

PERFORMANCE COMPARISON OF C4.5 AND K-NEAREST NEIGHBOR ALGORITHMS FOR MARKETPLACE SALES POTENTIAL ANALYSIS

Sya'roni

Information Systems & Technology, Institut Teknologi Dan Bisnis Mesuji, Lampung, Indonesia

Email: rsya1522@gmail.com

(Received: May 22, 2026; Revised: May 27, 2026; Published: May 30, 2026)

Abstract

The rapid growth of digital marketplace platforms such as Shopee, Tokopedia, and Bukalapak has transformed online business competition and increased the importance of data-driven sales analysis. Marketplace data, including product price, ratings, reviews, sales volume, views, and seller location, contain valuable information that can be utilized to predict product market potential. However, the large volume, heterogeneous characteristics, and dynamic nature of marketplace data make manual analysis inefficient. Therefore, this study aims to analyze and compare the performance of the C4.5 and K-Nearest Neighbor (KNN) algorithms in classifying marketplace sales potential. The dataset used in this research was collected through data scraping from Shopee, Tokopedia, and Bukalapak using the BigSeller application in March 2022, consisting of 21,750 product records with numerical and categorical attributes. Data preprocessing was conducted using Orange Data Mining, including data cleaning, missing value handling, normalization, feature transformation, and dataset partitioning. The classification process categorized products into three market potential levels: low, medium, and high. Model performance was evaluated using a confusion matrix based on accuracy, precision, recall, and F1-score metrics. The experimental results demonstrate that the C4.5 algorithm outperformed KNN, achieving an accuracy of 0.86, while KNN obtained an accuracy of 0.70. Moreover, C4.5 showed higher precision, recall, and F1-score values, indicating better classification consistency and stability. The findings suggest that C4.5 is more effective for marketplace sales potential classification due to its ability to identify influential attributes and manage heterogeneous marketplace datasets. This study contributes to marketplace sales prediction and supports data-driven decision-making in e-commerce environments.

Keywords: marketplace; C4.5; k-nearest neighbor; classification; sales prediction.

1. INTRODUCTION

The development of information technology over the past decade has transformed patterns of economic activity, particularly through the rapid growth of digital marketplace platforms such as Shopee, Tokopedia, and Bukalapak. Marketplaces now serve not only as a means of buying and selling, but have become complex digital ecosystems encompassing consumer interactions, marketing strategies, and data-driven sales performance analysis. In an increasingly competitive environment, the ability to accurately understand a product's market potential is crucial for businesses. Marketplace data such as prices, ratings, number of reviews, discounts, shipping costs, seller location, and number of products sold are important variables influencing purchasing decisions. However, this data is highly dynamic, large in volume, and non-linear, making manual analysis ineffective. Therefore, various studies emphasize the need to apply data mining and machine learning as approaches to extract insights from data and build accurate sales prediction models.

The analysis of previous research provides a strong theoretical basis for this study. Several studies confirm that ratings, reviews, and sales volume are determining factors in the success of product sales in the marketplace [1]. Other studies show that price and discounts also play a significant role in influencing buyer interest, especially in price-sensitive consumer segments[2]. Furthermore, a study on sales prediction based on historical data found that previous sales volume is a key predictor for estimating future sales[3].

In the context of classification methods, the K-Nearest Neighbor (KNN) algorithm has been proven to produce high accuracy in predicting sales potential, especially in datasets with non-linear patterns and high variance between attributes [4]. Other studies have shown that KNN is able to work well on multi-class datasets such as sales categories in marketplaces[5]. In addition, KNN is also commonly used because it is simple yet effective in various e-commerce studies[6].

The C4.5 algorithm is often applied in marketplace research due to its ability to produce easily interpretable decision tree models, thus explaining the factors that influence sales performance [7]. Other studies have shown that C4.5 is capable of achieving high accuracy on categorical and numerical data, making it suitable for use in the context of heterogeneous marketplace data [8]. In addition, C4.5 has the ability to perform pruning to reduce overfitting, thus making the model more stable [9]. The importance of preprocessing as a critical factor in improving model accuracy. Data normalization, extreme value handling, and categorical attribute encoding have been reported to improve model performance by more than 10% in several studies [10]. Furthermore, comprehensive model evaluation using precision, recall, and F1-score has been shown to be more representative than relying solely on accuracy, especially on class-imbalanced datasets [11]. Other studies suggest the application of cross-validation to ensure consistent model performance across various data subsets [12].

