Comparative Analysis Of K-Nearest Neighbors And Decision Tree Methods In Determining Students' Purchase Interest In Macbook Laptops

Intan Fadilla Hasibuan[1*], Deci Irmayani[2], Volvo Sihombing[3]

[1,2,3] Faculty of Science and Technology, Universitas Labuhanbatu, Sumatera Utara Indonesia.
[*]Corresponding Author:
Email: intanfadillahsb@gmail.com

Abstract.

In the context of increasingly competitive technology markets, companies need to know consumer preferences accurately to optimize product offerings and increase sales. Two classification methods that are often used in data mining, namely K-Nearest Neighbors and Decision Tree, have their own advantages and disadvantages. This study proposes a solution that involves processing student data using both classification methods to identify the most accurate and effective method for identifying purchase intentions. This study aims to compare the performance of the two methods in determining student purchase intentions for MacBook laptops. The research methodology includes collecting data from 100 students covering various factors such as user experience, design and portability, technical specifications, price, and security. This data is then classified using the K-Nearest Neighbors and Decision Tree methods. Furthermore, a confusion matrix is used to provide a more detailed picture of the performance of the two methods. The results of the study show that the Decision Tree method has a higher accuracy (91%) compared to K-Nearest Neighbors (88%). In addition, Decision Tree excels in other metrics such as precision (87.18% vs. 85.71%), recall (89.47% vs. 85.71%), specificity (91.94% vs. 89.66%), and F1-Score (88.31% vs. 85.71%). The decision tree also has a higher NPV value and lower FPR and FNR rates than K-Nearest Neighbors, indicating that it is superior in avoiding misclassification. The study's conclusion is that the Decision Tree method is more effective and accurate than K-Nearest Neighbors in determining students' purchase intentions for MacBook laptops. The decision tree shows better performance in almost all evaluation metrics, making it a more reliable method to use in consumer data analysis. The results of this study are expected to help companies choose a more appropriate and effective analysis method for their marketing strategies, as well as provide a basis for further research in the field of consumer purchase intention classification.

Keywords: *Data Mining, Decision Tree, K-Nearest Neighbors, MacBook Laptop and Purchase Interest.*

I. INTRODUCTION

Laptops are an important tool for students in their academic lives. With a laptop, students can access various digital resources, such as journals, articles, and e-books, that are very important to support learning and research. In addition, laptops make it easier for students to create and edit documents, make presentations, and complete assignments and projects efficiently. The ability to carry a laptop anywhere also allows students to study and work from various locations, whether on campus, in the library, or at home. Each student has their own preferences in choosing the brand and type of laptop they want, adjusted to their needs and lifestyle. Some students may prefer laptops from popular brands such as Acer, Asus, HP, and Lenovo because of the more affordable price factor and adequate features for academic and entertainment purposes. There are also those who prefer a gaming laptop with high specifications for playing games and running heavy applications, which can provide the best visual experience and performance. This preference is often based on personal experience, recommendations from friends, or reviews from other users. On the other hand, there are also students who choose laptops with a slim and thin design that offers high portability without sacrificing performance, such as the MacBook.The MacBook is known for its elegant design, long battery life, and stable and reliable performance [1]. These advantages make the MacBook a favorite choice for those who want a laptop with high capabilities that is still practical to carry anywhere. On the other hand, there are still some students who do not like MacBook laptops.

