Implementation Of K-Nearest Neighbors Algorithm In Analyzing Public Interest In Shoping At Supermarkets

Siti Kholijah[1*], Volvo Sihombing[2], Deci Irmayani[3]

[1,2,3] Faculty of Science and Technology, Universitas Labuhanbatu, Sumatera Utara Indonesia
[*]Corresponding Author:
Email: sitikholijah3004@gmail.com

Abstract.

People's shopping patterns and behaviors continue to develop along with technological advances and lifestyle changes, thus requiring retail business actors, especially supermarkets, to better understand their customers' interests and preferences. In this context, accurate analysis of customer shopping interests is very important to improve customer satisfaction and optimize marketing strategies. One solution that can be implemented to analyze people's shopping interests is the application of the K-Nearest Neighbors algorithm, a simple yet effective nearest neighbor-based classification method for recognizing patterns from existing data. This study aims to apply the K-Nearest Neighbors algorithm to classify people's interest in shopping at supermarkets. This study also evaluates the effectiveness and performance of the algorithm in the context of business decision-making in the retail sector. The research methodology includes collecting data on people's shopping interests, data pre-processing, implementing the K-Nearest Neighbors algorithm, and evaluating model performance using evaluation metrics such as accuracy, precision, recall, and F1-score. The results of this study indicate that the K-Nearest Neighbors algorithm is able to achieve an accuracy of 88%, with precision, recall, and F1-score all reaching 92.86%. These results indicate that the K-Nearest Neighbors model is very effective in classifying people's shopping interests, with a low error rate. The resulting confusion matrix also shows the model's ability to identify customers who are interested in shopping with little prediction error. This study concludes that we can rely on the K-Nearest Neighbors algorithm to analyze people's shopping interests in supermarkets. This model not only shows good performance in classification but also has great potential to be implemented in recommendation systems and customer segmentation in the real world. This study contributes to the development of consumer behavior analysis methods in the retail sector, as well as providing a basis for further research to explore other algorithms or combinations of techniques to improve the accuracy and effectiveness of classification models.

Keywords: *Classification, K-Nearest Neighbors, Public Interest,and Shoping Supermarket.*

I. INTRODUCTION

Brastagi Supermarket, located in Rantauprapat, Labuhanbatu Regency, is known as one of the main shopping centers in the area. This supermarket offers a wide range of daily necessities, from fresh food such as local fruits to household products and other daily supplies. With complete product availability and competitive prices, Brastagi Supermarket plays an important role in advancing the local economy and meeting consumer needs well. Brastagi Supermarket is known not only for its diverse product choices but also for its high standards of cleanliness. By prioritizing the cleanliness of the room and products, this supermarket provides a guarantee of the quality and safety of the goods sold to customers. In terms of hygiene, the products offered are always kept clean, be they fresh food, packaged food, or other household goods. In addition, friendly and professional service is another advantage of Brastagi Supermarket. The staff are trained to provide assistance to customers politely and responsively, creating a pleasant and comfortable shopping atmosphere for all visitors. The combination of quality products, maintained room cleanliness, and friendly service makes Brastagi Supermarket the main choice for the people of Rantauprapat and its surroundings in meeting their daily needs. Although Brastagi Supermarket is known for its high standards of cleanliness and service, it is undeniable that some shoppers occasionally experience less than satisfactory experiences.

Some complaints that may arise include the quality of certain products that are less fresh or wilted, which may affect the customer's shopping experience. Nevertheless, Brastagi Supermarket management continues to strive to improve quality control to ensure that all products available are sold in the best condition. It is important to remember that shopping experiences can vary from customer to customer and from time to time. Brastagi Supermarket management is always open to customer feedback and is committed to continuously improving service standards and product quality. By improving and managing customer feedback, Brastagi Supermarket aims to maintain its reputation as a reliable and quality shopping destination in Rantauprapat and surrounding areas. From the information that has been presented previously, it is clear that Brastagi Supermarket has a fairly strong reputation in Rantauprapat. The positive side includes the variety of products offered, the cleanliness of the room, and the friendly service. This makes this supermarket the main choice for many people for meeting their daily needs. However, as highlighted, there is also a negative side, which includes some complaints about the quality of certain products or less than satisfactory shopping experiences for some buyers.