Several studies have highlighted the need for statistical testing, such as paired t-tests, to determine whether performance differences between algorithms are truly significant and not due to dataset variation [13]. This approach is particularly important when comparing models on marketplace datasets with fluctuating characteristics. While various studies have examined the use of KNN and C4.5 in marketplace analysis, most studies have used only one algorithm without conducting comprehensive comparisons. The few existing comparative studies generally have not used representative marketplace datasets or have not been equipped with in-depth statistical testing. This leaves a research gap regarding how these two algorithms perform when applied to actual marketplace sales datasets.

Based on this background, this study aims to analyze and compare the performance of the C4.5 and K-Nearest Neighbor algorithms in classifying marketplace market potential for sales. In addition to evaluating the performance of both models using comprehensive evaluation metrics, this study also conducted tests to determine the significance of performance differences between the algorithms. This research is expected to contribute to the development of sales prediction analysis methods in marketplaces and serve as a reference for data-driven decision-making.

2. RESEARCH METHODS

2.1. System Design

The stages of this research for classification using the KNN and C4.5 algorithms are derived from the KDD (Knowledge Discovery in Databases) workflow, which is widely used in previous studies related to marketplace sales analysis. The stages in the implementation used start from data collection, selection of relevant data, data cleaning, transformation, dataset division, algorithm implementation, and evaluation of model results [14]. The following is the data analysis flow for KNN and C4.5 classification on marketplace sales data.

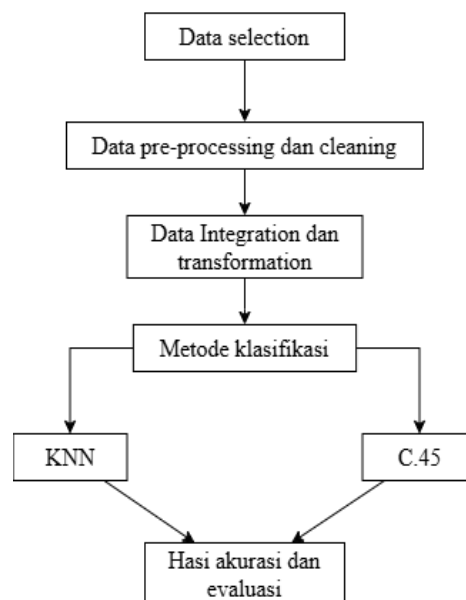


Figure 2. KNN and C4.5 Analysis Flow

2.2. K-Nearest Neighbors

The K-Nearest Neighbors algorithm is an instance-based learning method for classification or regression. KNN is one of several machine learning algorithms. The basic principle of KNN is to find the shortest distance between new data and data in a dataset whose class is known. This method is often used in predictive analytics due to its advantages in managing nonlinear data and its ease of use. The K-Nearest Neighbors algorithm begins by determining the value of K. After determining the value of K, the next step is to calculate the proximity value

between objects. Euclidean distance determines which objects are selected to meet the K value in determining the closest distance between two items [15].

The Euclidean formula can be used to determine the distance between two points in the training data and one point in the testing data as follows.

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Where:

- d : Euclidean distance
- x_{2i} : value of the i-th testing data
- x_{1i} : value of the i-th training data
- p : number of attributes

2.3. C4.5 Algorithm

The technique for creating a decision tree using the given training data is the C4.5 Algorithm. The C4.5 Algorithm is an improvement on ID3. The C4.5 Algorithm requires the following steps to build a decision tree [16]:

1. Prepare the training data. Training data is usually categorized and comes from historical data.
2. Determine the root of the tree by calculating the gain value for each attribute; the first root has the highest gain value. Determine the entropy value first, then the attribute gain value. The entropy value is calculated using the formula.

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Information:

- S : Case Set
- n : Number of partitions of S
- p_i : Proportion of S_i to S

The calculated entropy value is used in calculating the gain value as follows.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3)$$

Where:

- S : set of cases.
- A : Attribute.
- n : number of partitions of attribute A.
- $|S_i|$: number of cases in partition i.
- $|S|$: number of cases in S.

2.4. Model evaluation

In this study, classification performance was measured using a confusion matrix to test the methodology.