One of the main reasons is the price, which is quite expensive compared to laptops from other brands. Students with a limited budget may feel that investing in a MacBook is not worth the cost, especially when there are other laptop options that offer excellent specifications at a more affordable price. In addition, additional costs for accessories and repairs also tend to be higher on MacBooks, adding to their financial burden. In addition to the price issue, using a MacBook is also considered not as easy as laptops in general for some students. The MacOS operating system, which is different from Windows [2], which is more commonly used, can be a challenge. Students who are used to Windows may need time to adapt to the macOS interface and how it works. Applications that are often used on Windows may not be available or have different versions on macOS, which can interfere with productivity. These factors make some students prefer laptops that are more familiar and simple to use to support their academic and daily activities.Data mining is a series of processes used to extract and identify patterns in databases that are used to search for knowledge in large amounts of data. The process of data mining is known as Knowledge Discovery in Databases (KDD) because KDD deals with patterns in large amounts of data through integration techniques and scientific discovery, interpretation, and visualization [3]. Data mining is the process of extracting useful information from large and complex datasets through data analysis techniques [4]. This process involves various methods and algorithms to find patterns, trends, and relationships in data that may not be immediately apparent [5], [6].

Data mining is used in various fields such as business, health, social sciences, and technology to help make better decisions and more accurate predictions [7]. The stages in data mining usually include data cleaning, data transformation, pattern mining, pattern evaluation, and knowledge representation [8]. By utilizing data mining, organizations can transform raw data into meaningful and actionable insights [9]. Companies can use data mining to make faster and more accurate decisions in the face of fierce market competition. In the long run, this can help companies improve their performance and profitability [10].K-Nearest Neighbors (KNN) is one of the classification methods in data mining that is included in supervised learning, where classification is done based on attributes and learning data. Thus, the process of classifying new data is done based on a comparison of the majority similarity in the learning data [11]. KNN is a method that is easy to understand, simple, and very suitable for application to determine the level of public satisfaction, be it the quality of a product or the comfort of a place [12]. The KNN method can produce accurate and accountable results [13]. This method can calculate the closeness between new cases and old cases through weight matching [14]. Another classification approach is the Decision Tree method, which uses a series of rules to make decisions in a manner similar to a tree. The concept is to present an algorithm with conditional statements consisting of branches to indicate decision-making steps that can produce profitable results. This classification finds targets on branches by looking at nodes [15]. The Decision Tree method is very suitable for classifying interests and works well according to user desires and needs [16].

The ability of a decision tree to create models that are easy to understand and interpret is the main benefit that makes this method so attractive [17]. The results show that the decision tree model can be used in business applications, especially in the fields of consumer behavior prediction and market analysis, because it can predict customer interests very accurately and precisely [18].From the above view, the author is interested in conducting research on students' interest in MacBook laptops. This study aims to conduct a comparative analysis of the KNN and Decision Tree methods in determining students' purchasing interest in MacBook laptops. Each method will be evaluated for its performance using performance evaluation matrices such as accuracy, precision, recall, specificity, and f1-score. This study was conducted to understand the factors that influence students' preferences for MacBook laptops, even though they are more expensive and their use is considered less intuitive for some users. The analysis will cover various aspects, such as user experience, design and portability, technical specifications, price, and security. By understanding the motivation behind this choice, the study is expected to provide deeper insight into technology preferences among students. The results of this study are expected to provide recommendations for laptop manufacturers and campuses on how to provide technology that is more in line with students' needs and preferences.

II. METHODS

For the research stage, the author uses the Knowledge Discovery in Databases (KDD) approach. The KDD stages include the data selection process, preprocessing, transformation, data mining, and pattern evaluation. The data selection stage is carried out to collect the data used in this study. With this stage, the author can easily collect the data that will be used and needed in this study. For the dataset used in data mining, namely training data and testing data, Training data is training data that will help the data classification process. For the data used in training data, it is 20. Testing data is research sample data that will be processed in data mining using the KNN method and the Decision Tree method. The testing data that will be used in this study is 100. The data preprocessing stage is a stage that is carried out to select data that is suitable for use. The transformation stage is a stage that is carried out to change the format and form of existing data that has been obtained in the form of an Excel file. The data mining stage is a stage that is carried out to design a classification model that will be used to perform data analysis with the help of the KNN method and the Decision Tree method. The evaluation stage is useful for testing how much the method is capable of classifying data. In order to conduct an evaluation in this study, the author must also design an evaluation model that can be used to provide evaluation results. By using the KDD approach, this study will ensure that every step from data processing to final analysis is carried out systematically and structured, producing accurate and reliable insights into students' interest in MacBook laptops.

