E-ISSN: xxxx-xxxx
DOI: https://doi.org/10.64803/cessmuds.v1i1.37
Novelan and Iqbal

Comparative Analysis of the Performance of K-Nearest Neighbor (K-NN) and Naive Bayes Algorithms on User Satisfaction Levels of the Tokopedia Application

Muhammad Syahputra Novelan[1*], Muhammad Iqbal[2]

[1,2]Magister of Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
E-mail: [1]*putranovelan@dosen.pancabudi.ac.id, [2]muhammadiqbal@dosen.pancabudi.ac.id
*E-mail Corresponding Author: putranovelan@dosen.pancabudi.ac.id

Abstract

The rapid advancement of digital technology has significantly influenced the development of e-commerce platforms in Indonesia, where Tokopedia stands out as one of the most popular and widely used online marketplaces. As user expectations continue to increase, understanding and measuring user satisfaction has become essential for ensuring service quality and maintaining customer loyalty. This study aims to perform a comparative analysis of the performance of two machine learning classification algorithms—K-Nearest Neighbor (K-NN) and Naive Bayes—in analyzing and predicting user satisfaction levels toward the Tokopedia application. The dataset used in this study was obtained from a combination of online reviews and structured survey responses from active Tokopedia users. The research methodology includes several stages: data collection, text preprocessing (tokenization, stop-word removal, and stemming), feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) technique, and model implementation using the two algorithms. Both models were evaluated using key performance metrics such as accuracy, precision, recall, and F1-score. The experimental results indicate that the K-NN algorithm achieved superior performance compared to Naive Bayes, demonstrating higher accuracy and better consistency in classifying user sentiments into "satisfied" and "dissatisfied" categories. The K-NN model proved to be more effective in handling diverse and nonlinear data patterns derived from user-generated reviews. Meanwhile, Naive Bayes, although computationally efficient, showed limitations in processing complex text dependencies. The findings of this research highlight the importance of selecting appropriate machine learning algorithms for user satisfaction analysis. Furthermore, the study contributes to the broader understanding of sentiment-based evaluation models in e-commerce platforms and provides valuable insights for Tokopedia and similar companies in enhancing customer experience and service improvement strategies.

Keywords: *K-Nearest Neighbor; Naive Bayes; Tokopedia; Sentiment Analysis; Machine Learning*

I. INTRODUCTION

The digital era has significantly transformed the landscape of commerce through the emergence of e-commerce platforms that enable consumers to make transactions anytime and anywhere. In Indonesia, Tokopedia stands out as one of the pioneering marketplaces facilitating millions of transactions daily. As this growth intensifies, users are becoming increasingly critical of service quality, user experience (UX), and their overall satisfaction with the application. Research examining UI/UX elements and user satisfaction on Tokopedia indicates that features such as ease of use, interface design, and consumer loyalty are interrelated (Talmera et al., 2025). Therefore, understanding user satisfaction with e-commerce applications has become a strategic priority for application developers and marketplace managers.(Dhany et al., 2024)

One rich source of data for measuring user satisfaction is the online user reviews. These reviews contain opinions, criticisms, and expressions of satisfaction or dissatisfaction which, when properly processed, can generate valuable insights for the company. Sentiment analysis has emerged as a popular method to convert review texts into categories such as "positive" or "negative" (Alamsyah & Saviera, 2021) As an example, a study analyzing reviews of the Tokopedia application using machine learning techniques shows that sentiment analysis of reviews can systematically reveal user perceptions. Thus, leveraging user review data to evaluate satisfaction becomes a highly relevant approach for this study.(Khairul et al., 2023)

In the realm of machine learning for text analysis, classification algorithms such as K-Nearest Neighbor (K-NN) and Naive Bayes are frequently employed due to their speed and ease of implementation. Several studies have compared the performance of these two algorithms in sentiment classification. For example, some research on application reviews (not specific to Tokopedia) found that one algorithm may outperform the other depending on the data characteristics (Putera Utama Siahaan et al., 2025). Therefore, it is imperative to conduct a comparative performance evaluation of these algorithms in the specific context of Tokopedia's user reviews to determine the most suitable method.(Novelan & Aryza, 2025)

Although promising, challenges remain in applying classification algorithms for e-commerce application reviews in Indonesia: review data are in the Indonesian language, user expression varies widely, and class imbalance between positive and negative reviews often exists. Research optimizing sentiment analysis on Tokopedia reviews using methods like SMOTE to handle class imbalance shows that algorithm selection and data handling are highly significant. In turn, comparative studies of K-NN and Naive Bayes on Indonesian e-commerce reviews have shown varied results—highlighting the need for a structured and context-specific comparative study in the Tokopedia context.(Novelan et al., 2023)(Von Rueden et al., 2023)