1. Accuracy, the proportion of data correctly classified by the system using the formula

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

2. Precision, the percentage of accuracy between the requested data and the classification results using the formula

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

3. Recall, the success of the classification model in re-determining information or value results using a formula

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

4. F1-Score, a comparison of the average value of precision and recall using the formula

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (7)$$

3. RESULTS AND DISCUSSION

3.1. Dataset and Preprocessing Results

The initial dataset used was obtained directly from the marketplace websites Shopee (<https://shopee.co.id/>), Bukalapak (<https://www.bukalapak.com/>), and Tokopedia (<https://www.tokopedia.com>) in March 2022 through scraping using the BigSeller application. The dataset consists of 145 data sets, each containing 150 records, for a total of 21,750 records. Each data set has 11 attributes, each with categorical and numeric values: Product Link,

Category 1, Category 2, Category 3, Item Name, Price, Reviews, Sold, Views, Rating, Location, and Sales, as shown in Table 1.

Tabel 1. Initial dataset

NO	X1	X2	X3	X4	X5
1	160.000	3	5,0	10	0,00
2	18.000	50	5,0	376	0,00
3	19.150	19000	4,2	63000	0,65
4	19.500	196	4,8	574	0,00
5	7.000	183	4,9	10000	0,00
6	55.200	84000	4,7	10000	0,08
7	100.000	88	4,5	192	0,00
8	34.250	21000	4,7	59000	0,00
9	100.000	12000	4,7	25000	0,00
10	110.890	0	0,0	2	0,28
.....
21747	300.000	0	0,0	0	0,00
21748	45.000	1	5,0	1	0,10
21749	67.500	0	0,0	1	0,10
21750	58.500	10	4,9	15	0,10

The next step involved cleaning empty attribute values and removing missing and noisy data to improve dataset quality. After the preprocessing stage, data integration was performed by combining all dataset variables into a single table. Furthermore, the normalization process was applied to prevent data anomalies and inconsistencies, ensuring that the dataset was suitable for further classification analysis.

The preprocessing stage was carried out using the Orange Data Mining application to improve data quality before the classification process. Orange Data Mining was selected because it provides a visual workflow interface that facilitates preprocessing, transformation, and evaluation without extensive programming implementation. The preprocessing process consisted of several stages, including data cleaning, missing value handling, data integration, normalization, feature transformation, and dataset preparation. The workflow began by importing marketplace datasets using the File widget, followed by selecting relevant attributes through Select Columns and adjusting data types using Edit Domain. Missing or incomplete values were handled using the Impute widget to improve dataset consistency. Numerical attributes were then normalized using the Normalize widget to ensure uniform value ranges and reduce data inconsistencies. Furthermore, categorical attributes were transformed into numerical representations using the Continuize widget before being processed by the classification algorithms. The preprocessing results were inspected using the Data Table widget, while dataset partitioning was performed using Data Sampler.

Table 2. Dataset Results After Normalization

NO	X1	X2	X3	X4	X5
1	0,001848569	6,31579E-06	1	0,000101	0
2	0,000206948	0,000105263	1	0,003798	0
3	0,000220243	0,04	0,84	0,636364	0,698925
4	0,000224289	0,000412632	0,96	0,005798	0
5	7,97804E-05	0,000385263	0,98	0,10101	0
6	0,000637007	0,176842105	0,94	0,10101	0,086022
7	0,001154926	0,000185263	0,9	0,001939	0
8	0,00039481	0,044210526	0,94	0,59596	0
9	0,001154926	0,025263158	0,94	0,252525	0
10	0,001280822	0	0	2,02E-05	0,301075
.....
21747	0,003467068	0	0	0	0
21748	0,000519087	2,10526E-06	1	1,01E-05	0,107527
21749	0,000779203	0	0	1,01E-05	0,107527
21750	0,000675157	2,10526E-05	0,98	0,000152	0,107527

After the normalization process, marketplace products were categorized into three levels of market potential, namely low, medium, and high, according to their sales characteristics and product performance indicators. This categorization was intended to provide a clearer representation of product sales patterns and identify the potential

level of each marketplace category. The classification stage was conducted using the K-Nearest Neighbor (KNN) algorithm by calculating the Euclidean distance between training data and testing data based on Formula (1). This calculation was used to measure the similarity between data objects and determine the nearest neighboring instances with comparable characteristics.