III. RESULT AND DISCUSSION

In this study, KNN and decision tree methods are used to analyze students' interest in MacBook laptops. KNN helps in classifying student data based on attributes such as user experience, design and portability, technical specifications, price, security, and category. By measuring the proximity between students who are interested in MacBooks and those who are not, KNN can identify significant preference patterns. This method provides insight into which groups of students are more likely to choose a MacBook based on their similar characteristics, allowing for more focused and relevant analysis. In addition, Decision Tree is used to create an easy-to-understand prediction model regarding students' interest in the MacBook. This method will divide the data into branches based on certain conditions. With a clear visualization of the decision tree, the author can identify the most influential variables in the MacBook purchase decision. Both methods, KNN and Decision Tree, provide an in-depth complementary approach to data analysis, ensuring that every important aspect of student preferences can be revealed and analyzed thoroughly.

Table 1. A Part of Dataset

Student	User Experience	Design and Portability	Technical Specifications	Price	Security	Category
Student01	Ever	Good	Not Good	Expensive	Weak	Not Interested
Student02	Ever	Good	Not Good	Expensive	Weak	Not Interested
Student03	Never	Good	Not Good	Expensive	Weak	Not Interested
Student04	Often	Good	Good	Expensive	Strong	Interested
Student05	Never	Good	Not Good	Expensive	Weak	Not Interested
Student06	Never	Good	Not Good	Expensive	Weak	Not Interested
Student07	Ever	Good	Not Good	Expensive	Weak	Not Interested
Student08	Never	Good	Not Good	Expensive	Weak	Not Interested
Student09	Ever	Good	Not Good	Expensive	Weak	Not Interested
Student10	Ever	Good	Good	Expensive	Strong	Interested

Table 1 shows some of the datasets used in this study. User experience indicates whether students have used a MacBook before or frequently. Design and Portability: assess the design and portability of the MacBook according to students' views. Technical specifications categorize whether students consider the technical specifications of the MacBook to be good or not. Price indicates whether students consider the price of the MacBook to be expensive or not. Security indicates whether students consider the security of the MacBook to be strong or weak. The category indicates whether students are interested or not in purchasing a MacBook. Based on the influence of user experience, most students who have used a MacBook tend to be less interested in purchasing it, possibly because the price and technical specifications do not meet their expectations. Students who frequently use a MacBook and consider the technical specifications and security to be strong tend to be interested in purchasing it. All students consider the design and portability of the MacBook to be good, indicating that design is not the main determinant in purchasing decisions. Price is the main barrier.

All students consider the MacBook to be expensive, which has a negative impact on purchase intention. Technical specifications that are considered poor by the majority of students in the sample data cause their interest in the MacBook to decrease. Strong security increases purchase intention, as seen in Student 04 and Student 10. The main factors that influence students' purchase intentions for the MacBook are technical specifications and price. Although the design and portability are considered good, as well as the experience of users who have used MacBooks, the high price and inadequate technical specifications are the main deterrent factors. This analysis provides deep insights into students' preferences for MacBooks and can be a guide for manufacturers in improving their products to attract more potential buyers.