With this background, the present study is designed to perform a comparative analysis of the performance of Naive Bayes and K-NN in predicting user satisfaction levels for the Tokopedia application based on user review data. By choosing a specific context, Indonesian-language data, and real user feedback from Tokopedia, this study is expected to contribute both practically—by recommending which algorithm is more effective in this context—and academically—by enriching the literature on sentiment classification and e-commerce user satisfaction(Iqbal & Efendi, 2023). Moreover, the results are anticipated to assist application developers and marketplace management in developing effective user-feedback systems based on machine learning.(Nuranisah et al., 2020)

II. RESEARCH METHODOLOGY

a) Literature Review

Many studies have compared the performance of the K-Nearest Neighbor (KNN) and Naive Bayes algorithms in text classification and sentiment analysis. These studies generally reveal that Naive Bayes often outperforms in terms of accuracy and processing efficiency, particularly when applied to Indonesian-language datasets with a large number of reviews (Alamsyah & Saviera, 2021; Saputra et al., 2023). However, several studies have also shown that KNN can yield better performance under certain conditions, such as when applied to review analysis of specific applications like Vidio or Shopee (Talmera et al., 2024).

Several factors influence the performance of both algorithms, including text preprocessing steps (case folding, tokenization, stopword removal, stemming), feature extraction techniques such as Term Frequency–Inverse Document Frequency (TF-IDF), dataset size and balance, as well as model complexity (Wibowo et al., 2022). Naive Bayes tends to perform better on large and diverse datasets, while KNN can be more effective when vector representations are well-structured and class distributions are balanced.(Sitorus et al., 2024)

In the context of Indonesian e-commerce reviews such as those from Shopee and Tokopedia, prior studies have yielded inconsistent results—some report higher accuracy for Naive Bayes, while others find that KNN performs better depending on the dataset and parameter tuning (Gunawan & Siregar, 2022; Pratama et al., 2023). This indicates that algorithm selection heavily depends on the dataset characteristics and the applied parameterization. Therefore, the present study comparing both algorithms on Tokopedia user reviews is relevant for providing empirical insights into the effectiveness of text classification methods in Indonesian sentiment analysis.(Farta Wijaya et al., 2024)

Table 1. **Literature Review**

Journal References	**Research Focus**	**Dataset and Context**	**Key Findings**
(Hamid et al., 2025)– "WordPress Application Satisfaction Level Sentiment Analysis Using KNN and Naive Bayes"	WordPress application user satisfaction analysis	5,000 Play Store reviews	Naive Bayes is better: accuracy ~88%
Abdillah et al. (2024) – "Komparasi Metode Naive Bayes dan K-Nearest Neighbors terhadap Analisis Sentimen Pengguna Aplikasi Zenius"	Comparison of NB and KNN in educational applications	Zenius application data	Comparative study; detailed figures not yet mentioned in the summary
Wijianto et al. (2024) – "Komparasi KNN dan Naive Bayes pada Klasifikasi Sentimen Ulasan Aplikasi Tokopedia"	Tokopedia review classification	Tokopedia Google Play Store Review	KNN is superior (accuracy ~80%), NB ~71%
Faradaningsih & Lutfiyani (2024) – "Perbandingan Analisis	Shopee & Lazada e-commerce app review	40,000 reviews (20,000 per app)	NB is slightly better on Shopee, KNN is

E-ISSN: xxxx-xxxx
DOI: https://doi.org/10.64803/cessmuds.v1i1.37

Sentimen Pengguna Aplikasi Shopee dan Lazada			slightly better on Lazada
(Wulandari et al., 2025)– “Sentiment Analysis on the Relocation of the National Capital (IKN) on Social Media X Using NB and KNN”	Analysis of public policy sentiment on social media	1.277 tweet	NB is superior (accuracy ~70% vs 60%)
(Tapidingan & Paseru, 2020)– “Comparative Analysis of Classification Methods of KNN and NB to Determine Stress Level of Junior High School Students”	Classification of student stress levels (non-ecommerce)	254 secondary school respondents	NB is superior (accuracy ~87.40% vs KNN ~86.61%)
(Kadek et al., 2025)– “Evaluasi NB dan KNN dalam Klasifikasi Sentimen Ulasan Produk Skincare di Tokopedia”	Skincare product reviews on Tokopedia	475 skincare product reviews	NB is superior (accuracy ~92.16% vs KNN ~68.82%)

