

## APPLICATION OF THE K-MEANS AND DECISION TREE ALGORITHMS IN DETERMINING STUDENT ACHIEVEMENT

Nandya Rifki Jevintya<sup>1</sup>, Ucuk Darusalam<sup>2</sup>, Syahid Abdullah<sup>3</sup>

<sup>1,2,3</sup>Department of Informatics, Universitas Siber Asia

\*Email: [digojevintya9@gmail.com](mailto:digojevintya9@gmail.com), [ucukdarusalam@exe.unsia.ac.id](mailto:ucukdarusalam@exe.unsia.ac.id), [syahidabdullah@lecturer.unsia.ac.id](mailto:syahidabdullah@lecturer.unsia.ac.id)

(Received: 23 February 2024, Revised: 01 March 2024, Accepted: 28 March 2024)

### Abstract

Various factors influence student achievement, both internal and external; this makes it difficult for some teachers to detect every student in class. This research aims to determine student achievement in class among students at the SDS Kartika X-6 school. Data comes from SDS Kartika X-6, an elementary school owned by the Indonesian Army. By knowing the factors that influence the determinants of student learning achievement, steps can be taken to improve student learning achievement at SDS Kartika x-6. The methods used in this research are the K-Means algorithm and Decision Tree. This method will be chosen to determine student learning achievement. The process begins by determining clusters using the K-Means algorithm; then a classification process is carried out using a Decision Tree. The number of datasets in this research is 28, and the criteria are gender, mathematics grades, English, natural sciences, religion, class performance, and school achievement. The implementation results show that academic grades, class achievements, and school achievements play a role in determining student achievement for SDS Kartika X-6 students. Meanwhile, 3 clusters were formed: Fairly Good, Good, and Very Good. In the testing stage using the Decision Tree method, prediction accuracy was 71%, with an error of 29.

**Keywords:** *Clustering, Decision Tree, Google Colab, K-Means, Academic Value*

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



\*Corresponding Author: Ucuk Darusalam

## 1. INTRODUCTION

To keep up with the times, one must first strengthen oneself; education is a beautiful approach to achieving this goal. Education is necessary to produce quality human resources because its people's resources determine a nation's achievement [1]. In previous studies, in applications using k-means algorithm methods, it can process data entered in the form of values of nahwu, shorof, moral, memorization, presence, and honesty, which is a process in which students' values will be selected in the process of grouping students with achievements so that they are quick and accurate [2]. In a recent study, classification was carried out using a decision tree algorithm, so it was found that parents' income did not affect student achievement. However, the distance from home to school influenced student achievement[1].

Then the study entitled Implementation of K-Means and K-Nearest Neighbors in the Outstanding Student Category, uses the K-means and K-Nearest Neighbors algorithm methods to categorize

outstanding students. The result of the research is that it is hoped that more special attention will be paid to creating the best graduates at SMKN 3 Karawang [3]. This research decided to collect data at SDS Kartika X-6, which is a private elementary school belonging to the Army. SDS Kartika X-6 in determining student achievement with class achievement scores, and school achievement. With what is currently being done, there are many shortcomings and problems faced. Especially considering that there is a lot of data that will influence the process of determining student academic grades with class achievement, and school achievement in determining student achievement.

The application in making achievement is in terms of method and based on class achievement and school achievement, considering that previous research used parental income and the distance from home to school [1]. And where the majority only use the K-Means algorithm, by using two methods and academic grades, class achievement, and school achievement, you will find out how influential they are in determining student achievement. In this research, the K-Means

and Decision Tree algorithms will be used to solve this problem. Researchers can then help take steps toward student achievement in SDS Kartika X-6 and find out how much is involved in determining students' achievements with academic scores, class achievement, and school achievement.

**2. RESEARCH METHOD**

In this investigation, the Decision Tree method and the K-Means algorithm were used. Both the Decision Tree method and the K-Means algorithm can be used in the following way, which consists of the stages involved in Figure 1.