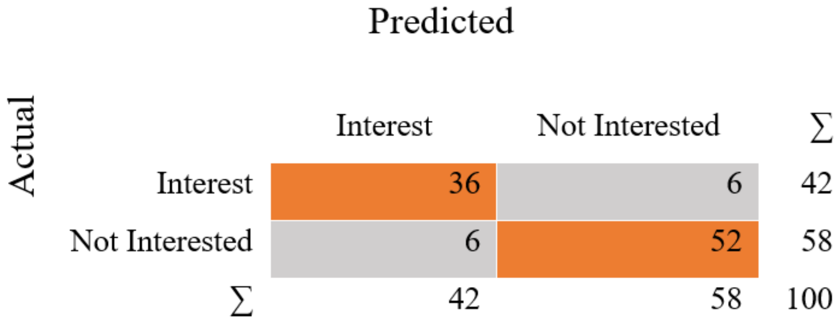

Fig 1. Confusion Matrix of KNN

Figure 1 shows a confusion matrix that illustrates the performance of the KNN method in predicting students' purchase interest in MacBook laptops. Based on the confusion matrix, we can conduct a more in-depth analysis to understand the strengths and weaknesses of the KNN method. True Positive (TP) shows that 36 students are indeed interested in MacBook laptops and are correctly classified by the model. True Negative (TN) shows that 52 students are not interested in MacBook laptops and are correctly classified by the model. False Positive (FP) shows that six students are actually not interested but are classified as interested by the model. False Negative (FN) shows that six students are actually interested but are classified as not interested by the model.

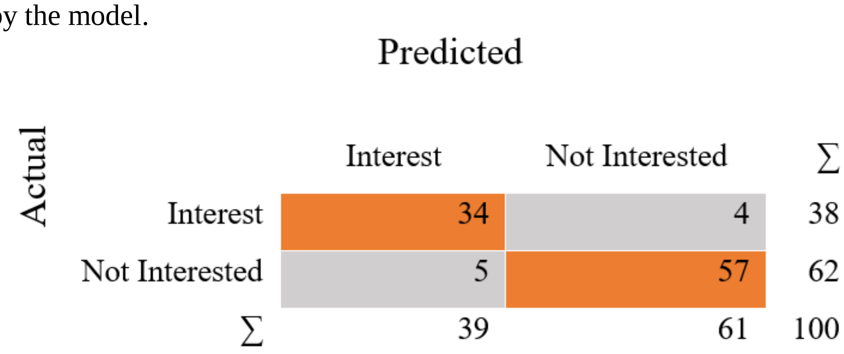

Fig 2. Confusion Matrix of Decision Tree

Figure 2 shows the confusion matrix that illustrates the performance of the Decision Tree method in predicting students' purchase interest in MacBook laptops. True Positive (TP) shows that 34 students are indeed interested in MacBook laptops and are correctly classified by the model. True Negative (TN) shows that 57 students are not interested in MacBook laptops and are correctly classified by the model. False Positive (FP) shows that 5 students are actually not interested but are classified as interested by the model. False Negative (FN) shows that 4 students are actually interested but are classified as not interested by the model.