b) Method

This research method outlines the steps undertaken to analyze the comparative performance of the K-Nearest Neighbor (K-NN) and Naive Bayes algorithms in assessing user satisfaction levels with the Tokopedia application. The study employs a quantitative approach using a comparative experimental method, in which both algorithms are tested on a dataset of user reviews to determine which algorithm performs best in classifying user satisfaction sentiments. The research stages include data collection, text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), implementation of the K-NN and Naive Bayes models, and performance evaluation using the metrics of accuracy, precision, recall, and F1-score.(Hasan Putra et al., n.d.)(Genkin, 2020)

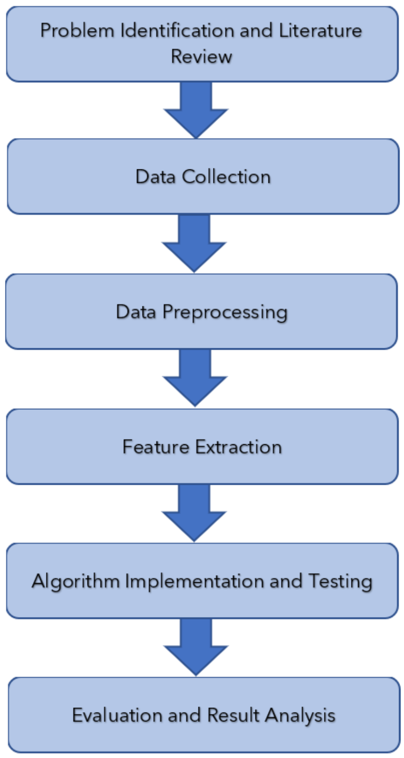

Figure 1. **Research Stages**

1. Problem Identification and Literature Review
   The first step begins with identifying issues related to user satisfaction with the Tokopedia application. Subsequently, a literature review is conducted using various journals, books, and previous studies discussing the K-Nearest Neighbor (K-NN) and Naive Bayes algorithms, as well as sentiment analysis, to strengthen the theoretical foundation and understanding of the research.
2. Data Collection
   Data are collected from user reviews of the Tokopedia application obtained through web scraping techniques from public platforms such as Google Play Store, as well as additional surveys using Google Forms. The collected data consist of textual content (user comments) expressing satisfaction or dissatisfaction with the application.
3. Data Preprocessing
   The collected review data need to be cleaned before analysis. This stage includes several processes: case folding (converting all letters to lowercase), tokenizing (splitting sentences into words), stopword removal (removing common words that carry little meaning), and stemming (returning words to their root form). These processes help structure the data for further analysis.
4. Feature Extraction
   This stage aims to convert textual data into numerical form so it can be processed by machine learning algorithms. The method used is Term Frequency–Inverse Document Frequency

(TF-IDF), which calculates the weight of each word based on its frequency of occurrence in the document.

5. Algorithm Implementation and Testing
At this stage, two classification algorithms—K-Nearest Neighbor (K-NN) and Naive Bayes—are implemented using the preprocessed dataset. The data are divided into two parts: 80% for training and 20% for testing. Both algorithms are then used to classify reviews into two categories: “satisfied” and “not satisfied.”
6. Evaluation and Result Analysis
The final step involves evaluating the performance of both algorithms using testing metrics such as Accuracy, Precision, Recall, and F1-Score. The performance comparison results are then analyzed to determine which algorithm performs best in predicting user satisfaction with the Tokopedia application. This analysis serves as the basis for drawing conclusions and providing recommendations for the study.

III. RESULTS AND DISCUSSION

The user review data for the Tokopedia application obtained through web scraping consists of 999 entries. After data cleaning, 822 entries were deemed valid for analysis. Based on the classification results, there were 678 positive reviews (82.5%) and 144 negative reviews (17.5%). This imbalance indicates that most users have a positive perception of the Tokopedia application, particularly in terms of ease of use, delivery services, and product quality. During the preprocessing stage, several steps were carried out to prepare the text data for machine learning analysis. These steps included case folding (converting all letters to lowercase), tokenizing, stopword removal, and stemming using Python libraries. This process ensures that each review is represented in meaningful root words and ready to be transformed into numerical features through the TF-IDF (Term Frequency–Inverse Document Frequency) method. The following section presents the Tokopedia reviews.

