Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison of the K-Means Algorithm and C4.5 Against Sales Data

Abdi Dharma[1)*], Eko Bambang Wijaya[2)], Daniel Heyneker[3)], Jeff Vanness[4)]
[1,2,3,4)] Universitas Prima Indonesia, Indonesia
abdidharma@unprimdn.ac.id,[1)*] ekowijaya0904@gmail.com,[2)] danielheynekertalk@gmail.com,[3)]
jeff.vanness999@gmail.com[4)]

Submitted : Feb 27, 2023 | **Accepted** : Mar 22, 2023 | **Published** : Apr 1, 2023

Abstract: In general, the process of collecting and grouping data requires a long process. And if it has to be grouped manually it takes a very long time. Therefore, data mining is a solution for clustering data - a lot of data to classify it. In this research conducted at CV.Togu - Togu On Medan Branch, data mining is applied using the K-Means process model and the C4.5 algorithm which provides a standard process for using data mining in various fields used in classification because the results of this method easy to understand and easy to interpret. . The K-means method is a non-herarical method which is an algorithmic technique for grouping items into k clusters by minimizing the distance of the SS (sum of square) to the cluster centroid. In the K-means method, the number of clusters can be determined by the researcher himself. And the testing methods used to measure cluster quality are the Silhouette Coefficient and the Elbow Method. Based on the research conducted, there are significant differences before and after using the two methods. The results of the K-Means algorithm will be compared with the results of the C4.5 algorithm in the form of rules (decision trees). This research produces data on goods that have the highest level of sales/behavior.
Keywords: Data Mining; K-Means; Elbow Method; Silhouette Coefficients; C4.5 algorithm.

INTRODUCTION

Research (Fani Mulyana Nasution, 2019) applies the K-Means algorithm to rank food security with the aim of increasing food crop production in urban and local communities in North Sumatra province, technically. This is used to identify areas with food security potential to support food demand using the method K-Means examined based on harvested area, yield and planted area, the use of the K-Means method aims to classify areas with high, medium and low yields of food crops. This study leads to the grouping of several food crops with a total of 3 clusters, where the first cluster is a group with high food security potential, the second cluster is a group with moderate food security potential and the third cluster is a group with moderate food security potential. low food security potential. security potential. potential for food security [12].This study (Gabriella Amelia Prasetyo, R. Gunawan Santosa 2019) was conducted to compare the prediction accuracy of the C4.5 and k-Means algorithms in predicting the first semester GPA of UKDW FTI students. The data used is the 2008-2016 UKDW FTI Student Dataset as training data and the 2017 Batch as test data. The attributes used will be distinguished according to the paths achieved and not achieved. Passing paths use category, state, location and ICE level attributes, while non-passing paths use type, state, location, ICE level, number, speech attributes, space, similar. Accuracy will be calculated by cross table. The C4.5 algorithm achieves the best result of 77.45 n the k-Means algorithm achieves the best result of 60.78%. Scenarios with successful routes get an average accuracy score of 55.27n. Conditions with incomplete paths have an average accuracy score of 38.95%. [13]

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

One that can be used is data mining, which is a tool or process that extracts important information from big data by turning it into interesting information patterns. Information can be found in databases using certain techniques. K-Means and C4.5 are currently the most widely used clustering and classification algorithms. The K-Means algorithm is used for iterative clustering by dividing the dataset into predetermined CV.Togu Togu Medan Branch is a company that moves and starts a businessin the field of E-Commerce / sales in the online field. The existence of this company is compatible with the situation in Indonesia that is currently being developed. In determining the sales data collection process there are constraints including time efficiency, the number of comparisons of variables tested, clarifying sales data requires valid calculations so that comparison methods are needed to obtain information that is truly guaranteed authenticityyes. Based on this, it is hoped that it will make it easier for CV.Togu Togu on the Medan Branch to record sales from existing piles of data.