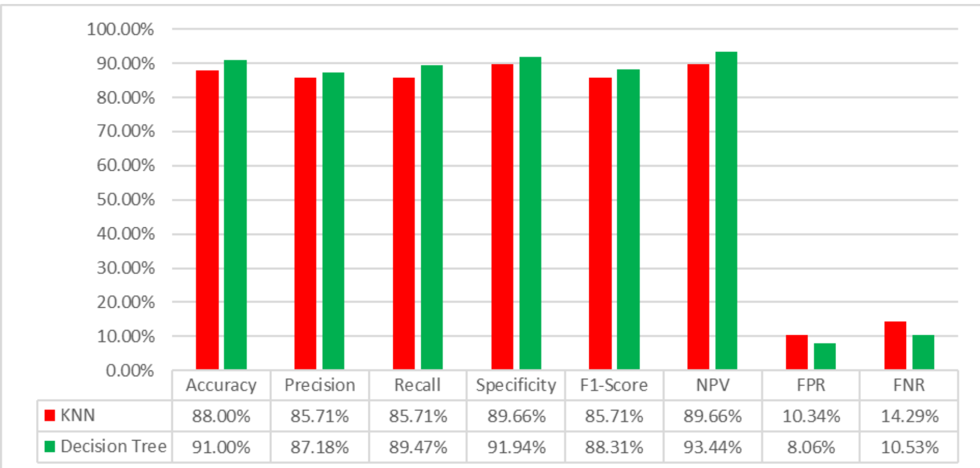

Fig 3. Comparison of Performance Evaluation Matrix

Figure 3 is a comparison of the performance evaluation matrix results of two classification methods, namely K-Nearest Neighbors (KNN) and Decision Tree, in determining students' purchase interest in MacBook laptops. The compared matrices include various evaluation metrics: accuracy, precision, recall, specificity, F1-score, negative predictive value (NPV), false positive rate (FPR), and false negative rate (FNR). Accuracy indicates how often the model provides correct predictions. Decision Tree has a higher accuracy (91%) than KNN (88%), indicating that Decision Tree is better at classifying the sample data as a whole. Precision measures how well the model is at identifying students who are truly interested in MacBook laptops. Decision Tree has a slightly higher precision (87.18%) than KNN (85.71%), indicating that Decision Tree is better at minimizing false positives. Recall indicates the model's ability to capture all students who are truly interested in MacBook laptops. Decision Tree has a higher recall (89.47%) than KNN (85.71%), indicating that Decision Tree is better at identifying students who are truly interested. Specificity measures the ability of the model to identify students who are not interested in MacBook laptops. Decision Tree has a higher specificity (91.94%) than KNN (89.66%), indicating that Decision Tree is better at avoiding false positives. F1-Score is the harmonic mean of precision and recall. Decision Tree has a higher F1-Score (88.31%) than KNN (85.71%), indicating that Decision Tree has a better balance between precision and recall. NPV indicates how well the model is at identifying students who are not interested in MacBook laptops. Decision Tree has a higher NPV (93.44%) than KNN (89.66%), indicating that Decision Tree is better at minimizing false negatives.

FPR measures how often the model incorrectly identifies uninterested students as interested. Decision Tree has a lower FPR (8.06%) than KNN (10.34%), indicating that Decision Tree is better at avoiding false positives. FNR measures how often the model fails to identify interested students. Decision Tree has a lower FNR (10.53%) than KNN (14.29%), indicating that Decision Tree is better at avoiding false negatives.Accuracy is one of the main evaluation metrics used to assess the performance of a classification model. With 91% accuracy for Decision Tree compared to 88% for KNN, this shows that Decision Tree makes more correct predictions than KNN. However, it is important to note that accuracy alone is not enough in situations where the classes are imbalanced or where the costs of false positives and false negatives are very different. In this context, the higher accuracy of the decision tree indicates that the model is more effective in handling the dataset at hand. The higher precision of Decision Tree (87.18%) compared to KNN (85.71%) indicates that the Decision Tree model is more reliable in minimizing false positives, i.e., cases where students are classified as interested in MacBooks when they are not. This is very important in the context of marketing and targeting, as it reduces the waste of resources on irrelevant targets. The higher recall of Decision Tree (89.47%) compared to KNN (85.71%) indicates that Decision Tree is better at capturing all students who are truly interested in MacBooks. This means that the decision tree model is more efficient in identifying potential market targets, reducing the risk of losing potential customers. Specificity measures the ability of the model to identify students who are not interested.

With higher specificity in Decision Tree (91.94%) compared to KNN (89.66%), it shows that Decision Tree is more effective in avoiding false positives. This can help in saving costs and increasing the efficiency of marketing campaigns. The higher F1-Score in Decision Tree (88.31%) compared to KNN (85.71%) shows that Decision Tree has a better balance between precision and recall. In this context, a higher F1-Score means that the decision tree provides more consistent performance in identifying students' purchase interest in the MacBook. The higher NPV in Decision Tree (93.44%) compared to KNN (89.66%) shows that Decision Tree is more reliable in predicting students who are not interested in MacBook. This reduces the risk of overestimating the market potential, which can lead to losses. Lower FPR in Decision Tree (8.06%) compared to KNN (10.34%) and lower FNR in Decision Tree (10.53%) compared to KNN (14.29%) indicate that Decision Tree is better at avoiding errors in classification. This is very important because false positives and false negatives can cause significant negative impacts on business decisions.Decision trees are better at handling high data complexity due to their ability to capture non-linearities in the data. KNN, while effective in many cases, struggles in situations where the data has many features interacting in a complex manner. Decision trees tend to perform better with larger datasets because they are able to leverage information from more examples to make more informed decisions. KNN can perform well on small to medium datasets but can experience performance degradation as the dataset size increases due to increased computational costs.

