Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

Comparison of the K Mean Algorithm with the C 45 Algorithm in Dataming Applications

Erwin Panggabean[1)*],Agustina Simangunson[2)],Dedi Sinaga[3)],Agus Putra Emas Sihombing[4)],Tri Evalina Aritonang[5)]
[1,2,3,4,5)]STMIK Pelita Nusantara Medan, Sumatera Utara, Indonesia
[1)]erwinpanggabean9@gmail.com,[2)]agustinasimangunsong93@gmail.com,[3)]dedisinaga27@gmail.com,
[4)]agusputraemassihombing17@gmail.com, [5)]trievalinaaritonang@gmail.com

ABSTRACT

This research topic discusses the comparison of the K-Means and C4.5 algorithms in the application of data mining to predict aquarium sales in a company. K-Means is a clustering algorithm that functions to group data based on similarity, for example grouping customers based on frequency or type of purchase. This helps companies understand market segments and design marketing strategies accordingly. Meanwhile, C4.5 is a classification algorithm that builds decision trees based on important attributes that influence sales, such as price, season, or promotions. This algorithm is able to predict sales categories, such as increases or decreases, based on historical data. By comparing these two algorithms, the research sought to find out which algorithm is more effective in helping companies predict sales and make strategic decisions. A combination of the two can also be used, with K-Means grouping the data first, then C4.5 classifying each segment formed. These results can provide more accurate sales predictions and more effective marketing strategies. This research is important to understand the effectiveness of algorithms in data mining to improve business decision making.

Keywords: K-Means, C4.5, Data Mining, Sales Prediction, Clustering and Classification Algorithms.

INTRODUCTION

In the increasingly rapidly developing digital era, data has become an important asset for companies to make strategic decisions. Effective use of data can provide a competitive advantage through in-depth and predictive analysis. One of the methods used in data analysis is data mining, namely the process of extracting useful information from large and complex data. Through data mining, hidden patterns or knowledge in data can be discovered to provide new insights for decision making. K-Means algorithm and C4.5 algorithm are two popular algorithms in data mining that are used for different purposes. The K-Means algorithm is a clustering method used to group data based on similarity, where similar data will be grouped in one cluster. This algorithm is often used in market segment analysis, determining customer targets, and product recommendations. On the other hand, the C4.5 algorithm is a decision tree algorithm used for classification, and is able to form a model that can explain the relationship between input variables and output variables. This algorithm is effectively used in classifications such as fraud detection, disease prediction, and credit analysis. In this research, a comparison was carried out between the K-Means algorithm and the C4.5 algorithm to see the effectiveness and performance of the two methods in applying data mining. This comparison aims to identify which algorithm is more optimal in producing accurate and relevant analysis results in accordance with research objectives, such as predicting product sales. Aspects to be compared include accuracy, computing time efficiency, and interpretability of results. With the results of this comparison, it is hoped that it can provide recommendations for researchers or practitioners to choose the right algorithm according to the needs and characteristics of the data they have.