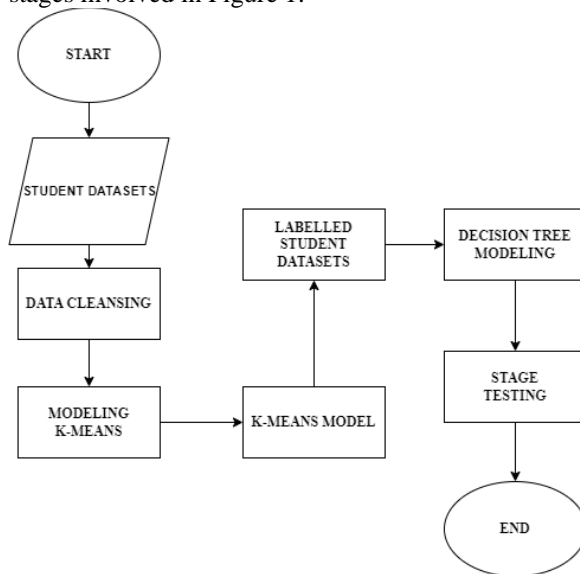


Figure 1. Research Stages implementation K-Means and Decision Tree

To proceed to the next step. This is where the work Clearing Data does with student data sets begins. After interviewing the principal of SDS Kartika X-6, an Excel dataset was obtained. After the data clearing method is complete, the data will be modeled using the K-Means approach with parameter k = 3. This allows the cluster findings to be used as labels for further modeling. A labeling process will be carried out on the dataset before the Decision Tree method is used for the modeling process. Using measurements such as, classification reports, matrix of confusion, and accuracy, the final step is to analyze the Decision Tree algorithm to determine which students achieved academic scores, class achievement, and school achievement.

**2.1 K-Means Algorithm**

It is said that clustering is a process that tries to organize data according to the similarities between the data in the cluster. This process aims to make data within a cluster very similar to each other, while data between clusters is very different from each other [4], [5], [6]. A group of objects or components that are connected is called a cluster. Therefore, clustering analysis will result in the formation of many groups [2]. Here are some well-known and widely used types

of clustering algorithms for processing all types of data, such as density-based, centroid, hierarchical, and K-means [7].

The K-Means algorithm is one of the algorithms that is widely used in clustering because of its simplicity and efficiency and is recognized as one of the top 10 data mining algorithms by the IEEE. Among the methods often used in clustering, the K-Means algorithm is one of the algorithms that is often used due to its ease of use, speed, and recognition as a data mining tool [8], [9]. With the clustering strategy, researchers often use the K-Means method in combination with it [3], [10]. This is because the clustering approach is very effective in finding groups of data. The K-Means Clustering approach divides data into many groups, assigning data that is similar to one group and data that is different to another group [11], [12], [13].

In the K-Means Clustering algorithm, data is divided into several groups, where data in one group has the same characteristics and data in other groups has different characteristics. The K-Means Clustering method begins with the following stages [11]:

1. Determine k, which is the total cluster you want to create.
2. K centroid (cluster center point) prefix that will be used randomly.

$$v = \sum_k^n = 1 \frac{xi}{N} \tag{1}$$

i = 1,2,3 ... n

The formula determines the initial centroid. v is the centroid and i is the iteration.

3. The distance of each object to each centroid is calculated based on each cluster. The distance between the object and the centroid is calculated using Euclidian Distance.

$$d(x,y) = |x-y| = \sqrt{\sum_{i=1}^n (xi - yi)^2} \tag{2}$$

The formula takes into account the distance between centroids.

4. Allocate each object to the nearest centroid.
5. Carry out the iteration. Next, using the equation, determine the position of the new centroid. If the centroid locations are not similar, it means repeating stage 3 [1].

Where :

x: First data point

y: Second data point

n: Number (attributes) in data mining

d(x,y): Euclidean distance, namely the distance between data at points x and y using mathematical calculations [14].

**2.2 Decision Tree**

This modeling can be used to determine the type of information that can be extracted from the data that

has been generated. In making decisions with the help of the Decision Tree model, it is at this stage that the Decision Tree algorithm can be used [1]. To make predictions, a Decision Tree that has been trained using the C4.5 approach can use a collection of decision rules [15]. In a Decision Tree, internal and root nodes are labeled with attribute names, edges are labeled with attribute values and leaf nodes are labeled with multiple classes. This makes the Decision Tree a fundamental example of a classification technique that can be used for many limited classes [16], [17].