Decision trees have a higher risk of overfitting if not pruned properly. However, with the right pruning techniques, this model can produce excellent results. KNN tends to be more robust against overfitting, but its performance is highly dependent on the correct selection of the K parameter. Decision trees provide a model that is easier to interpret and visualize, which can be an advantage in communicating analysis results to stakeholders. KNN, while intuitive, does not provide clear insight into the decision structure due to its instance-based nature. With higher precision and recall, decision trees can be used to create more effective marketing strategies, targeting college students who are more likely to be interested in MacBooks. Decision trees can also help with more accurate market segmentation, allowing for more personalized and relevant marketing campaigns. With lower FPR and FNR, decision trees help reduce errors in target identification, which can save costs and increase ROI from marketing campaigns. With higher accuracy and NPV, businesses can be more confident in decisions made based on model predictions. The clarity of interpretation of decision tree models can help management understand buying interest patterns and make better strategic decisions. The use of decision trees can support the development of products and offerings that are more in line with customer preferences.

IV. CONCLUSION

This study has compared two classification methods to determine students' purchase interest in MacBook laptops. Based on the results of the analysis of the performance evaluation matrix, several important points can be concluded. The decision tree method shows a higher accuracy (91%) compared to KNN, which has an accuracy of 88%. The Decision Tree method also excels in other metrics such as precision, recall, specificity, and F1-Score, indicating better performance in classifying students' purchase interest. Decision Tree has higher precision (87.18%) and recall (89.47%) than KNN, indicating a better ability to identify students who are truly interested in MacBook laptops and minimize misclassification. The decision tree has a higher specificity (91.94%), indicating a better ability to identify students who are not interested in MacBook laptops. Decision Tree's F1-Score (88.31%) is also higher, indicating a better balance between precision and recall. Decision Tree has a lower FPR (8.06%) and FNR (10.53%) than KNN, indicating that Decision Tree is better at avoiding detrimental misclassifications.

Future research can use larger and more diverse datasets to improve the generalizability of the model and ensure that the results obtained are applicable to a wider population. In addition to KNN and Decision Tree, future research can consider testing other classification methods such as random forest, support vector machine (SVM), and neural networks to evaluate their performance and compare their results. To reduce the risk of overfitting in a decision tree, more sophisticated pruning techniques and hyperparameter tuning can be applied to optimize the model's performance. Further research can conduct feature analysis to understand which features are most influential in determining students' purchasing intentions. This can help improve the interpretability of the model and provide deeper insights. The use of ensemble methods such as bagging and boosting can be investigated to see if a combination of multiple models can improve the accuracy and robustness of predictions.

REFERENCES

[1] D. Arradian, "**MacBook vs Windows: 10 Keunggulan MacBook yang Bikin Penggunanya Klepek-klepek!**," sindonews. [Online]. Available: https://tekno.sindonews.com/read/1345375/123/macbook-vs-windows-10-keunggulan-macbook-yang-bikin-penggunanya-klepek-klepek-1711080183

[2] W. Kurniasih, "**Perbedaan Mac OS dan Windows OS**," gramedia.com. [Online]. Available: https://www.gramedia.com/best-seller/perbedaan-mac-os-dan-windows-os/

[3] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," **Simetris**, vol. 11, no. 2, 2020, doi: 10.24176/simet.v11i2.5482.