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

LITERATURE REVIEW

Based on the results of applying data mining to lecturer performance assessment data based on student satisfaction using the K-Means algorithm, it can be concluded: The K-Means algorithm can be implemented in the clustering process using the Rapid Miner tool. Cluster testing is grouped into 4 clusters, namely: very good, good, quite good, and not good. In general, the assessment of lecturer performance by students showed that the number of assessment clusters was very good 312 (31.74%) student data, good assessment clusters 401 (40.79%) student data, quite good assessment clusters 189 (19.23%) student data , and the poor assessment cluster was 81 (8.24%) student data. So that a greater amount of student data is found in the good assessment cluster. The evaluation results using the Davies-Bouldin Index were obtained at 0.270 or 27%. So the level of accuracy of the cluster results is good, because the DBI value will be close to zero (Dewi, Purnama, and Utami 2022). Several conclusions were obtained based on the results of the analysis and discussion that the researchers had carried out. The following is the conclusion. Clustering was carried out using K-Means based on 5 target attributes, namely School Type, School Department, Entrance Test Score, Semester 1 and 2 Achievement Index on data at the STMI Jakarta Polytechnic, resulting in information describing student characteristics according to study program. Clustering carried out on the Automotive Business Administration (ABO) study program data resulted in three student clusters from a total of 355 items. Clustering carried out on the Automotive Industrial Information Systems (SIIO) study program data resulted in two student clusters from a total of 349 items. Clustering carried out on data from the Automotive Industrial Engineering (TIO) study program resulted in two student clusters from a total of 233 items. Clustering was carried out on data from the Polymer Chemical Engineering (TKP) study program. produced six student clusters from a total of 139 items (Isnanto and Widodo 2021). Based on the results of data grouping which was carried out using the K-Means method using default parameters, the smallest results were obtained, namely in cluster k-3 with a DBI value of 0.530. So cluster k-3 is the best cluster out of a total of k of 20. With a total of 1972 items in cluster 0, 1612 items in cluster 1, and 140 items in cluster 2 (Permadi and Wiyaja 2023). Based on the results of the research and discussion, several conclusions were obtained, namely that data mining using the K-means Clustering algorithm which is implemented in perfume sales, can be used by IM Perfume Stores as a method to determine the direction of its marketing policy. The results of perfume sales analysis obtained from this system show that the data is most frequently requested by consumers and is welcomed by teenagers and adults. From the clustering results, three clusters were obtained, the first letter cluster (C0) is the best-selling perfume group, the second cluster (C1) is the best-selling perfume group and the second letter three cluster (C2) is the unsold fragrance group. The cluster results obtained from the K-Means algorithm show that group C0 of the best-selling products includes 9 products, group C1 of the best-selling products includes 3 products, and group C2 includes 13 products. The product is not selling well (Riadi and Mesran 2023). Based on the research conducted, it can be concluded that data mining used with the C45 algorithm method can be applied. With this research, the results of the algorithm obtained from the data obtained can provide recommendations for decision makers so that the Head of Regei Sustainable Village can provide BLT assistance evenly based on the model provided (Bachtiar and Mahradianur 2023).

METHOD

K Mean Algorithm

The K-Means algorithm is a clustering method that divides data into a number of groups or clusters based on similarities between the data. This algorithm works by initializing a central point (centroid) for each cluster, then grouping data into these clusters by calculating the distance between each data and the

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

nearest centroid. This process repeats until convergence, namely when the centroid position does not experience significant changes or there is no further data movement between clusters (Aggarwal 2020; Reddy and Vinzamuri 2021).
Following are the main steps in the K-Means algorithm:
1. Centroid Initialization: Randomly select a starting point as the centroid for each cluster.
2. Data Clustering: Calculates the distance of each data to the nearest centroid using a metric, such as Euclidean distance, and groups the data into clusters with the nearest centroid.
3. Centroid Update: Recalculates the centroid position for each cluster based on the average of all data in that cluster.
4. Repeat Until Convergence: Repeats the centroid grouping and updating process until the centroid position does not change or the change is very small.

The advantage of the K-Means algorithm is its speed and simplicity, which makes it effective for large datasets. However, this algorithm also has several weaknesses, such as being sensitive to the initial centroid selection and having difficulty grouping data that is non-spherical or has varying sizes and densities (Sharma and Kumar 2020; Mahapatra et al. 2022).

$$d(x_i, c_j) = \sqrt{\sum_1^n (xij - cjk)^2} \quad \text{.....1}$$

Where:
d(xi,cj) is the distance between data point xi and centroid cj.
xik is the kkkth feature value of data point xi.
cjk is the kkk feature value of the centroid cj.
n is the number of features in the data.