LITERATURE REVIEW

Study (Nasution, 2019) apply the K-Means algorithm to classify levels of food security with the aim of increasing production of food crops in local communities and urban areas in North Sumatra Province, this technique is used to determine areas with food security potential to help food needs by utilizing the K-Means method reviewed based on harvested area, production, and planted area, the use of the K-Means method is aimed at classifying areas with high, medium, and low yields of food crops. This research resulted in a grouping of several types of food crops with a total of 3 clusters where the first cluster is a group with high food security potential, the second cluster is a group with medium food security potential, and the third cluster is a group with low food security potential. Research (Fadilah Salsabila, Sheila Maulida Intani 2021) applies the K-Mean method and the C4.5 algorithm to determine the spread of Covid-19 in Indonesia. This study aims to sort the distribution of Covid-19 infection by province in Indonesia. Clustering uses a combination of the K-Means clustering algorithm and the C4.5 clustering algorithm. The KMeans algorithm works to group data into regional clusters in Indonesia by province. The grouping of results uses the C4.5 algorithm to see the rules in the form of a decision tree [14].The study (Relita Girsang1, Erika Fahmi Ginting, Masyuni Hutasuhut 2022) applies the C4.5 algorithm to identify beneficiaries of local government assistance programs, because the system is still used manually, so errors often occur when entering data about potential beneficiaries. As a result, this results in inaccuracies in obtaining support. Therefore we need a system that can correct performance errors and minimize errors at the village head's office, because it uses data mining. The result of this research is the creation of an application that is able to predict acceptance program assistance precisely and accurately, thereby assisting the village head's office in overcoming the problem of acceptance program assistance from the local government.[15]. (Prasetyo et al., 2017) explains the prediction accuracy of the C4.5 algorithm And k-Means to predict the first semester GPA of FTI UKDW students. The data used is UKDW FTI student data from 2008 to 2016, as training data and 2017 as exam data. The attributes used includeyes iscategory, state, location, and ICE level, whereas paths that do not pass use category, state, location, ICE level, number, language, space, and analogy. Accuracy will be calculated by cross table. Algorithm C4.5 obtains resultsthat is77.45% and the k-Means algorithm achieves the best results of 60.78%. The average accuracy rate of scenes with achievement paths is 55.27%, and the average accuracy rate of scenes without achievement paths is 38.95%.

METHOD

In the Clustering assessment of sales data using the K-Mean method, the following are:

(1) Data Collection

The procedure for collecting data for this study was based on sales records of the Medan branch of CV.Togu Togu. The data used is from 2020-2021. Data taken as many as 1000 records. There is as well as information from several journal literature to books related to this research topic. All data collected is entered into previously processed data. Data is selected at the pre-processing stage according to specified criteria. Data quality is influenced by several factors, among othersyesaccuracy, completeness, consistency, reality and interpretation.

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(2) Using the K-Means Algorithm

The algorithm is a group analysis technique that produces (N) objects divided into (K) groups (clusters), where each object in a group is grouped by its average(Sutoyo, 2019). Idea The goal of this algorithm is to divide the data into several groups by placing groups that have similar characteristics into one group and groups that have different characteristics into another group. Steps of the K-Means algorithm process in the implementation of the K-Means algorithm(Siregar, 2018):

1. Determine the number of cluster criteria.
2. Divide the data randomly into several groups.
3. Calculating the cluster center (midpoint) using the method of each group.
4. How to find the shortest center of gravity using the formula:

$$D(X_2 - X_1) = \|X_2 - X_1\|^2 = \sqrt{\sum_{i=1}^{p} |x_2 - x_1|^2} \quad (1)$$

5. Align the data to the nearest centroid. The formula for aligning data from the nearest centroid :
6. Then repeat step 3, repeat the process until the resulting centroid value is fixed and no cluster members move to another cluster.

(3) Using the C4.5 Algorithm

Algorithm C4.5 is a method for building decision trees based on the training data provided. The C4.5 algorithm is a further development of ID3(Hidayanti et al., 2020). There are several steps to build a decision tree using the C4.5 algorithm, namely(Narulita et al., 2021):

1. Prepare training data. Training data usually comes from historical data from the past and is divided into certain categories.
2. Find the root of the tree, the root is taken from the selected attribute by calculating the gain value of each attribute, the highest gain value is the first root. Before calculating the attribute gain value, first calculate the entropy value. The formula is used to calculate the entropy value

$$Entropy(S) - \text{pi}. \log 2\ \text{pi} \sum_{i=1}^{n} - \quad (3)$$

Calculating the gain value using the formula:

$$gains(A) = \text{Entropy}(S) - x\ \text{Entropy}(Si) \sum_{i=1}^{n} |Si|/|S| \quad (4)$$

Repeat step 2 until all records are partitioned, the decision tree partitioning process will stop when:

a) All records in node N get the same class.
b) There are no attributes in the partitioned record anymore.
c) There are no records in the branch that are empty.

