation of the model more accurate. Validation was conducted with variations of K ranging from 1 to 10 to measure the stability of the model against different data splits. The validation results of the Naïve Bayes algorithm are shown in the following table 5:

Table 5. Validation Results of the Naïve Bayes Algorithm

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	59,1	59,4	59,1	59,3
2	58	59,6	58	58,3
3	58,6	60,2	58,6	58,9
4	62,4	63,7	62,4	62,8
5	59,7	59,2	59,7	59,4
6	66,3	66,8	66,3	66,3
7	63	62,9	63	62,8
8	61,3	61,6	61,3	61,3
9	57,8	59	57,8	58,2
10	60	63,5	60	61,2
Average	60,62	61,59	60,62	60,85

The validation results indicate that the performance of the Naïve Bayes algorithm varies depending on the applied K value. The average accuracy of the Naïve Bayes algorithm in this study is 60.62%. The highest accuracy was recorded at K=6, where the model achieved an accuracy level of 66.30%. To ensure the accuracy of the Random Forest algorithm in classifying data, validation was conducted using the K-Fold Cross-Validation method. For the Random Forest algorithm, the model was configured with 100 trees. Validation was performed with K values ranging from 1 to 10. The validation results for the Random Forest algorithm can be seen in Table 6 below:

Table 6. Validation Results of the Random Forest Algorithm

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	64,6	67,5	64,6	60,8
2	66,3	63,4	66,3	63,6
3	69,6	70	69,6	66,7
4	59,7	54,7	59,7	53,8
5	60,8	55,7	60,8	55,8
6	70,7	72,5	70,7	66,7
7	72,4	73,6	72,4	69,1
8	56,9	58,6	56,9	52,1
9	64,4	61,2	64,4	60,4
10	62,2	59,9	62,2	59,2
Average	64,76	63,71	64,76	60,82

The validation results show that the performance of the Random Forest algorithm varies depending on the applied K value. The average accuracy of the Random Forest algorithm in this study is 64.76%. The highest accuracy was recorded at K=7, where the model achieved an accuracy of 72.4%. To ensure the accuracy of the K-Nearest Neighbors (KNN) algorithm in data classification, validation was conducted using the K-Fold Cross-Validation method. For the KNN algorithm, the model was configured with the number of neighbors set to 5. Validation was performed with K values ranging from 1 to 10. The validation results for the KNN algorithm can be seen in Table 7 below:

Table 7. Validation Results of the KNN Algorithm

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	62,4	60,5	62,4	60,9
2	62,4	61	62,4	61,5
3	61,9	62,1	61,9	62
4	63	61,5	63	62,1
5	56,4	55,9	56,4	56,1
6	66,9	65,7	66,9	66,1
7	60,2	58,1	60,2	58,7
8	58	55,5	58	55,5
9	62,2	62,1	62,2	62

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
10	62,8	62,3	62,8	62,5
Average	61,62	60,47	61,62	60,74

The validation results show that the performance of the KNN algorithm varies depending on the applied K value. The average accuracy of the KNN algorithm in this study is 61.62%, with the highest accuracy recorded at K=6, where the model achieved an accuracy of 66.9%.

Word visualization was performed for each positive, negative, and neutral sentiment using word cloud to identify the most frequently occurring words in each sentiment category. The goal is to understand the topics or issues that are frequently discussed by the public regarding the launch of the iPhone 16 in Indonesia. The results of this visualization are divided into three parts: positive data visualization, negative data visualization, and neutral data visualization. Each visualization provides an overview of the dominant words that appear in each sentiment category.

Based on the visualization of positive sentiment in Figure 4, the words that often appear include “iphone” 1028 times, “beli” 179 times, “pakai” 76 times, “kasih” 61 times, and “terima” 49 times. Words such as “beli”, “pakai”, “kasih”, and “terima” reflect the public’s enthusiasm for having the iPhone 16 since its launch.

Figure 4. Positive Sentiment Word Cloud

Based on the negative sentiment visualization in Figure 5, the words that often appear include “iphone” 444 times, “larang” 50 times, “blokir” 42 times, and “tolak” 18 times. Words such as “larang”, “blokir”, and “tolak” reflect the public’s rejection and objection to the presence of the iPhone 16 product in Indonesia.