Calculating objective function(SSE-Sum of Square Error):

$$SSE = \sum_{j=1}^{k} \sum_{xi \epsilon cj} d(xi, cj)^2 \quad \text{.....2}$$

Where:
SSE = Total squared error in all Clusters
d(xi,cj) = Euclidean distance between data points xi and cj

C 45 algorithm

The C4.5 algorithm is a decision tree-based machine learning algorithm developed to perform classification. This algorithm is a development of the ID3 (Iterative Dichotomiser 3) algorithm and is used to produce decision trees from datasets with various attributes. C4.5 works by selecting the most informative attributes as decision nodes at each level of the tree, using the information gain ratio criterion to determine the best attribute. This algorithm overcomes several limitations of ID3, including handling attributes with continuous values and dealing with missing data (Patel and Patel 2020; Kaur and Arora 2021).

Following are the main steps in the C4.5 algorithm:
1. Best Attribute Selection: The algorithm calculates the gain ratio for each attribute and selects the attribute with the highest gain ratio as the decision node.
2. Dataset Splitting Based on Attributes: Once the best attributes are selected, the data is split into subsets based on the values of those attributes.

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

3. Handling Continuous Values and Missing Values: The C4.5 algorithm can handle attributes with continuous values by finding the best dividing point and handling missing values with the estimation method.
4. Branch Formation: For each subset, the process is repeated by selecting the best attribute to further split the data until all data in the subset has the same class or no attributes remain.
5. Tree Pruning: To reduce overfitting, C4.5 also performs pruning by removing less significant nodes based on a certain error size or confidence level.

The advantages of C4.5 are its ability to handle continuous data, missing data, and its ability to produce more stable decision trees. However, C4.5 can be slow on large datasets, especially if the dataset has many attributes or attributes with many values (Singh et al. 2022; Zhang and Wei 2023).The C4.5 algorithm is a decision tree formation algorithm used for classification. This algorithm works by dividing data based on attribute values that provide maximum information at each stage of tree formation. Following are some of the main formulas in the C4.5 algorithm. Entropy measures the uncertainty or impurity of a data set and is used to determine how well an attribute divides the data. For an attribute A with class nnn, entropy is defined as:

$$\text{Entropy}(S) = -\sum_{i=1}^{n} pi \text{ Log}_2 pi \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

Where:
S is the data set being evaluated
pi is the probability or proportion of I in data S
n is the number of classes in data S
Information Gain is a reduction in entropy produced by dividing a dataset based on an attribute. The attribute with the highest Information Gain value is selected as the divider:

$$\text{Gain}(S,A) = \text{Entgropy} - \sum_{u \in \text{Value}(A)} \frac{|Sv|}{|S|} \text{ x Entropy}(Sv) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

Where:
S is a data set.
A is the attribute being evaluated
Value(A) is the set of unique values of attribute A.
Su is a subset of S where attribute A has a value v.
|Sv|and |S| is the size of the subset Sv and the dataset S as a whole.
The C4.5 algorithm uses the Gain Ratio to improve the Information Gain so that it is not biased towards attributes with many unique values. Gain Ratio takes into account Split Information from attributes:

$$\text{Gain Ratio}(S,A) = \frac{\text{Gain}(S,A)}{\text{Split Information}(S,A)}, \text{dengan Split Information}(S,A) = -\sum_{v \in \text{Values}(A)} \frac{|Sv|}{|S|} \log 2 \frac{|S|}{|S|} \ldots 5$$

Sales Patterns
Sales patterns refer to the analysis of patterns or trends in sales data to understand customer behavior, market trends, and factors that influence the increase or decrease in sales of a product. This sales pattern is usually used to help companies in designing marketing strategies, inventory planning, and making decisions regarding products and services. The following are some of the main concepts in sales patterns and their references:

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

1. Seasonal Patterns

Seasonal patterns refer to sales fluctuations that occur during certain periods of the year. This pattern is often influenced by factors such as weather, holidays, and holidays (Grewal, D., & Roggeveen, A. L.., 2020).

2. Cyclical Patterns

Cyclical patterns refer to sales fluctuations that are influenced by economic cycles, such as recessions or economic expansions. Sales can fluctuate longer than seasonal patterns (Donthu, N., & Gustafsson, A. 2020).