3.2. KNN Classification Results

The implementation of KNN started by defining the value of K, which represents the number of neighboring data points involved in the classification process[13]. Subsequently, distance calculations were performed using marketplace attributes such as product price, number of reviews, ratings, sales quantity, and product views. Products with the closest distance values were considered to have higher similarity and were used as references in assigning market potential categories. The experimental results indicate that the KNN algorithm was able to classify marketplace products into their respective market potential levels by utilizing similarity relationships among data records. This approach provided useful insights into marketplace sales behavior and supported the identification of product categories with different market potential characteristics.

Tabel 3. KNN Classification Results

No	Sales Category	Product Category	KNN Prediction	KNN (Low)	KNN (Medium)	KNN (High)
1	High	Fashion	High	0	0	1
2	High	Fashion	High	0	0	1
3	High	Fashion	Low	0.666667	0	0.333333
4	High	Fashion	High	0	0	1
...
2249	Low	Automotive	Low	1	0	0
2250	Low	Automotive	Low	1	0	0

3.3. C4.5 Decision Tree Analysis

Implementation of the C4.5 algorithm. In this process, the entropy value is first calculated using Formula (2) to determine the level of data uncertainty. Next, the entropy value is used to calculate the gain value using Formula (3) to determine the best attributes in forming a decision tree. The processing results produce a C4.5 level 4 decision tree.

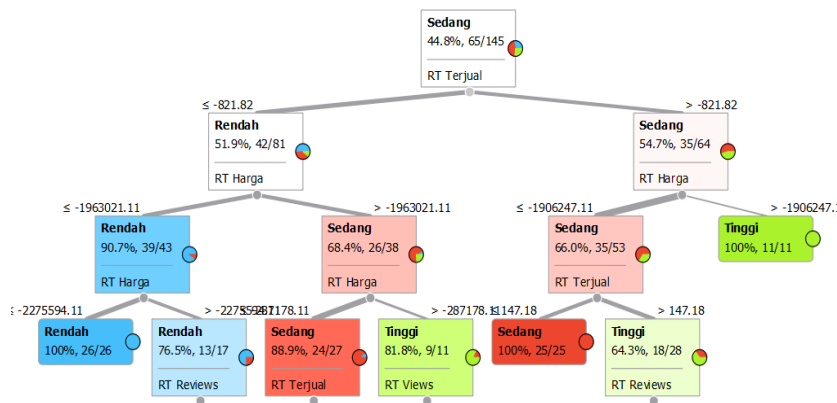


Figure 2. C4.5 level 4 decision tree

Based on calculations using the C4.5 algorithm, the sold attribute had the highest gain value of 44.8%, representing 65 product categories out of a total of 145 product categories used in the study. This high gain value indicates that the sold attribute has the best ability to differentiate and group data compared to other attributes. Therefore, this attribute is the primary and most influential factor in the classification process of marketplace product sales potential. A high gain value for the "sold" attribute indicates that the number of products sold significantly determines the market potential of a product category. The higher the number of products sold, the greater the likelihood that the product category has high market potential. Therefore, the "sold" attribute was chosen as the root or primary node in the decision tree because it provides the most optimal information for the decision-making process.

3.4. Performance Evaluation of KNN and C4.5

Based on data processing using the C4.5 and KNN methods, distinct sales patterns were found in the Fashion, Electronics, and Automotive categories on the Shopee, Tokopedia, and Bukalapak marketplaces. The classification results indicate that each product category has varying levels of market potential based on sales volume, product price, and consumer interest levels in each marketplace.

In the classification testing process using the K-Nearest Neighbor (KNN) and C4.5 algorithms, an evaluation was conducted based on a confusion matrix to determine the level of prediction accuracy of both methods. The dataset used has a sufficiently large number of records, providing a realistic picture of each algorithm's ability to recognize data patterns. The confusion matrix is the primary analysis tool because it can show in detail how effectively the model distinguishes positive and negative classes. The four core components of the confusion matrix—True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)—are direct determinants of the model's prediction quality. High TP and TN indicate that the model is able to classify data accurately, while low FP and FN describe model errors that can significantly impact analysis results, particularly in the context of business or market predictions.

The test results show that the C4.5 algorithm has superior performance compared to KNN. The C4.5 TP value reached 115 and TN was 260, significantly higher than the KNN TP of only 79 and TN of 224. This difference indicates that C4.5 is more effective in accurately recognizing data categories, both in the positive and negative classes. In contrast, KNN recorded higher FP and FN values, each amounting to 66, indicating that the model frequently misclassifies data. These errors are very likely because KNN operates based on distance between data and is therefore very sensitive to data distribution, value scales, and the presence of outliers. If the data has a large variation between item prices and sales quantities, KNN is more likely to produce prediction errors.