[4] D. R. Y. TB, "Analisa Penerapan Data Mining pada Penerimaan Mahasiswa Politeknik Negeri Lhokseumawe Menggunakan Algoritma K-Means," ***J. Informatics Comput. Sci.***, vol. 4, no. 2, pp. 208–216, 2018, [Online]. Available: http://jurnal.uui.ac.id/index.php/jics/article/view/749

[5] E. Mardiani *et al.*, "Membandingkan Algoritma Data Mining dengan Tools Orange untuk Social Economy," **Digit. Transform. Technol.**, vol. 3, no. 2, pp. 686–693, 2023, doi: 10.47709/digitech.v3i2.3256.

[6] P. Kasandra, K. Vadilla, V. Audila, F. Zainal, and A. A. Dermawan, "Prediksi Nilai Emas Menggunakan Algoritma Regresi Linear," ***IESM J.***, vol. 5, no. 1, pp. 152–165, 2024, doi: 10.22303/iesm.1.1.2021.01-10.

[7] K. Silvi Amalia, A. Purbaningrum, and R. Kusumastuti, “Analisis Pengelompokan Data Karyawan Terbaik Perusahaan Menggunakan Algoritma K-Means Clustering,” in **Seminar Nasional Amikom Surakarta (SEMNASA)**, 2023, pp. 11–24.

[8] E. S. Rahmadina, B. Irawan, and A. Bahtiar, “Penerapan Algoritma Fuzzy C-Means pada Data Penjualan Distro,” ***JATI (Jurnal Mhs. Tek. Inform***., vol. 8, no. 2, pp. 2348–2354, 2024, doi: 10.36040/jati.v8i2.8467.

[9] M. F. Haryanti *et al.*, “Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan,” ***J. Manaj. dan Bisnis***, vol. 3, no. 1, pp. 71–90, 2024.

[10] P. Primadona and R. Fauzi, “Penerapan Data Mining Pada Penjualan Produk Elektronik,” Comput. Sci. Ind. Eng., vol. 9, no. 4, 2023, doi: 10.33884/comasiejournal.v9i4.7712.

[11] A. P. Permana, K. Ainiyah, and K. F. H. Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up,” ***JISKa (Jurnal Inform. Sunan Kalijaga)***, vol. 6, no. 3, pp. 178–188, 2021, doi: 10.14421/jiska.2021.6.3.178-188.

[12] D. J. Triani, M. H. Dar, and G. J. Yanris, “Analysis of Public Purchase Interest in Yamaha Motorcycles Using the K-Nearest Neighbor Method,” ***Sink. J. dan Penelit. Tek. Inform***., vol. 7, no. 3, pp. 1238–1254, 2023, doi: 10.33395/sinkron.v8i3.12433.

[13] 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.

[14] S. Sumarlin, “Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM,” ***J. Sist. Inf. Bisnis***, vol. 5, no. 1, pp. 52–62, 2015, doi: 10.21456/vol5iss1pp52-62.

[15] E. Mardiani *et al.*, “Komparasi Metode KNN, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar SMA,” ***Innov. J. Soc. Sci. Res***., vol. 3, no. 2, pp. 13880–13892, 2023, [Online]. Available: http://j-innovative.org/index.php/Innovative/article/view/1949%0Ahttp://j-innovative.org/index.php/Innovative/article/download/1949/1468

[16] S. Maizura, V. Sihombing, and M. H. Dar, “Analysis of the Decision Tree Method for Determining Interest in Prospective Student College,” ***Sink. J. dan Penelit. Tek. Inform***., vol. 7, no. 2, pp. 956–979, 2023, doi: 10.33395/sinkron.v8i2.12258.

[17] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” ***J. Appl. Sci. Technol. Trends***, vol. 2, no. 01 SE-, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[18] R. Surya, M. H. Dar, and F. A. Nasution, “Implementation Of Decision Tree Method To Predict Customer Interest In Internet Data Packages,” ***Int. J. Sci. Technol. Manag***., vol. 5, no. 4, pp. 947–952, 2024, doi: 10.46729/ijstm.v5i4.1155.