3. Cohort Analysis

Cohort analysis involves grouping customers based on the time they first made a purchase to study changes in their behavior over time (D'Haen, J., & Van den Poel, D. (2019). S.

RESULT

K-Means Clustering algorithms

The results of the two K-Means Clustering algorithms and the C 4.5 algorithm in their application to datamining are as follows:

Data used:

No.	Code	Name	Size	Type	Price	Month-Year	Prediction
1	A01	Akuarium	60	Tanaman air	450000	Maret 2024	……..?
2	V03	Vivarium	40	Darat & air	400000	Maret 2024	……..?
3	P02	Paludarium	50	Rawa	400000	Januari 2024	……..?

To apply the K-Means algorithm, numeric attributes are required. In this data, the relevant attributes are Size and Price. If necessary, categorical data such as Type can be encoded into numerical form using encoding methods. The K-Means algorithm works by grouping data into several clusters based on the Euclidean distance between the data and the center point (centroid) of each cluster. The K-Means process can be explained as follows:

Select the Number of Clusters (K): For example, we choose K=2 to group the data into two clusters: products that tend to be "Constant" and "Smooth".

Centroid Initialization: Determine two initial center points (centroids) randomly.

Iteration: Iterate to calculate the distance of each data to the centroid, then assign the data to the closest cluster. Then update the centroid until the centroid position is stable or reaches the maximum number of iterations.

After clustering is carried out, each product will be grouped into one of two clusters. The following are hypothetical clustering results based on the Size and Price attributes:

Cluster 1: Products with smaller prices or sizes, may tend to "Stuck".

Cluster 2: Products with larger prices or sizes, may tend to be "Smooth".

After clustering is complete, we can compare the prediction results with the labels provided by the C4.5 Prediction algorithm to validate whether the clusters formed match the expected sales pattern (congested or smooth).

Initial Grouping Based on Closest Distance

Based on the distance calculation results:

Data 1 (Aquarium) is closer to Centroid 1 (distance = 0).

Data 2 (Vivarium) is closer to Centroid 2 (distance = 0).

Data 3 (Paludarium) is closer to Centroid 2 (distance = 10).

Initial Cluster Results:

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

Cluster 1: Data 1 (Aquarium)
Cluster 2: Data 2 (Vivarium) and Data 3 (Paludarium)

Update Centroid

After initial clustering, we update the centroid position based on the average of points in each cluster.
1. New Centroid 1 for Cluster 1 (Data 1):
Size=60, Price=450000
2. New Centroid 2 for Cluster 2 (Data 2 and Data 3):
Average Size = (40+50)/2=45
Average Price = (400000+400000)/2=400000
New Centroid 2: (Size = 45, Price = 400000)
Re-Iteration
Calculation of the distance to the new centroid until the centroid is stable (does not change anymore). This process is called iteration in K-Means, which will be repeated until the data in the cluster does not change any more. K-Means: Based on Size and Price, K-Means will group products into two clusters, for example, products with large sizes and high prices in one cluster, and products with smaller sizes and lower prices in the other cluster. However, K-Means does not consider the “Current” or “Congestion” labels directly.

C 45 Algorithm in the Application of Dataming

That the C4.5 algorithm is a popular decision tree algorithm for classification. This algorithm builds a decision tree based on the attributes that provide the most information to separate data into certain categories. In this case, it will be used to predict whether a product has a sales status of "Sluggish" or "Current" based on existing attributes. Before applying the C4.5 algorithm, we need to do some preparatory steps:
Category Attribute Encoding: The Type and Month-Year attributes need to be converted into a numeric or categorical form that can be read by the model. For example:
Type: “aquatic plants” = 0, “land & water” = 1, “swamp” = 2.
Month-Year: Use months or convert them into seasonal categories, such as “Jan-Mar” = 1, “Apr-Jun” = 2, and so on.
Data Splitting: Use Size, Type, Price, and Month-Year as input attributes (features) to predict the Predicted_Sales columnC4.5.
Building a Decision Tree with the C4.5 Algorithm
The C4.5 algorithm works by selecting the attribute that has the highest Information Gain value to divide the data, then forming a tree structure based on this attribute. The attribute with the highest Information Gain will be selected as the root of the tree, and this process continues recursively for subsequent branches until the tree is complete. C4.5: Based on existing data and the C4.5 Sales_Prediction label, the C4.5 algorithm will generate rules to determine whether product sales are "Smooth" or "Sluggish" by considering Size, Price, Type, and possibly Month-Year. For example, C4.5 can generate a rule like: If Size > 50 and Price > 400000, then Sales Prediction is " Sluggish ".