Table 4. Analysis using KNN and C4.5

Model	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
KNN	79	224	66	66
C4.5	115	260	30	30

The differences in the characteristics of the two algorithms also contribute to their final performance. C4.5 works by forming a decision tree through information gain calculations, thus selecting the most relevant attributes to distinguish data categories. This approach makes C4.5 more stable and able to handle complex patterns. In contrast, KNN is a lazy learner algorithm that doesn't build a model but instead calculates distances directly during predictions. When data is large or non-standardized, KNN's performance easily declines. This is evident in its accuracy, precision, recall, and F1-score values, which are consistently lower than those of C4.5.

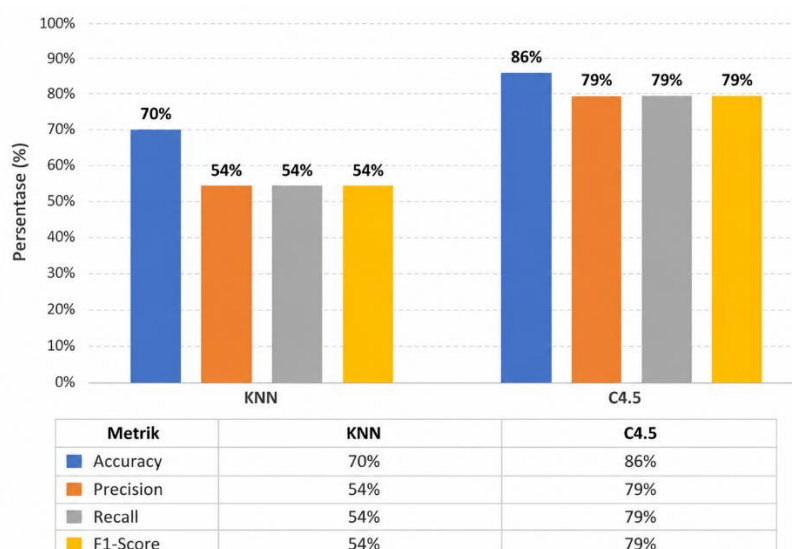


Figure 3. comparison of algorithm matrix evaluation

Overall, the metric values show a quite striking difference between the two methods. KNN only achieved an accuracy of 0.70, with precision, recall, and F1-score values of 0.54. Meanwhile, C4.5 achieved an accuracy of 0.86, with precision, recall, and F1-score values of 0.79 each. These metrics illustrate that C4.5 is more capable of providing balanced predictions between its ability to detect positive and negative data. The consistency of precision

and recall values in C4.5 also indicates that this algorithm is not only accurate in recognizing positive data, but also has minimal errors in avoiding false positive predictions.

These findings confirm that C4.5 is more suitable for use on datasets with complex patterns, especially those involving numeric variables such as product price and sales volume. The model's ability to automatically identify key attributes makes C4.5 more adaptive and stable in generating predictions. On the other hand, KNN can still be used, but it requires data normalization, selecting the right k value, and noise reduction to achieve optimal performance. Therefore, in the context of this research, the C4.5 algorithm is proven to be superior and worthy of being the primary choice for marketplace data-based classification needs.

3.5. Discussion

The findings of this study demonstrate that both the K-Nearest Neighbors (KNN) and C4.5 algorithms are capable of classifying marketplace product sales potential; however, significant differences were observed in their performance. The experimental results revealed that the C4.5 algorithm outperformed KNN across all evaluation metrics, including accuracy, precision, recall, and F1-score. Specifically, C4.5 achieved an accuracy of 0.86, substantially higher than KNN, which obtained an accuracy of 0.70. This indicates that the decision tree approach employed by C4.5 is more effective in identifying complex sales patterns within marketplace data characterized by heterogeneous attributes such as price, ratings, reviews, number of products sold, and product views. The superior performance of C4.5 can be attributed to its ability to select the most informative attributes through information gain calculations, thereby producing a more structured and interpretable classification model.

Furthermore, the results indicate that the "sold" attribute had the highest contribution to marketplace sales potential classification, accounting for 44.8% of the gain value and representing 65 product categories. This finding suggests that product sales quantity is the most influential factor in determining market potential. The decision tree generated by C4.5 effectively identified this dominant attribute and used it as the root node, enabling more accurate categorization of low, medium, and high market potential products. This result aligns with previous studies emphasizing that historical sales volume is a strong predictor of future product performance in digital marketplaces. The ability of C4.5 to prioritize influential variables enhances its robustness when handling large-scale and mixed-type marketplace datasets.