RESULT

There are several procedures in the research stages used in this study, the first is the procedure for identifying the problem, then collecting data, implementing the algorithm, and finally testing the two methods used in the procedure. basic classification This algorithm is implemented using two algorithms, K-Means and C4.5 algorithms.

Calculations Using the K-Means Algorithm

Calculation of K-Means

In the Clustering assessment of sales data in the *K-Mean method* what is done is to make a table of criteria, so that it can be seen more clearly in table 1 as follows:

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1 Sample Stock Data for July 2019 – July 2019

No	Name of goods	First stock	Stock Sold	Last stock
1	Fancytime Umbrella	6	3	3
2	One Two Cups Tea Bags	10	10	0
3	TaffHOME Towel Cloth	3	3	0
4	TaffLED Headlight LED Flashlight	104	76	28
5	Taffware HUMI Essential Oils	4	1	3

Table 2 Initial Centroid

Information	First stock	Stock Sold	Last stock
C1	104	76	28
C2	100	137	13
C3	3	3	0

In this process to determine the initial centroid C1 is taken from the 4th and C2 data taken from the 3rd data based on the highest number and lowest number. Calculating the distance of each existing data to the centroid value, shown for the following first iteration data:

$$D(X1, C1) = \sqrt{(\mathrm{SA}_1 - \mathrm{C1}_1)^2 + (\mathrm{ST}_1 - \mathrm{C1}_1)^2 + (\mathrm{SAK}_1 - \mathrm{C1}_1)^2}$$
$$= \sqrt{(6-104)^2 + (3-76)^2 + (3-28)^2}$$
$$= 124.7317$$

And so on until data 10 to obtain:

Table 3 Data Cluster 1 Iteration 1

No	Name of goods	First stock	Stock Sold	Last stock	C1	C2	C3	Minimum Value
1	Fancytime Umbrella	6	3	3	124.7317	4.242640687	163.9878	4.242640687
2	One Two Cups Tea Bags	10	10	0	118.2201	9.899494937	156,1986	9.899494937
3	TaffHOME Towel Cloth	3	3	0	127.7263	0	165.9337	0

The next step is to determine the location of the cluster by comparing the two clusters, the minimum value is the selected value. If a value is found the smallest (minimum) then it can be included in the cluster. For more details see table 4 below:

Table 4 Determination of New Clusters

No	Name of goods	First stock	Stock Sold	Last stock	C1	C2	C3
1	Fancytime Umbrella	6	3	3		1	
2	One Two Cups Tea Bags	10	10	0		1	
3	TaffHOME Towel Cloth	3	3	0		1	
4	TaffLED Headlight LED Flashlight	104	76	28	1	1	
5	Taffware HUMI Essential Oils	4	1	3		1	

Then determine the new centroid value, this value is determined by the data that enters the cluster, based on the table above (data 1-10) the following values are obtained:

1. Cluster 1 has 2 data
2. Cluster 2 has 7 data

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

3. Cluster 3 has 1 data

To determine a new centroid value (for example, in Cluster 1 there is 1 data, this can be done in the following way:

$$\boldsymbol{CK} = \frac{\text{jumlah dari nilai yang masuk kedalam cluster}}{\text{Jumlah data yang masuk}}$$

For the first centroid (C1):

$$\boldsymbol{CK_1} = \frac{104 + 100}{1} = 102$$

For the second centroid (C2) there are 7 data:

$$\boldsymbol{CK_1} = \frac{6 + 10 + 3 + 4 + 4 + 12 + 4}{7} = 6{,}142$$

For the second centroid (C1) there is 1 data:

$$\boldsymbol{CK_1} = \frac{100}{1} = \mathbf{100}$$

Overall, the centroid value is obtained:

Table 5 New Clusters

Information			
C1	102	106,2	20.5
C2	5,875	3,875	2
C3	100	137	0

To find the next centroid value, repeat the steps above. Based on the cluster results above, it is done by calculating the average of each cluster member based on predetermined categories, the following analysis results can be described in graphical form as follows:

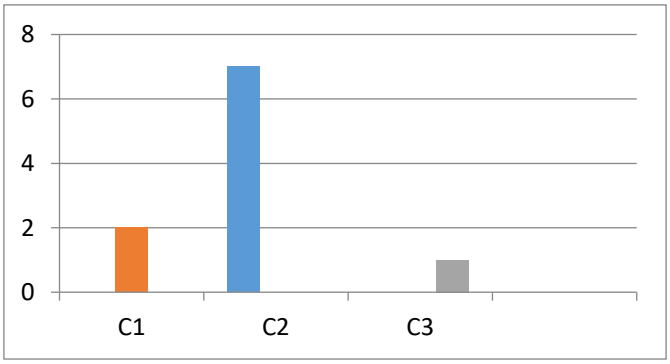

Figure 1 Cluster Distribution Graph

How the Elbow Method Works

Elbow method to determine the optimal or best number of clusters. Elbow method steps:

Literacy 1 :

SSE(1)

$= (6 - 4{,}242640687)^2 + (3 - 4{,}242640687)^2 + (3 - 4{,}242640687)^2$

= 3.0976 + 1.5376 + 1.5376

= 6.1728

And so on up to the 10th data .

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 6 SSE Calculation Results

literacy	SSE	Distance	Information
Literacy 1	27076.87	-	-
Literacy 2	59288.25	32211 ,38	C1 to C2
Literacy 3	45922,15	13366 ,1	C2 to C3

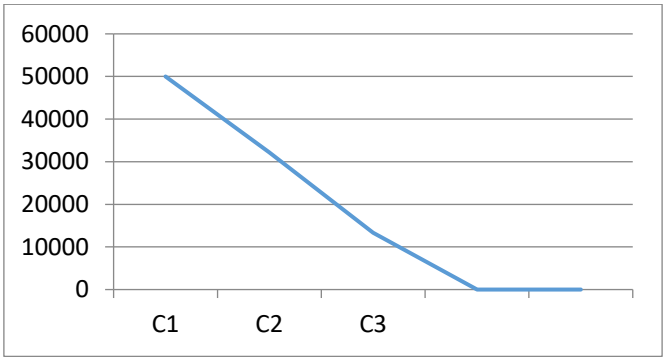

Figure 2 Results of the Elbow method

The cluster value taken as the optimal cluster in the Elbow method is the point that forms the elbow. C SSE Distance . The value of the distance from 1 cluster to these 2 clusters is the distance value that has experienced the most significant or greatest decrease, followed by a relatively constant distance value, so that 2 clusters are the optimal or best clusters. the points that form the elbow are at the 2 cluster points, as shown in Figure 3.4

How the Silhoutte Coefficient Method works

The following is an example of data that has been formed and grouped in each cluster.

Table 7 Silhoutte Coefficient Results

No	Name of goods	C1	C2	C3	(ai)	b(i) =Min(d)	S(i)
1	Fancytime Umbrella	142.03	1.33	7.35	141.03	7.35	140.03
2	One Two Cups Tea Bags	134.68	7.65	14,14	133.68	14,14	132.68
3	TaffHOME Towel Cloth	144.45	3.61	4,243	143.45	4,243	142.45

The results of the silhouette coefficient results for the accuracy of the silhouette coefficient can be seen in table 3.9. due to the results of the silhoutte value in the cluster, the data grouping in the cluster is categorized as Good where the value in table 3.8 is greater than -1.

Table 8 Silhoutte Coefficient Results

Silhoutte Accuracy Results
912.0369

C.45 algorithm and the objective attribute is opportunity with the classification of the opportunity attribute being Applicable (L) and Not Applicable (TL).

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 9 Data for Classification of Goods

No	Name of goods	Types of goods	Sold	Price	Category
1	Fancytime Umbrella	Accessories & More	761	Normal	L
2	One Two Cups Tea Bags	Food Equipment	29	Cheap	TL
3	TaffHOME Towel Cloth	Hygiene kits	25	Cheap	L

The following is a more detailed explanation of each step in constructing a decision tree using the C.45 algorithm .

Table 10 Establishment Decision Tree

No	Quantity	Price	Category
1	Lots	Tall	L
2	Lots	Low	TL
3	A little	Tall	TL

The decision tree is made after calculating the total *entropy* , *the entropy* of each attribute and calculating the gain and determining the highest gain. *Entrophy* quest key :

1. If one of the "Applicable" or "Not Applicable" columns has a value of 0, then *the entropy* is also confirmed to be 0.
2. If the "Applicable" and "Not Applicable" columns have the same value, then *the entropy* is also confirmed to be 1.