This modeling stage is useful for finding out what patterns or information can be used as useful information from the prepared data. At this modeling stage, the Decision Tree algorithm is used, and this algorithm will be useful for making decisions whose modeling is in the form of a decision tree. In formulating the Decision Tree algorithm, the formula used is as follows [1]:

$$Entropy(S) = \sum_{i=1}^n - P_i \log 2 p_i \tag{3}$$

The formula for calculating entropy

Explanation :

S = Case Collection

S = Number of Partitions S

S = Proportion of

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} entropy(S_i) \tag{4}$$

The formula for calculating Gain

Explanation :

S = Case Collection

S = Attribute

S = Number of Attribute Partitions A

Si = Number of Cases in Partition i

S = Number of Cases in S [1]

### 3. RESULTS AND DISCUSSION

#### 3.1 Data Collection

This research data has been collected at the SDS Kartika X-6 school. Various characteristics included in the data include number, gender, Mathematics subjects (MTK), Indonesian Language (BI), Natural Sciences (IPA), Religion, class achievement, and school achievement (with a total of data obtained from 28 students ). The data collection process consisted of interviewing the Principal of SDS Kartika X-6 with students.

No.	Jenis Kelamin	MTK	BI	IPA	Agama	Prestasi Kelas	Prestasi Sekolah	
0	1	2	88	86	85	82	0	0
1	2	1	75	79	82	78	0	0
2	3	2	92	87	88	85	1	0
3	4	2	87	84	86	84	0	0
4	5	2	83	84	86	82	0	0
5	6	1	88	82	88	81	0	0
6	7	1	87	81	81	78	0	0
7	8	2	87	84	80	80	0	0
8	9	1	80	83	88	80	0	0
9	10	1	82	84	85	80	0	0
10	11	2	84	82	88	82	0	0
11	12	1	82	75	82	76	0	0
12	13	1	88	82	84	79	0	0
13	14	2	92	86	92	87	0	1
14	15	2	82	82	88	84	0	0
15	16	2	84	75	82	75	0	0
16	17	2	74	75	78	75	0	0
17	18	2	87	84	90	80	0	0
18	19	2	78	89	90	81	0	0
19	20	2	82	79	81	78	0	0
20	21	1	88	80	85	81	0	0
21	22	1	88	92	80	90	1	0
22	23	1	88	77	78	87	0	0
23	24	1	89	89	91	80	0	1
24	25	2	87	85	75	80	0	0
25	26	2	81	80	83	82	0	0
26	27	1	76	84	77	84	0	0
27	28	2	80	89	87	81	0	0

Figure 2. Research Dataset

Explanation of Figure 2. above:

Gender

1 Boy

2 Women

Class Achievement

0 None

1 Yes

School Achievement

0 None

1 Yes

#### 3.2 Data Cleansing

```

df_null = round(100*(df.isnull().sum())/len(df),2)
df_null

No.          0.0
Jenis Kelamin  0.0
MTK          0.0
BI           0.0
IPA          0.0
Agama       0.0
Prestasi Kelas  0.0
Prestasi Sekolah  0.0
dtype: float64
    
```

Figure 3. Data Cleansing

#### 3.3 K-Means uses Google Colab

As soon as the data is ready, use Figure 3. The first step of this research was to develop a model based on the student data set. Resources provided by Google Colab were used in this investigation. This is the first step to finding labels in existing data, namely using K-Means modeling (Figure 4).

```
from sklearn import cluster
df = pd.read_csv("/content/data/akhir.csv")
df
```

No.	Jenis Kelamin	MTK	BI	IPA	Agama	Prestasi Kelas	Prestasi Sekolah
0	1	2	88	86	85	82	0
1	2	1	75	79	82	78	0
2	3	2	92	87	88	85	1
3	4	2	87	84	86	84	0
4	5	2	83	84	86	82	0
5	6	1	88	82	88	81	0
6	7	1	87	81	81	78	0
7	8	2	87	84	80	80	0
8	9	1	80	83	88	80	0
9	10	1	82	84	85	80	0
10	11	2	84	82	88	82	0
11	12	1	82	75	82	76	0
12	13	1	88	82	84	79	0
13	14	2	92	86	92	87	0
14	15	2	82	82	88	84	0
15	16	2	84	75	82	75	0
16	17	2	74	75	78	75	0
17	18	2	87	84	90	80	0
18	19	2	78	89	90	81	0
19	20	2	82	79	81	78	0
20	21	1	88	80	85	81	0
21	22	1	88	92	80	90	1
22	23	1	88	77	78	87	0
23	24	1	89	89	91	80	0
24	25	2	87	85	75	80	0
25	26	2	81	80	83	82	0
26	27	1	76	84	77	84	0
27	28	2	80	89	87	81	0