In contrast, KNN showed lower classification performance, mainly due to its dependence on distance calculations between data points. Although normalization was applied during preprocessing, KNN remained sensitive to data distribution, attribute scaling, and the presence of noise or outliers. The relatively high false positive and false negative values indicate that KNN frequently misclassified products, particularly in categories with overlapping characteristics. Since KNN relies on neighborhood similarity rather than rule extraction, its predictive capability becomes limited when dealing with highly dynamic and nonlinear marketplace data. Nevertheless, KNN still demonstrated acceptable performance and may remain suitable for simpler classification problems or datasets with more homogeneous characteristics.

From a practical perspective, the results of this study suggest that C4.5 is more suitable for marketplace sales analysis and product market potential classification due to its higher predictive stability and interpretability. Decision tree visualization generated by C4.5 can also assist business stakeholders in understanding the key factors influencing product sales performance, thereby supporting data-driven decision-making. However, future studies are recommended to integrate ensemble approaches or hybrid machine learning models to further improve predictive accuracy. Additionally, implementing cross-validation and statistical significance testing on larger and more recent marketplace datasets may provide stronger evidence regarding the comparative effectiveness of classification algorithms in e-commerce environments.

4. CONCLUSION

Based on the results of classification tests using the K-Nearest Neighbor (KNN) and C4.5 algorithms, it can be concluded that the C4.5 algorithm provides superior performance compared to KNN in predicting data categories in sales or marketplace datasets. This is evident from the significantly higher True Positive (TP) and True Negative (TN) values in C4.5, as well as the relatively lower False Positive (FP) and False Negative (FN) values. This difference in performance indicates that C4.5 is better able to recognize complex patterns in data and produce more accurate decisions. Overall, the C4.5 algorithm achieved an accuracy of 0.86, while KNN only achieved 0.70. The precision, recall, and F1-score of C4.5, all at 0.79, also confirm the model's consistency and stability in classification. In contrast, KNN only achieved a score of 0.54 across all three metrics, indicating that the model is more susceptible to misclassification due to its sensitivity to distance and data distribution.

Thus, it can be concluded that C4.5 is more suitable for use in the context of this research, especially when the dataset contains diverse variable patterns such as product prices and sales volumes. While KNN can still be used, its performance is highly dependent on parameter selection and the quality of data preprocessing. These results reinforce the point that selecting an appropriate algorithm is crucial for the success of the classification process in market potential analysis and data-driven decision-making.