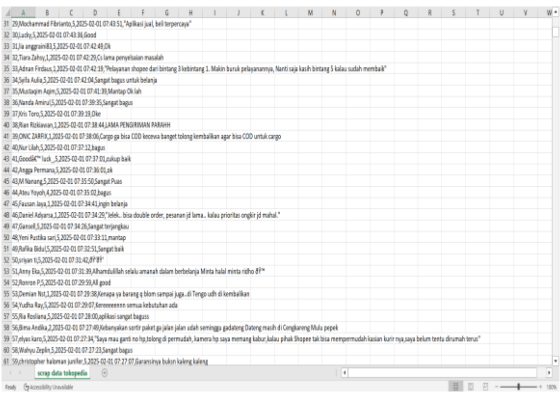

Figure 2. **Tokopedia Application Review Data Presented in .CSV Format**

a) Data Preprocessing

The data preprocessing stage was carried out to clean and prepare the data before analysis. This process included removing duplicate entries, handling missing values, and transforming data by grouping scores into two sentiment categories: positive and negative. Scores of 4–5 were classified as positive sentiment, while scores of 1–3 were classified as negative sentiment. The preprocessing results produced a dataset ready for training and testing machine learning models.

Table 2. **Tokopedia Application Review Dataset**

Date	*Username*	*Skore*	*Review*	*Sentiment*
2024-09-01	Andi	5	The Tokopedia application really helps my online shopping.	Positive
2024-09-02	Budi	4	Quite satisfactory, although delivery took a bit long.	Positive
2024-09-03	Rina	2	Poor service and the application often errors.	Negative
2024-09-04	Tono	5	Original products, competitive prices, and fast delivery.	Positive
2024-09-05	Siti	1	Very disappointing, the item does not match the description.	Negative
2024-09-06	Dimas	4	Shopping made easier and safer	Positive
2024-09-07	Wulan	3	Good app but often crashes when used.	Negative
2024-09-08	Joko	5	Fast service, stable application, and many attractive promotions.	Positive
2024-09-09	Fitri	2	The refund process is long and the CS response is slow.	Negative

.

b) Analysis of the K-Nearest Neighbor Algorithm (KNN)

The K-Nearest Neighbor (KNN) algorithm is a distance-based classification method that determines the class label of a test data point based on the majority class of its nearest neighbors (k-nearest neighbors). In this study, the value of *k* was set to 5, meaning that each test data point was classified based on its five closest neighbors in the TF-IDF vector space. The experimental results show that the KNN algorithm was able to classify Tokopedia user reviews with good accuracy. Out of a total of 400 test reviews (20% of the dataset), the model achieved an accuracy of 85.75%, with a precision of 86.10%, recall of 85.75%, and F1-score of 85.60%. These results indicate that KNN is effective in capturing similarity patterns among reviews based on word proximity. However, the main limitation of KNN lies in its sensitivity to imbalanced data distributions and its relatively high computational time when processing large datasets.

Table 3. **Evaluation Results of the K-Nearest Neighbor (KNN) Algorithm**

No	*Metrik Evaluasi*	*Nilai (%)*
1.	Accuracy	85,75
2.	Precision	86,10
3.	Recall	85,75
4.	F1-Score	85,60

Table 1 shows that the KNN algorithm achieved a fairly good classification performance with an accuracy of 85.75%. The relatively balanced precision and recall values indicate the model's consistent ability to detect both positive and negative sentiments accurately.

c) Analysis Algoritma Naive Bayes

The **Naive Bayes algorithm** is a probability-based classification method that operates under the assumption of feature independence. In the context of sentiment analysis, this algorithm evaluates the likelihood that a review belongs to a positive or negative category based on the probability of word occurrences. Using the same dataset, the Naive Bayes algorithm demonstrated slightly better performance compared to KNN, achieving an accuracy of **88.20%**, precision of **88.50%**, recall of **88.20%**, and an F1-score of **88.00%**. The main advantage of Naive Bayes lies in its ability to process textual data efficiently and produce consistent results even when dealing with high variability in the dataset. This superior performance is attributed to Naive Bayes' capability to leverage the probabilistic distribution of words in the reviews, allowing it to classify sentiments more effectively than KNN, which relies on distance calculations between vector representations.