Table 11 *Node* 1 Criteria Classification Table

nodes	Number of Cases		In demand	Not sold
1	Total	10	6	4
	Types of goods :			
	Accessories & more	5	3	2
	Food Equipment	1	0	1
	Hygiene kits	4	3	1
	Sold :			
	<100	5	3	2
	>100	5	4	1
	Price :			
	Tall	8	6	2
	Low	2	1	1
	quantity :			
	Lots	6	3	3
	A little	4	1	3

Total Entropy = Entropy (S) = $\sum - n\ i = l$ pi*log $_2$ pi Total Entropy
= ((-6/10 * log 2 (-6/10) + (-4/10 * log 2 (-4/10)) = **0.15653559774**
Service Gain 1= 0.15653559774 - (((5/10*0 + ((1/ 10*0)) + ((4/10 x 0.15653559774)) = **0.09392135864**

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 12. *Node* 1 Calculation Results

nodes	Number of Cases		Entropy	gains
1	Total	10	0.15653559774	
	Types of goods :			0.09392135864
	Accessories & more	5	0.15653559774	
	Food Equipment	1	0.30102999566	
	Hygiene kits	4	0.18814374729	
	Sold :			0.15653559774
	<100	5	0.15653559774	
	>100	5	0.15653559774	
	Price :			0.625613
	Tall	8	0.970951	
	Low	2	0.650022	
	quantity :			0.695577
	Lots	6	0.468996	
	A little	4	0	

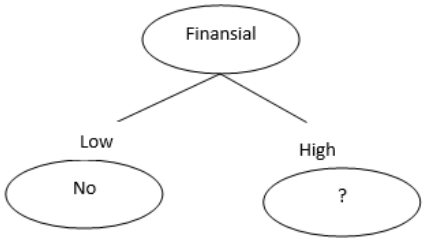

Figure 3 Decision Tree of *Node* 1 Calculation Results

the Entropy and *Gain* values are calculated for each attribute to become the root *node* of the Height attribute, in the same way as above by calculating the *entropy values* of the remaining attributes, namely player quality, age, and coach quality.

Table 13. *Node* Criteria Classification Table 1.1

nodes	Number of Cases		In demand	Not sold
1.1	Total	10	9	1
	Types of goods :			
	Accessories & more	5	3	2
	Food Equipment	1	0	1

*Eko Bambang Wijayes

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12224
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	Hygiene kits	4	3	1
	Sold :			
	<100	5	3	2
	>100	5	4	1
	Price :			
	Tall	8	6	2
	Low	2	1	1

Table 14. *Node* Calculation Results 1.1

nodes	**Number of Cases**		**Entropy**	**gains**
1.1	Total	10	0.468996	
	Types of goods:			0
	Accessories & more	10	0.468996	
	Food Equipment	0	0	
	Hygiene kits			
	Price :			0.468996
	Tall	2	0	
	Low	7	0	
	Sold			0.108032
	>100	5	0.721928	
	<100	5	0	

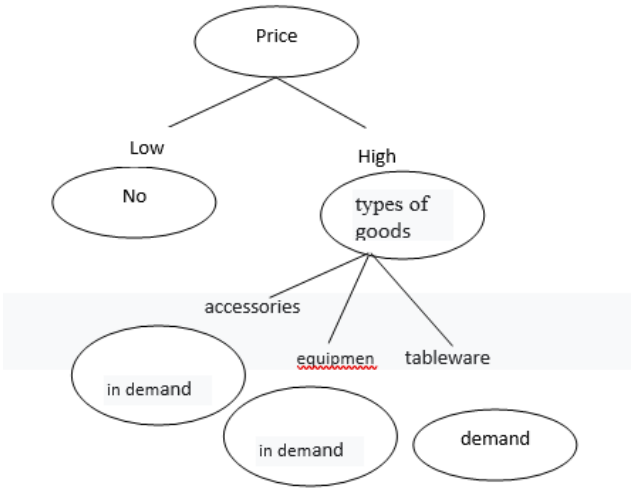

Figure 4. Node Decision Tree 1.1

*Eko Bambang Wijayes

From the results in Table 3.18. it can be seen that the attribute with the highest *Gain* is types of goods , which is equal to 0.468996. Thus Item Type can be a branch *node* of the High attribute value. There are three attribute values of Type of Goods, namely accessories, Cleaning Equipment, and Tableware. Of these three attributes, the Accessories attribute value has classified the case into 1, namely the decision is valid, the Average attribute value has also classified the case as 1, namely the decision is valid and the Cutlery attribute value has also classified the case into 1, namely the decision is Not Sellable, so no further calculations are necessary. Thus a decision tree is obtained that looks like in Figure 3.17.