Figure 5. Negative Sentiment Word Cloud

Based on the visualization of neutral sentiment in Figure 6, the words that frequently appear include "iphone" 617 times, "kemenperin" 58 times, "pemerintah" 53 times, "imei" 49 times, and "tkdn" 48 times. Words such as “kemenperin”, “pemerinta”, “imei”, and “tkdn” reflects a more factual or informative discussion regarding government regulations and policies regarding the launch of the iPhone 16. This shows that people discuss administrative and technical aspects, such as IMEI rules and TKDN requirements.

Figure 6. Neutral Sentiment Word Cloud

Discussion

In the testing of the three algorithms, the results showed that the Random Forest algorithm recorded the highest accuracy of 72.4% at K=7, with an average accuracy of 64.76%. The KNN algorithm ranked second with the highest accuracy of 66.9% at K=6 and an average accuracy of 61.62%. Meanwhile, the Naïve Bayes algorithm achieved the highest accuracy of 66.3% at K=6, with an average accuracy of 60.62%. Based on these results, it can be concluded that the Random Forest algorithm has the best accuracy among the three algorithms tested.

In a previous study conducted by Sherin et al. (2019) on sentiment analysis of iPhone products, the performance of several classification algorithms was evaluated. The sentiment prediction accuracy using the Random Forest algorithm reached 93.25%, while the Naïve Bayes algorithm achieved an accuracy of 40.5%. The findings of this study align with previous results, where the Random Forest algorithm also demonstrated superior performance in data classification compared to other algorithms.

The frequency of keywords that often appear in positive sentiments include “beli”, “pakai”, “kasih”, and “terima”, while in negative sentiments there are words such as “larang”, “blokir”, and “tolak”. While in neutral sentiments, the words that often appear are “kemenperin”, “pemerintah”, “imei”, and “tkdn”. This shows that these words are expressions that are frequently used by Twitter or X users to convey their opinions in various sentiment contexts.

Conclusion

This study collected 2,524 tweets, which, after cleaning, resulted in 1,808 tweets. Based on sentiment labeling, 51.49% of the tweets were positive, 28.15% were neutral, and 20.35% were negative. This shows that the majority of people welcomed the launch of the iPhone 16 with enthusiasm and appreciation. This response reflects the dominance of positive views and strong support for the product on social media. Accuracy testing using the Random Forest, KNN, and Naïve Bayes algorithms showed that Random Forest achieved the highest accuracy (72.4%), followed by KNN (66.9%) and Naïve Bayes (66.3%). These results indicate that the majority of the Indonesian public has a positive sentiment toward the launch of the iPhone 16, reflecting their enthusiasm and expectations for the product. This study demonstrates that the Random Forest algorithm is more effective in classifying sentiment compared to the other two algorithms.

References

Adepu Rajesh, M. R., & Hiwarkar, D. T. (2023). Exploring preprocessing techniques for natural language text: A comprehensive study using Python code. *International Journal of Engineering Technology and Management Sciences, 7*(5), 390-399.

Aliyah, N. A., Azzahra, N. N., Putri, N. A. I., & Rakhmawati, N. N. A. (2024). Analisis sentimen Twitter terhadap tren penyebaran informasi pelaku kejahatan menggunakan algoritma Naives Bayes. *Bridge, 2*(2), 85-97. https://doi.org/10.62951/bridge.v2i2.63

Atmaja, R. M. R. W. P. K., & Yustanti, W. (2021). Analisis sentimen customer review aplikasi ruang guru dengan metode BERT (Bidirectional Encoder Representations from Transformers). *Analisis Sentimen Customer Review Aplikasi Ruang Guru Dengan Metode BERT (Bidirectional Encoder Representations From Transformers)*, *2*(3).