To better understand the dynamics of public interest as buyers at Brastagi Supermarket, research can focus on collecting data on consumer preferences, their perceptions of product and service quality, and the factors that influence their purchasing decisions. By using methods such as surveys or direct interviews with customers, this research can produce valuable insights for supermarket management on improving the shopping experience and building stronger relationships with their loyal customers.Research on the public interest as buyers at Brastagi Supermarket will be conducted using the K-Nearest Neighbors (KNN) method [1], [2]. The KNN method was chosen because of its simplicity and effectiveness in classifying data based on the proximity between new data samples and existing data [3], [4]. In the context of this research, KNN will be used to analyze the public's shopping patterns and preferences by grouping respondents based on certain characteristics, such as shopping frequency, types of products purchased, and levels of satisfaction with the quality and service at Brastagi Supermarket [5]. By applying the KNN method, this study will identify consumer groups with similar preferences, making it possible to understand the factors that most influence shopping interest at Brastagi Supermarket. The results of this analysis are expected to provide a clearer picture of the different market segments and their specific needs, which in turn can help supermarket management design more effective marketing strategies and improve service quality in accordance with customer expectations.

II. METHODS

Research on public interest in Brastagi Supermarket will be conducted using the K-Nearest Neighbors (KNN) method within a data mining framework, with stages that follow the Knowledge Discovery in Databases (KDD) process [6], [7]. The KDD process begins with the data selection stage, where relevant data on public preferences and shopping behavior is collected from various sources [8]. Next, the data pre-processing stage is carried out to clean and integrate the collected data, ensuring that the data is free from noise and ready to be analyzed. The preprocessing stage is the stage carried out to select the data that will be used in this study. With preprocessing, the problems that will arise will also be small. So for this stage, the data will be truly selected based on the eligibility of the data that has been set or become the provisions in this study.

Data transformation is then carried out to change the raw data into a format that is suitable for further analysis. The transformation stage is the stage carried out to change the data format used in this study, adjusting to the needs of this study. So in this study, the data used, namely Excel data, will be changed according to the provisions and needs of the study. After the data is ready, the data mining stage using the KNN method is carried out to identify patterns and trends in people's shopping preferences [9], [10]. The KNN method will group consumers based on their similar characteristics, such as shopping frequency, types of products often purchased, and levels of satisfaction with Brastagi Supermarket services. The final stage of KDD is the evaluation and interpretation of the patterns found, which will provide valuable insights into the factors that influence people's interests. The results of this study are expected to help Brastagi Supermarket develop more effective marketing strategies and improve services to better meet customer needs and preferences.

III. RESULT AND DISCUSSION

The selection stage is a stage used to collect the data that will be used in this study. So this study will use data from people who are buyers at Brastagi Supermarket. For the data used in this study, there are 2 data sets: the first is training data, and the second is testing data. Training data in data mining is a subset of data used to train the model to recognize patterns and make accurate predictions. This data contains examples that have been labeled, which helps the algorithm understand the relationship between input and expected output. The training data is shown in Table 1.

Table 1. Training Data

Respondents	Location	Product Quality	Service	Cleanliness	Parking	Category
Respondent01	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent02	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent03	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent04	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent05	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent06	Easy to Reach	Not Good	Not Friendly	Not Clean	Narrow	Not Interested
Respondent07	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow	Not Interested
Respondent08	Hard to Reach	Not Good	Friendly	Not Clean	Narrow	Not Interested
Respondent09	Easy to Reach	Good	Not Friendly	Clean	Wide	Interest
Respondent10	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent11	Hard to Reach	Good	Friendly	Clean	Wide	Interest
Respondent12	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent13	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow	Not Interested
Respondent14	Hard to Reach	Good	Not Friendly	Not Clean	Narrow	Not Interested
Respondent15	Easy to Reach	Good	Friendly	Clean	Narrow	Interest
Respondent16	Easy to Reach	Good	Friendly	Not Clean	Wide	Interest
Respondent17	Hard to Reach	Not Good	Not Friendly	Not Clean	Wide	Not Interested
Respondent18	Easy to Reach	Not Good	Friendly	Clean	Wide	Interest
Respondent19	Easy to Reach	Good	Friendly	Clean	Wide	Interest
Respondent20	Hard to Reach	Not Good	Not Friendly	Clean	Narrow	Not Interested

Table 1 above is the training data that will be used to help the data classification process, so the training data above will later train the testing data to obtain the classification results. For the training data used, there are 20 respondent data points.