DISCUSSION

The results showed that with Based on a comparison of the calculations of the 2 methods, it can be concluded that the two methods of calculating K-Means and the C.45 Algorithm get the same conclusion where there are two attributes/categories of goods with the highest sales, namely the K-Means method produces 2 clusters, namely 1 and cluster 2 and refined with the C.45 Algorithm where the Type of Goods Accessories and Cleaning Equipment has a high level of sales. The accuracy of the K-Means Method is 912.0369%.

CONCLUSION

Based on the comparison of the calculation of the 2 methods, it can be concluded that the two K-Means calculation methods and the C.45 Algorithm get the same conclusion where there are two attributes/categories of goods with the highest sales, namely the K-Means method produces 2 clusters namely 1 and cluster 2 and is refined with the C.45 Algorithm where the Types of Accessories and Cleaning Tools have a high level of sales. The resulting accuracy of the K-MEANS model is 912.0369%. The accuracy produced by the C4.5 algorithm model is 0.468996%.

REFERENCES

Narulita, S., Oktaga, A. T., & Susanti, I. (2021). Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4 . 5. *Jurnal Media Aplikom*, *13*(2), 15–29.

Nasution, F. M. (2019). Penerapan Metode K-Means Clustering Untuk Mengelompokkan Ketahanan Tanaman Pangan Kabupaten/Kota Diprovinsi Sumatera Utara. In *Skripsi*.

Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48. https://doi.org/10.51557/pt_jiit.v6i2.659

Prasetyo, G. A., Santosa, R. G., & Chrismanto, A. R. (2017). *Memprediksi Kategori Indeks Prestasi Mahasiswa*. *5*. https://doi.org/10.21460/jutei.2019.32.185

Pungky, A. (2019). Penerapan Metode K-Nn Untuk Memprediksi Hasil Pertanian Di Kabupaten Malang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *3*(1), 235–242.

Rozana, L., & Musfikar, R. (2020). Analisis Dan Perancangan Sistem Informasi Pengarsipan Surat Berbasis Web Pada Kantor Lurah Desa Dayah Tuha. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, *4*(1), 14. https://doi.org/10.22373/cj.v4i1.6933

Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, *1*(2), 83–91. https://doi.org/10.36378/jtos.v1i2.24

Sutoyo, M. N. (2019). *Algoritma K-Means*. *1*, 1–7.

Wahyudi, I., Sulthan, M. B., & Suhartini, L. (2021). Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan. *Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM)*, *2*(2), 72–81. https://doi.org/10.31102/jatim.v2i2.1274

Widiyati, D. K., Wati, M., & Pakpahan, H. S. (2018). Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *2*(2), 125. https://doi.org/10.30872/jurti.v2i2.1864

Syafii, A., Dwilestari, G., & Ajiz, A. (2022). Komparasi Algoritma Naïve Bayes Dan Algoritma C4.5

*Eko Bambang Wijayes

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Dalam Klasifikasi Pelanggan Produk Indihome. *Jurnal Sistem Informasi Dan Manajemen(JURISMA)*, *10*(2), 60–70. https://ejournal.stmikgici.ac.id/

Lee, J. S. (2019). AUC4.5: AUC-Based C4.5 Decision Tree Algorithm for Imbalanced Data Classification. *IEEE Access*, *7*, 106034–106042. https://doi.org/10.1109/ACCESS.2019.2931865

Shi, X., Li, G., Li, K., Liu, J., & Wang, R. (2020). Customer Classification Method of Logistics Enterprises Based on BP-AdaBoost. *Journal of Physics: Conference Series*, *1670*(1). https://doi.org/10.1088/1742-6596/1670/1/012018

Kartikawati, L. (2022). Analisis Kualitas Pengelompokkan Algoritma K-Means di Knime dan Excel untuk PTMT Pasca Vaksinasi Covid-19. *Ideguru: Jurnal Karya Ilmiah Guru*, *7*(1), 70–79. https://doi.org/10.51169/ideguru.v7i1.316

Arasa, R. A., & Setiawanb, N. A. (2022). Comparison Of Data Mining Classification Techniques For Heart Disease Prediction System. *Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, *2*(2), 85–90.

*Eko Bambang Wijayes