REFERENCES

- [1] Sya'roni, "Analisis Potensi Marketplace Terhadap Penjualan Menggunakan K-Nearest Neighbors," *JURNAL SISTEM DAN INFORMATIKA (JSI)*, vol. 19, no. 2, pp. 1–9, Oct. 2025, doi: <https://doi.org/10.30864/jsi.v19i2.724>.
- [2] M. Utami and V. Ayumi, "Prediksi Penjualan Produk Terlaris Menggunakan Algoritma K-Nearest Neighbour (KNN)," *JCOSIS (Journal Computer Science and Information Syetem)*, vol. 1, no. 2, pp. 43–47, 2024, doi: <https://doi.org/10.61567>.
- [3] F. U. Aulya and K. Kusnawi, "Evaluating Classification Models for Predicting Product Success in Indonesian E-Commerce," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 2723–2739, Aug. 2025, doi: [10.52436/1.jutif.2025.6.4.5071](https://doi.org/10.52436/1.jutif.2025.6.4.5071).
- [4] L. Nadhia Ningsih, R. Septiani, A. Pramadjaya, and S. Nuralisah, "Implementing the C4.5 Algorithm for Customer Satisfaction Classification," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 4, pp. 1470–1480, Sep. 2025, doi: [10.57152/malcom.v5i4.2211](https://doi.org/10.57152/malcom.v5i4.2211).
- [5] N. Ubaedilah, Puji Isyanto, and Asep Darajatul Romli, "Analisis Faktor-Faktor Yang Mempengaruhi Keputusan Pembelian Impulsif Pada Pengguna Tiktok Shop," *Journal of Trends Economics and Accounting Research*, vol. 4, no. 1, pp. 46–56, Sep. 2023, doi: [10.47065/jtear.v4i1.875](https://doi.org/10.47065/jtear.v4i1.875).
- [6] M. Yasir Alghifari *et al.*, "COMPARISON OF SVM AND NAIVE BAYES ALGORITHMS IN SENTIMENT ANALYSIS OF USER REVIEWS ON BUKALAPAK," *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, vol. 10, no. 3, pp. 1623–1633, Nov. 2025, doi: <https://doi.org/10.35314/dqhpkb12>.
- [7] E. Purwanto, B. P. Cipto Utomo, H. Permatasari, and F. Mohd, "Comparative Analysis of Classification Models for Sales Prediction in E-commerce: Decision Tree, Random Forest, SVM, Naive Bayes, and KNN," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 6, pp. 5899–5915, Jan. 2026, doi: [10.52436/1.jutif.2025.6.6.5224](https://doi.org/10.52436/1.jutif.2025.6.6.5224).
- [8] A. Sonita and A. Lestari, "Implementasi Metode K-Nearest Neighbor Untuk Prediksi," *JSAI: Journal Scientific and Applied Informatics*, vol. 7, no. 3, pp. 544–552, Nov. 2024, doi: [10.36085](https://doi.org/10.36085).
- [9] N. F. Octavia and Berlilana, "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Ulasan Produk Elektronik pada Platform E-Commerce," *Jurnal Algoritma*, vol. 22, no. 2, pp. 2110–2121, Nov. 2025, doi: [10.33364/algoritma/v.22-2.3083](https://doi.org/10.33364/algoritma/v.22-2.3083).
- [10] M. Redinal Muktar, M. Rifqi Faidhil Syam, A. Kunda, M. Syahlan Natsir, and J. K. Sistem Informasi Universitas Dipa Makassar Jln Perintis Kemerdekaan, "Analisis Penerapan Data Mining dalam Klasifikasi Penjualan Pakaian Pada Toko Online Shopee Menggunakan Algoritma C4.5," *JURNAL DIPANEGARA KOMPUTER SISTEM INFORMASI*, vol. 17, no. 1, pp. 1–8, Jun. 2023, doi: <https://doi.org/10.36774/dipakomsi.v17i1.1388>.
- [11] J. Nangi and R. Rinaldi Hadistio, "PENERAPAN DATA MINING UNTUK MEMREDIKSI PENJUALAN MENGGUNAKAN METODE KNEAREST NEIGHBOR (STUDI KASUS: THRIFTING SECOND 3)," *ANIMATOR*, vol. 3, no. 3, pp. 1–5, Dec. 2025, Accessed: Mar. 03, 2026. [Online]. Available: <https://animator.uho.ac.id/index.php/journal/article/view/1284/56>
- [12] M. Danny, A. Muhidin, and A. Jamal, "Application of the K-Nearest Neighbor Machine Learning Algorithm to Product Sales of Best-Selling Products," *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 1, pp. 255–264, Jun. 2024, doi: [10.47709/brilliance.v4i1.4063](https://doi.org/10.47709/brilliance.v4i1.4063).
- [13] A. Verma, C. Nagar, S. Haryani, and S. Jain, "Prediction of E-Commerce Shoppers' Purchasing Intention using Knn Algorithm," in *Atlantis Press*, 2025, pp. 63–72. doi: [10.2991/978-94-6463-716-8_6](https://doi.org/10.2991/978-94-6463-716-8_6).
- [14] Z. Fatah and M. Syafiq, "Prediksi Penjualan Sepeda Motor Menerapkan Metode K-Nearest Neighbor," *JAMASTIKA*, vol. 5, no. 1, pp. 333–342, Apr. 2026, doi: <https://doi.org/10.35473/jamastika.v5i1.4701>.
- [15] Roni S and C. Crysdiandian, "Jurnal Teknologi dan Manajemen Informatika Studi Literature Analisis Potensi Pasar Marketplace terhadap Penjualan Article Info ABSTRACT," *Jurnal Teknologi dan Manajemen Informatika (JTMI)*, vol. 8, no. 2, pp. 134–142, 2022, [Online]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [16] F. Wijaya, R. Rahmansyah, and N. A. Hasibuan, "Prediksi Penjualan Barang Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Sistem Cerdas dan Rekayasa (JSCR)*, vol. 7, no. 2, pp. 2656–7504, Oct. 2025, doi: <https://doi.org/10.61293/jscr.v7i2.856>.