No.	**Code**	**Name**	**Size**	**Type**	**Price**	Month-Year	Prediction
1	A01	Akuarium	60	Tanaman air	450000	Maret 2024	Sluggish
2	V03	Vivarium	40	Darat & air	400000	Maret 2024	Smooth
3	P02	Paludarium	50	Rawa	400000	Januari 2024	Sluggish

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

DISCUSSIONS

K-Means is an unsupervised learning algorithm used to group data based on similar numerical attributes. In this case, the attributes used for clustering are product size and price. The selection of these two attributes is quite appropriate because they both have a significant influence on product purchasing decisions. In the first stage, the number of clusters is selected (K=2). This is an important step, because choosing the right K can affect the accuracy and interpretability of the clustering results. In this case, K=2 is chosen to group products into two categories: "Current" and "Congestion". After that, random centroid initialization is carried out. Centroid is the center point of each cluster. Each data is then calculated for its distance to the centroid, and the product will be sent based on its proximity to the centroid.Iteration: In the first iteration, the aquarium product (with size 60 and price 450,000) sends to the cluster with centroid 1, which has a larger size and price. Meanwhile, vivarium and paludarium are included in the cluster with centroid 2, which has a smaller size and price. After that, the centroid position is updated based on the average size and price in each cluster.

Advantages: K-Means allows us to find groups of data that have similar characteristics based on numerical attributes, without requiring labels first. This is suitable when we only have data without clear labels and want to understand hidden patterns or trends.

Limitations: One of the limitations of K-Means is its dependence on selecting the number of clusters (K). If K is chosen too small or large, the clustering results may not be optimal. Additionally, K-Means only considers numeric attributes and cannot handle categorical data directly (such as Type or Month-Year attributes), although this can be overcome by using encoding for categorical attributes.

Interpretation of Results: The clustering results show that products with larger sizes and prices tend to be in one cluster, which may be associated with “Current” sales status. Meanwhile, products with smaller sizes and prices tend to be in other clusters, which may be associated with "Loss" sales status. However, K-Means clustering results do not provide explicit classification labels (such as "Smooth" or "Congestion"), because K-Means only groups data without considering categories.Advantages: C4.5 is very effective for data classification and produces easy-to-understand decision rules, such as "If Size > 50 and Price > 400,000, then the sale is 'Loss'". Another advantage is that C4.5 can handle both numeric and categorical data well.

Limitations: The decision trees generated by C4.5 can become too complex if the dataset is large and has many attributes. Very large trees can make the model difficult to understand, and this can also lead to overfitting if proper tree pruning is not done.Interpretation of Results: Based on the resulting decision tree results, C4.5 is able to provide clear rules for predicting product sales status based on existing attributes. For example, products with larger sizes and higher prices tend to have “Slow” sales, while products with smaller sizes and lower prices tend to have more “Slow” sales. This provides a clear understanding of what factors influence sales status.

CONCLUSION

K-Means is more suitable for initial exploration and data segmentation without specific classification goals, K-Means will only group based on similarity of Size and Price, without considering sales predictions. C4.5 is more suitable if the goal is to predict or classify sales as "Current" or "Loss" based on other attributes. C4.5 will produce rules that support decision making. So, for the purpose of predicting “Current” or “Loss” sales, C4.5 is a more suitable algorithm than K-Means because it produces predictive rules that can be used in decision making.