Table 2. Testing Data

Respondents	Location	Product Quality	Service	Cleanliness	Parking
Respondent01	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent02	Easy to Reach	Good	Friendly	Clean	Narrow
Respondent03	Easy to Reach	Good	Friendly	Clean	Wide
Respondent04	Easy to Reach	Good	Friendly	Clean	Wide
Respondent05	Hard to Reach	Good	Friendly	Clean	Wide
Respondent06	Easy to Reach	Good	Friendly	Clean	Wide
Respondent07	Easy to Reach	Good	Friendly	Clean	Wide
Respondent08	Hard to Reach	Not Good	Not Friendly	Clean	Narrow
Respondent09	Easy to Reach	Not Good	Friendly	Clean	Wide
Respondent10	Hard to Reach	Good	Not Friendly	Not Clean	Narrow
Respondent11	Easy to Reach	Good	Friendly	Clean	Wide
Respondent12	Easy to Reach	Good	Not Friendly	Clean	Wide
Respondent13	Hard to Reach	Not Good	Friendly	Not Clean	Narrow
Respondent14	Easy to Reach	Good	Friendly	Clean	Wide
Respondent15	Easy to Reach	Good	Friendly	Clean	Wide
Respondent16	Easy to Reach	Good	Friendly	Clean	Wide

Respondent17	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent18	Easy to Reach	Good	Friendly	Not Clean	Wide
Respondent19	Easy to Reach	Good	Friendly	Clean	Wide
Respondent20	Easy to Reach	Good	Friendly	Clean	Wide
Respondent21	Easy to Reach	Not Good	Friendly	Clean	Wide
Respondent22	Easy to Reach	Good	Friendly	Clean	Wide
Respondent23	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent24	Easy to Reach	Good	Friendly	Clean	Wide
Respondent25	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent26	Easy to Reach	Good	Friendly	Clean	Wide
Respondent27	Easy to Reach	Good	Friendly	Clean	Wide
Respondent28	Easy to Reach	Good	Friendly	Clean	Wide
Respondent29	Easy to Reach	Good	Friendly	Clean	Narrow
Respondent30	Easy to Reach	Good	Friendly	Clean	Wide
Respondent31	Easy to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent32	Easy to Reach	Good	Friendly	Clean	Wide
Respondent33	Easy to Reach	Good	Friendly	Clean	Wide
Respondent34	Easy to Reach	Good	Friendly	Not Clean	Wide
Respondent35	Easy to Reach	Good	Friendly	Clean	Wide
Respondent36	Easy to Reach	Good	Friendly	Clean	Wide
Respondent37	Easy to Reach	Good	Friendly	Clean	Wide
Respondent38	Hard to Reach	Good	Friendly	Clean	Wide
Respondent39	Easy to Reach	Good	Friendly	Clean	Wide
Respondent40	Easy to Reach	Good	Friendly	Clean	Wide
Respondent41	Easy to Reach	Good	Friendly	Clean	Wide
Respondent42	Easy to Reach	Good	Friendly	Clean	Wide
Respondent43	Easy to Reach	Good	Friendly	Clean	Wide
Respondent44	Easy to Reach	Good	Friendly	Clean	Wide
Respondent45	Easy to Reach	Good	Not Friendly	Clean	Wide
Respondent46	Easy to Reach	Good	Friendly	Clean	Wide
Respondent47	Hard to Reach	Not Good	Not Friendly	Not Clean	Wide
Respondent48	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent49	Hard to Reach	Not Good	Not Friendly	Not Clean	Narrow
Respondent50	Easy to Reach	Good	Friendly	Clean	Wide

Table 2 shows the testing data used in this study. The data used is 50 sample data points. Testing data is sample data used to be classified using the KNN method. So this data will later be processed in data mining using the KNN method.