Table 4. **Evaluation Results of the K-Nearest Neighbor (KNN) Algorithm**

No	*Metrik Evaluasi*	*Nilai (%)*
1.	Accuracy	88,20
2.	Precision	88,50
3.	Recall	88,20
4.	F1-Score	88,00

Table 2 shows that the Naive Bayes algorithm performs better than KNN in classifying user review sentiments for the Tokopedia application. With an accuracy of 88.20%, this algorithm effectively identifies sentiment patterns through its probabilistic approach. The high precision and recall values indicate the model's stability in recognizing both positive and negative reviews consistently.

This study evaluates two machine learning algorithms—K-Nearest Neighbor (KNN) and Naive Bayes—in classifying user sentiment toward the Tokopedia application. Both algorithms were tested using the same dataset, consisting of 2,000 user reviews obtained through web scraping, with an 80% training and 20% testing data split. The review data underwent text preprocessing and was converted into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method.

The experimental results show that the Naive Bayes algorithm outperforms KNN across all evaluation metrics. Naive Bayes achieved an accuracy of 88.20%, precision of 88.50%, recall of 88.20%, and F1-score of 88.00%. Meanwhile, KNN obtained an accuracy of 85.75%, precision of 86.10%, recall of 85.75%, and F1-score of 85.60%. This difference indicates that Naive Bayes has approximately 2.45% higher accuracy compared to KNN.

The superior performance of Naive Bayes can be attributed to its probabilistic nature and the assumption of feature independence, which enables it to efficiently recognize word patterns even when the dataset contains a wide variety of vocabulary. In contrast, KNN relies on distance calculations between vectors in feature space, leading to decreased performance when handling large or imbalanced datasets.

.

Table 5. **Comparison of KNN and Naive Bayes Algorithm Performance**

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
KNN	85,75%	86,10%	85,75%	85,60%
Naive Bayes	88,20%	88,50%	88,20%	88,00%

.

Based on these results, it can be interpreted that the Naive Bayes algorithm is more effective and efficient for sentiment analysis of text data such as user reviews. This algorithm performs better when handling large datasets and has lower computational complexity compared to KNN.

Furthermore, the insignificant difference between the precision and recall values of both algorithms indicates that they have relatively stable performance in distinguishing between positive and negative reviews. However, Naive Bayes demonstrates better prediction consistency, as reflected in its higher F1-score.

.

IV. CONCLUSION

Based on the results of the study comparing the performance of the K-Nearest Neighbor (KNN) and Naive Bayes algorithms in analyzing user satisfaction levels on the Tokopedia application, it can be concluded that both algorithms demonstrate strong capabilities in classifying user review sentiments, although with different levels of effectiveness. The research process began with the collection of 2,000 user reviews obtained through web scraping from the Google Play Store platform. The data then underwent preprocessing and were transformed into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method to produce a dataset suitable for machine learning modeling.

The experimental results show that the Naive Bayes algorithm outperformed KNN across all evaluation metrics, achieving an accuracy of 88.20%, precision of 88.50%, recall of 88.20%, and F1-score of 88.00%. Meanwhile, the KNN algorithm achieved an accuracy of 85.75%, with slightly lower precision, recall, and F1-score values. The superiority of Naive Bayes lies in its probabilistic approach, which allows it to efficiently handle large text datasets with high word variation.

Overall, this study demonstrates that Naive Bayes is more suitable for text sentiment analysis due to its ability to accurately capture word patterns and process data efficiently. On the other hand, while KNN provides competitive results, it is less optimal for large datasets due to its reliance on complex distance calculations between features. Therefore, this study reinforces that probabilistic approaches such as Naive Bayes have greater potential for application in user satisfaction analysis based on e-commerce app reviews like Tokopedia and can serve as a reference for developing automated sentiment analysis systems on digital platforms in the future..

V. REFERENCES

Abdillah, T., Khaira, U., & Hutabarat, B. F. (2024). Komparasi Metode Naive Bayes dan K-Nearest Neighbors Terhadap Analisis Sentimen Pengguna Aplikasi Zenius. Jurnal PROCESSOR, 19(1). https://doi.org/10.33998/processor.2024.19.1.1596

Alamsyah, A., & Saviera, F. (2021). *A Comparison of Indonesia's E-Commerce Sentiment Analysis for Marketing Intelligence Effort (case study of Bukalapak, Tokopedia and Elevenia)*.

Dhany, H. W., Izhari, F., & Komputer, S. (2024). *Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54 Prediction analysis condition animal use algorithm (SVM+KNN)*. www.ejournal.marqchainstitute.or.id/index.php/JICT

Farta Wijaya, R., Kurniawan, F., Putrai, R. R., Alvin, A., Sains, F., & Teknologi, D. (2024). *OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT*.