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

REFERENCES

Bachtiar, Lukman, and Mahradianur Mahradianur. 2023. "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai." *Jurnal Informatika* 10(1):28–36. doi: 10.31294/inf.v10i1.15115.

Dewi, Kiki Rosita, Kemal Farouq Mauladi, and Masruroh. 2020. "Analisa Algoritma C4.5 Untuk Prediksi Penjualan Obat Pertanian Di Toko Dewi Sri." *Seminar Nasional Inovasi Teknologi* 25:109–14.

Dewi, Ni Luh Putu Purnama, I. Nyoman Purnama, and Nengah Widya Utami. 2022. "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)." *Jurnal Ilmiah Teknologi Informasi Asia* 16(2):105. doi: 10.32815/jitika.v16i2.761.

Islahudin, Rizal Abi, Sidik Rahmatullah, Asep Afandi, and Sriyani Safitri. 2022. "Algoritma C4.5 Untuk Memprediksi Kelayakan Penerima Bantuan Pangan Non Tunai." *Jurnal Informatika* 22(2):147–59. doi: 10.30873/ji.v22i2.3367.

Isnanto, Septian, and Suryarini Widodo. 2021. "Penerapan Data Mining Pada Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Clustering." *Jurnal Teknik Informasi Dan Komputer (Tekinkom)* 4(2):158. doi: 10.37600/tekinkom.v4i2.367.

Nas, Chairun. 2021. "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5." *Jurnal Manajemen Informatika (JAMIKA)* 11(2):131–45. doi: 10.34010/jamika.v11i2.5506.

Nurahman, Nurahman, Agung Purwanto, and Sigit Mulyanto. 2022. "Klasterisasi Sekolah Menggunakan Algoritma K-Means Berdasarkan Fasilitas, Pendidik, Dan Tenaga Pendidik." *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer* 21(2):337–50. doi: 10.30812/matrik.v21i2.1411.

Permadi, Anjar, and Yudhistira Arie Wiyaja. 2023. "Pengelompokan Terbaik Menggunakan Algoritma K-Means Pada Dataset Bus Biskita Bogor." *INTERNAL (Information System Journal)* 6(1):88–100. doi: 10.32627/internal.v6i1.689.

Pratiwi, Hesti, and Ari Purno Wahyu Wibowo. 2022. "Implementasi Algoritma K-Means Untuk Mengklaster Kelompok Sektor Perkebunan Di Indonesia." *JOISIE Journal Of Information System And Informatics Engineering* 6(1):1–10.

Riadi, Restu, and Mesran. 2023. "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume." *Journal of Informatics, Electrical and Electronics Engineering* 2(4):138–45. doi: 10.47065/jieee.v2i4.1181.

Yunus, Muhamad, Hanandriya Ramadhan, Dimas Rizki Aji, and Agus Yulianto. 2021. "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indon[1] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, 'Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 1, Januari 2025
https://doi.org/10.47709/cnahpc.v7i1.5319

Submitted : Jan 14, 2025
Accepted : Jan 14, 2025
Published : Jan 17, 2025

Indonesia Pintar (KIP),' Paradig. - J. Komput. Dan Inform., Vol." *Paradigma - Jurnal Komputer Dan Informatika* 23(2).

Kaur, Gagandeep, dan Meenu Arora. 2021. "A Comparative Study on Decision Tree Algorithms for Classification." *International Journal of Advanced Computer Science and Applications* 12(6):184–90.

Patel, Jignesh, dan Preeti Patel. 2020. "Performance Analysis of C4.5 and ID3 Decision Tree Algorithm for Data Classification." *International Journal of Engineering Research and Technology* 9(6):87–92.

Singh, Amandeep, Naveen Kumar, dan Aruna. 2022. "A Hybrid C4.5 Decision Tree Model with Random Forest for Predictive Analysis." *Journal of King Saud University - Computer and Information Sciences* 34(3):142–50.

Zhang, Li, dan Xueping Wei. 2023. "Enhanced C4.5 Decision Tree Algorithm Based on Pruning and Feature Selection." *Journal of Big Data* 10(1):1–15

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).