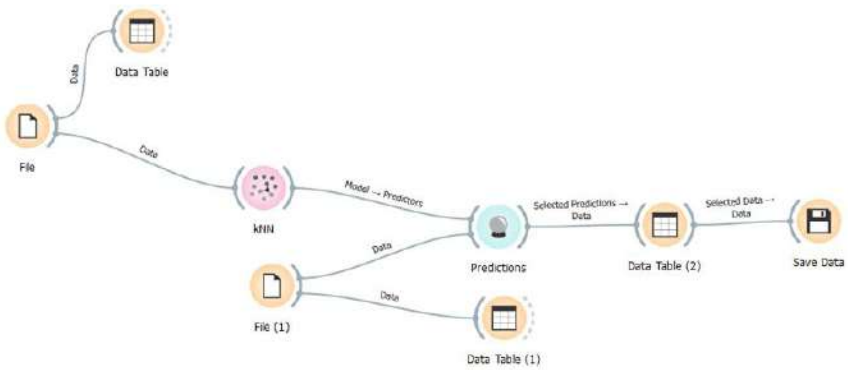

Fig 1. Design of a Data Classification Model

Figure 1 is a workflow that illustrates the data classification process using the KNN algorithm to analyze people's shopping interests in supermarkets. The figure shows two nodes labeled "File" and "File (1)," respectively. These nodes represent raw data sources imported from external files, which are likely to be shopping history data or supermarket customer preferences. This data is then converted into a data table so that it can be processed further. After the data is imported, the file is converted into a "data table," which means that the data has been organized into a table for further analysis. In this table, each row may represent one sample (e.g., one customer), and each column represents a feature or attribute (e.g., shopping frequency, products purchased, etc.). The "kNN" node is the core of the classification process, where the KNN algorithm is applied. The KNN algorithm works by identifying several nearest neighbors of a sample based on feature distance, then making predictions based on the majority of the nearest neighbor classes. At this stage, the kNN model is trained using the imported data. Data from the "Data Table" was previously used to build a KNN model that aims to predict the class or category of people's shopping interests.

After the KNN model is trained, the "Predictions" node is used to generate predictions for new data or data whose class is unknown. These prediction results are the output of the KNN model, indicating where the new data sample should be classified based on its features. The figure shows two different "Data Table" nodes, namely "Data Table (1)" and "Data Table (2)". "Data Table (1)" may store the original unprocessed data, while "Data Table (2)" contains the data that has been predicted by the KNN model. This node is used to view or analyze the prediction results and compare them with the original data. The last node is "Save Data," which indicates that the prediction results are saved in a file. This allows the user to save the prediction data that can be used for further analysis or reporting research results. This workflow illustrates the complete classification process from input data to saving the prediction results, with an emphasis on using the KNN algorithm to understand people's shopping interests in supermarkets.

Predicted

Actual	Interest	Not Interested	∑
Interest	39	3	42
Not Interested	3	5	8
∑	42	8	50

Fig 2. Confusion Matrix

Figure 2 shows the confusion matrix used to measure the performance of the classification model, in this case, the KNN algorithm. Let's explain each part of this confusion matrix in full and in detail and do an in-depth analysis. The "Predicted" column shows the prediction results of the KNN model. There are two prediction categories: interest and not interest. The left row of the matrix shows the actual label or real class of the data being tested, which is also divided into two categories: interest and not interest. True Positive (TP) is the number of cases where the model predicts "interest," and in fact it is "interest.". This means that out of 42 cases that are truly interesting, 39 of them are correctly predicted by the model. False Negative (FN) is the number of cases where the model predicts "not interested," but in reality it is "interested.". There are three cases where the model incorrectly predicts customers who are actually interested as not interested. False Positive (FP) is the number of cases where the model predicts "interest," but in reality it is "not interested." There are 3 cases where customers who are actually not interested are predicted by the model to be interested. True Negative (TN) is the number of cases where the model predicts "not interested," and in fact it is "not interested." There are 5 cases where the model correctly predicts that a customer is not interested.