Bourequat, W., & Mourad, H. (2021). Sentiment analysis approach for analyzing iPhone release using support vector machine. *International Journal of Advances in Data and Information Systems*, *2*(1), 36–44. https://doi.org/10.25008/ijadis.v2i1.1216

Chen, X., Liu, Y., & Gong, H. (2021). Apple Inc. Strategic Marketing Analysis and Evaluation. *Advances in Economics, Business and Management Research/Advances in Economics, Business and Management Research*. https://doi.org/10.2991/assehr.k.211209.499

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) untuk klasifikasi seleksi penerima beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2). https://doi.org/10.31294/ijcit.v6i2.10438

Dzulhijjah, A. N., & Anraeni, S. (2021). Klasifikasi kematangan citra labu siam menggunakan metode KNN (K-Nearest Neighbor) dengan ekstraksi fitur HSV (Hue, Saturation, Value). *Buletin Sistem Informasi dan Teknologi Islam*, *2*(2), 103–110.

Farhi, F., Jeljeli, R., Zahra, A., Saidani, S., & Feguiri, L. (2023). Factors behind virtual assistance usage among iPhone users: Theory of reasoned action. *International Journal of Interactive Mobile Technologies*, *17*(2), 42–61.

Fauzianto, R. A., & Supatman. (2023). Analisis sentimen opini masyarakat terhadap tech winter pada Twitter menggunakan natural language processing. *Jurnal Syntax Admiration*, *3*(9), 1577–1585. https://doi.org/10.46799/jsa.v3i9.909

Fitriyah, N., Warsito, B., & Maruddani, D. A. I. (2020). Analisis sentimen GoJek pada media sosial Twitter dengan klasifikasi support vector machine (SVM). *Jurnal Gaussian*, *9*(3), 376–390. https://doi.org/10.14710/j.gauss.v9i3.28932

Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis sentimen aplikasi Ruang Guru di Twitter menggunakan algoritma klasifikasi. *Jurnal Teknoinfo*, *14*(2), 115. https://doi.org/10.33365/jti.v14i2.679

Hakim, B. (2021). Analisa sentimen data text preprocessing pada data mining dengan menggunakan machine learning. *JBASE - Journal of Business and Audit Information Systems*, *4*(2). https://doi.org/10.30813/jbase.v4i2.3000

Hidayah, L., & Rosadi, M. I. (2024). Penerapan algoritma Random Forest untuk memprediksi jumlah santri baru. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1), 5237. https://doi.org/10.23960/jitet.v12i3s1.5237

Kartika, L. G. S., Utama, P. K. L., Budiastawa, I. D. G., & Rinartha, K. (2023). Comparison of the sentiment analysis model's code complexity and processing time. *Sinkron*, *8*(1), 109–118. https://doi.org/10.33395/sinkron.v8i1.11894

Nugroho, N. A., & Amrullah, N. A. (2023). EVALUASI KINERJA ALGORITMA K-NN MENGGUNAKAN K-FOLD CROSS VALIDATION PADA DATA DEBITUR KSP

GALIH MANUNGGAL. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, *5*(2), 294–300. https://doi.org/10.51401/jinteks.v5i2.2506

Prayogo, A., Fauziah, F., & Winarsih, W. (2023). PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI JUDUL ARTIKEL PADA JURNAL ILMIAH. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *8*(4), 1327–1338. https://doi.org/10.29100/jipi.v8i4.4141

Rifaldi, D., Fadlil, N. A., & Herman, N. (2023). Teknik preprocessing pada text mining menggunakan data tweet 'Mental Health.' *Decode Jurnal Pendidikan Teknologi Informasi, 3*(2), 161–171. https://doi.org/10.51454/decode.v3i2.131

Sherin, M. J., & Kartheeban, K. (2019). Sentiment scoring and performance metrics Examination of various supervised classifiers. *International Journal of Innovative Technology and Exploring Engineering*, *9*(2S2), 1120–1126. https://doi.org/10.35940/ijitee.b1111.1292s219

Simatupang, M. S., & Purba, H. (2023). THE IMPACT SOCIAL MEDIA MARKETING, SOCIAL INTERACTIVITY AND PERCEIVED QUALITY OF BRAND LOYALTY ON IPHONE USERS. *Jurnal Riset Manajemen Dan Akuntansi*, *3*(1), 125–136.

Tantyoko, H., Sari, D. K., & Wijaya, A. R. (2023). Prediksi potensial gempa bumi Indonesia menggunakan metode random forest dan feature selection. *Indonesia Journal Information System, 6*(2), 83-89.

Wulandari, N. F., Haerani, N. E., Fikry, N. M., & Budianita, N. E. (2023). Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, *4*(1), 88–96.