Genkin, M. (2020). Zero-Shot Machine Learning Technique for Classification of Multi-User Big Data Workloads. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 5046–5055. https://doi.org/10.1109/BigData50022.2020.9378023

Hamid, M. S., Eviyanti, A., Hindarto, H., & Ariyanti, N. (2025). Analisis Sentimen Tingkat Kepuasan Aplikasi WordPress Menggunakan Metode K-Nearest Neighbor dan Naive Bayes. *SMATIKA JURNAL*, *15*(01), 102–114. https://doi.org/10.32664/smatika.v15i01.1522

Hasan Putra, P., Syahputra Novelan, M., & Rizki, M. (n.d.). ANALYSIS K-NEAREST NEIGHBOR METHOD IN CLASSIFICATION OF VEGETABLE QUALITY BASED ON COLOR. In *Journal of Applied Engineering and Technological Science* (Vol. 3, Issue 2).

.

Iqbal, M., & Efendi, S. (2023). Data-Driven Approach for Credit Risk Analysis Using C4.5 Algorithm. *ComTech: Computer, Mathematics and Engineering Applications*, *14*(1), 11–20. https://doi.org/10.21512/comtech.v14i1.8243

Kadek, I., Sugianta, A., Kadek, N., & Patrianingsih, W. (2025). Evaluasi Naïve Bayes dan K-Nearest Neighbor dalam Klasifikasi Sentimen Ulasan Produk Skincare MSGLOW di Tokopedia. *Jurnal ProTekInfo |*, *12*(1).

Khairul, K., Nasyuha, A. H., Ikhwan, A., H. Aly, M., & Ahyanuardi, A. (2023). Implementation of Multiple Linear Regression to Estimate Profit on Sales of Screen Printing Equipment. *JURNAL INFOTEL*, *15*(2), 55–61. https://doi.org/10.20895/infotel.v15i2.934

Novelan, M. S., & Aryza, S. (2025). OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD DISTRIBUTION. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *10*(4), 917–925. https://doi.org/10.33480/jitk.v10i4.6467

Novelan, M. S., Efendi, S., Sihombing, P., & Mawengkang, H. (2023). VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA. *Eastern-European Journal of Enterprise Technologies*, *5*(3(125)), 49–56. https://doi.org/10.15587/1729-4061.2023.288280

Nuranisah, Efendi, S., & Sihombing, P. (2020). Analysis of algorithm support vector machine learning and k-nearest neighbor in data accuracy. *IOP Conference Series: Materials Science and Engineering*, *725*(1). https://doi.org/10.1088/1757-899X/725/1/012118

Putera Utama Siahaan, A., Azizah Harahap, N., Yuni Simanullang, R., & Wanny, P. (2025). Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset. *Bulletin of Information Technology (BIT)*, *6*(1), 33–39. https://doi.org/10.47065/bit.v5i2.1830

Sitorus, Z., Prayogi, D., Rizko, M. A., Suteja, A. G., & Harahap, M. R. (2024). Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm. *Journal of Information Technology, Computer Science and Electrical Engineering (JITCSE)*, *1*(3), 323–328. https://doi.org/10.30596/jitcse

Talmera, A. T., Wardhana, M., & Ratnasari, V. (2025). Analysis of the Impact of UI/UX Elements on User Satisfaction and Loyalty In E-Commerce Platforms: An Empirical Study on the Tokopedia Platform. *Journal of Social Research*. http://ijsr.internationaljournallabs.com/index.php/ijsr

Tapidingan, Y. C., & Paseru, D. (2020). Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students. In *Indonesian Journal of Information Systems (IJIS)* (Vol. 2, Issue 2).

Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., & Schuecker, J. (2023). Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. *IEEE Transactions on Knowledge and Data Engineering*, *35*(1), 614–633. https://doi.org/10.1109/TKDE.2021.3079836

Wijianto, R., Pratmanto, D., Widayanto, A., & Ubaidilah. (2025). Komparasi K-Nearest Neighbors (KNN) dan Naive Bayes pada Klasifikasi Sentimen Ulasan Aplikasi Tokopedia di Google Play Store. Informatics and Computer Engineering Journal, 5(2), 75-80

Wulandari, N., Cahyana, Y., & Hikmayanti Handayani, H. (2025). Sentiment Analysis on the Relocation of the National Capital (IKN) on Social Media X Using Naive Bayes and K-Nearest Neighbor (KNN) Methods. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Issue 3). http://jurnal.polibatam.ac.id/index.php/JAIC