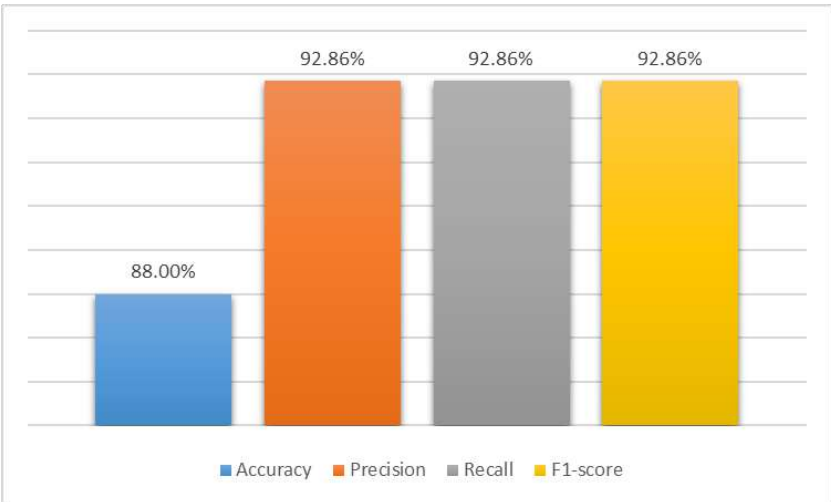

Fig 3. Evaluation Matrix

Figure 3 shows the performance evaluation matrix of the KNN algorithm in analyzing people's interest in shopping at supermarkets. This matrix is used to assess the performance of the KNN classification model by displaying four main evaluation metrics: accuracy, precision, recall, and F1-score. Accuracy is a measure that shows the percentage of total correct predictions compared to the overall predictions made by the model. In this context, an accuracy of 88% indicates that the KNN model successfully classifies 88% of all tested cases. Although 88% accuracy is considered quite good, it also means that there are 12% wrong predictions. However, accuracy alone is not enough to fully understand the performance of the model, especially if the dataset is imbalanced. Precision measures the accuracy of the positive predictions made by the model, namely how many of all samples predicted as "interest" are actually "interest.". High precision (92.86%) indicates that the model rarely provides incorrect "interest" predictions. High precision indicates that the KNN model is very good at identifying customers who are truly interested in shopping at supermarkets, and only a few cases are incorrectly predicted as interested when in fact they are not. Recall measures the ability of the model to find all true positive samples, i.e., how many of all samples that are truly “interesting” were successfully found by the model. A recall value of 92.86% indicates that the model is able to identify most of the interested customers.

The high recall indicates that the KNN model is effective in detecting almost all of the true “interest” cases, with only a few cases missed. F1-score is a measure of the harmony between precision and recall. It is an important metric when we want to consider both simultaneously, especially in the context of an imbalanced dataset. An F1-score of 92.86% indicates a good balance between precision and recall. A high F1-score indicates that the KNN model has a solid and balanced performance in identifying customers who are interested in shopping, both in terms of prediction accuracy and detection coverage.From the research results, we can see that the KNN model used in this study shows very good performance with evaluation metrics that are close to perfect, except for slightly lower accuracy. The high precision, recall, and F1-score indicate that the model is very effective in classifying customer shopping interests in supermarkets. The difference between the accuracy value (88%) and other metrics (92.86%) may indicate that although the model has good performance in detecting the "interest" class, there are still some cases that may cause a decrease in accuracy. This could be due to the small number of "not interested" cases, which affect the final accuracy results. In the case of an imbalanced dataset, precision, recall, and F1-score become more representative metrics than accuracy. With equally high precision, recall, and F1-score, this model shows a good balance between avoiding false positives and identifying all true positives.

This is especially important in applications where misclassification can have negative impacts, such as in shopping recommendations. Although 88% accuracy is still considered good, there is room for improvement, for example, through adjusting KNN parameters, using feature selection techniques, or even combining KNN with other algorithms to improve the prediction of more difficult classes. This evaluation matrix provides an excellent performance picture of the KNN model in classifying people's shopping interests in supermarkets. The high precision, recall, and F1-score values indicate that this model is very reliable in predicting shopping interest, with a good balance between detecting interested customers and avoiding false positive predictions. 88% accuracy is a strong indicator, although slightly lower than other metrics, but this does not reduce the reliability of the model in the tested application.

IV. CONCLUSION

This study aims to apply the KNN algorithm to analyzing people's interest in shopping at supermarkets. The KNN algorithm used in this study showed very good performance, with an accuracy of 88% and precision, recall, and F1-score all reaching 92.86%. This shows that KNN is effective in classifying people's shopping interest data at supermarkets. These results reflect that this model is able to distinguish well between customers who are interested and not interested in shopping, making it a reliable tool for consumer behavior analysis in this context. The resulting confusion matrix indicates that the model is able to identify the majority of customers who are interested in shopping with few errors in classification. There are only a few cases of false positives and false negatives, indicating that the model rarely makes errors in positive and negative classification. The reliability of this model is mainly seen in its high precision, which means that the model makes very few errors in identifying customers who are truly interested. With evaluation results showing high performance, the KNN algorithm can be implemented as part of a recommendation system or customer segmentation in supermarkets.

This can help in making business decisions related to marketing strategies and more targeted product offerings.Although the evaluation metrics show good results, it is important to consider the possibility of class imbalance in the dataset used. Further research can explore techniques such as oversampling or undersampling to address data imbalances, as well as the use of additional metrics such as ROC-AUC for a more holistic evaluation. To improve the accuracy of the already good model, researchers can explore techniques such as feature selection or dimensionality reduction (e.g., using PCA) to further improve the model's performance. In addition, adjusting KNN parameters, such as the number of nearest neighbors (k), can also be explored to find a more optimal configuration. Future research can expand the scope of the data by including other variables that may affect people's shopping intentions, such as promotions, customer loyalty, or online shopping behavior. With richer data, predictive models can be more accurate and provide deeper insights.

REFERENCES

[1] H. A. Pratama, G. J. Yanris, and M. N. S. Hasibuan, "Implementation of Data Mining for Data Classification of Visitor Satisfaction Levels," ***Sink. J. dan Penelit. Tek. Inform***., vol. 8, no. 3, pp. 1832–1851, 2023, doi: 10.33395/sinkron.v8i3.12674.

[2] P. Arsi, L. N. Hidayati, and A. Nurhakim, "Komparasi Model Klasifikasi Sentimen Issue Vaksin Covid-19 Berbasis Platform Instagram," ***J. Media Inform. Budidarma***, vol. 6, no. 1, p. 459, 2022, doi: 10.30865/mib.v6i1.3509.

[3] M. Ula, R. Zulhusna, R. Putra Fhonna, and A. Pratama, "Penerapan Model Klasifikasi K Nearest Neighbor Dalam Pencarian Kesesuaian Pekerjaan," ***Metik J***., vol. 6, no. 1, pp. 18–23, 2022, doi: 10.47002/metik.v6i1.343.

[4] B. Hssina, A. Grota, and M. Erritali, "Recommendation system using the k-nearest neighbors and singular value decomposition algorithms," ***Int. J. Electr. Comput. Eng.***, vol. 11, no. 6, pp. 5541–5548, 2021, doi: 10.11591/ijece.v11i6.pp5541-5548.

[5] T. A. Assegie, "An optimized K-Nearest neighbor based breast cancer detection," ***J. Robot. Control***, vol. 2, no. 3, pp. 115–118, 2021, doi: 10.18196/jrc.2363.

[6] A. N. Z. Hidayah and A. F. Rozi, "Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus : Universitas Mercu Buana Yogyakarta)," ***J. Inf. Syst. Artif. Intell.***, vol. 1, no. 2, pp. 117–127, 2021, doi: 10.26486/jisai.v1i2.24.

[7] B. S. Pranata and D. P. Utomo, "Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," ***Bull. Inf. Technol***., vol. 1, no. 2, pp. 83–91, 2020.

[8] W. Cholil, A. R. Dalimunthi, and L. Atika, "Model Data Mining Dalam Mengidentifikasi Pola Laju Pertumbuhan Antar Sektor Ekonomi di Provinsi Sumatera Selatan dan Bangka Belitung," **Teknika**, vol. 8, no. 2, pp. 103–109, 2019, doi: 10.34148/teknika.v8i2.181.

[9] R. Ali, M. M. Yusro, M. S. Hitam, and M. Ikhwanuddin, "Machine Learning With Multistage Classifiers For Identification Of Of Ectoparasite Infected Mud Crab Genus Scylla," **Telkomnika (Telecommunication Comput. Electron. Control**., vol. 19, no. 2, pp. 406–413, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16724.

[10] I. P. Putri, "Analisis Performa Metode K- Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular," ***Indones. J. Data Sci***., vol. 2, no. 1, pp. 21–28, 2021, doi: 10.33096/ijodas.v2i1.25.

[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.