Figure 4. Dataset for K-Means Modeling

By using MTK, BI, IPA, Religion, and school achievement data as a basis, clustering was carried out. In dividing the k-values into three different groups, we will later use these groupings as labels in the data set relating to students. The clustering process is used to get the average value of the current dataset. This is the reason why the value of k is divided by three. This results in a model that appears as follows in Figure 5 and 6:

```
from sklearn.cluster import KMeans
kmean = KMeans(n_clusters=3)
kmean

[473] y_cluster = kmean.fit_predict(x_train)

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning
warnings.warn(

[474] y_cluster
array([1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 2, 1, 0, 0, 1, 1, 0, 1, 1,
       0, 2, 0, 0, 0, 1], dtype=int32)
```

Figure 5. K-Means Algorithm Model

```
kmean.cluster_centers_
array([[4.49494949e-01, 2.62032086e-01, 2.88770053e-01, 2.90909091e-01,
        1.38777878e-17],
       [6.11111111e-01, 5.68627451e-01, 6.78431373e-01, 4.75555556e-01,
        1.38777878e-17],
       [9.16666667e-01, 7.35294118e-01, 9.70588235e-01, 5.66666667e-01,
        1.00000000e+00]])
```

Figure 6. Centroid K-Means

### 3. 4 Decision Tree using Google Colab

Next, the student dataset will be labeled according to the results of the clustering model trained using K-Means before being applied to the data. Next, a classification method is used, namely the Decision Tree algorithm, to model the student dataset, which has been labeled in Figure 7 [1].

```
df["Cluster"] = df["Cluster"].replace([0,1,2],["Cukup Baik","Baik", 'Sangat Baik'])
```

No.	Jenis Kelamin	MTK	BI	IPA	Agama	Prestasi Kelas	Prestasi Sekolah	Cluster
0	1	2	88	86	85	82	0	Baik
1	2	1	75	79	82	78	0	Cukup Baik
2	3	2	92	87	88	85	1	Baik
3	4	2	87	84	86	84	0	Baik
4	5	2	83	84	86	82	0	Baik
5	6	1	88	82	88	81	0	Baik
6	7	1	87	81	81	78	0	Cukup Baik
7	8	2	87	84	80	80	0	Cukup Baik
8	9	1	80	83	88	80	0	Baik
9	10	1	82	84	85	80	0	Baik
10	11	2	84	82	88	82	0	Baik
11	12	1	82	75	82	76	0	Cukup Baik
12	13	1	88	82	84	79	0	Baik
13	14	2	92	86	92	87	0	Sangat Baik
14	15	2	82	82	88	84	0	Baik
15	16	2	84	75	82	75	0	Cukup Baik
16	17	2	74	75	78	75	0	Cukup Baik
17	18	2	87	84	90	80	0	Baik
18	19	2	78	89	90	81	0	Baik
19	20	2	82	79	81	78	0	Cukup Baik
20	21	1	88	80	85	81	0	Baik
21	22	1	88	92	80	90	1	Baik
22	23	1	88	77	78	87	0	Cukup Baik
23	24	1	89	89	91	80	0	Sangat Baik
24	25	2	87	85	75	80	0	Cukup Baik
25	26	2	81	80	83	82	0	Cukup Baik
26	27	1	76	84	77	84	0	Cukup Baik
27	28	2	80	89	87	81	0	Baik

Figure 7. labeled Datasets

The attributes used are MTK, BI, IPA, Religion, school achievement. With clusters that have been labeled from K-Means Figure 8.

```
x = df.iloc[:,2:7]
y = df.iloc[:,8]
```

	MTK	BI	IPA	Agama	Prestasi Kelas
0	88	86	85	82	0
1	75	79	82	78	0
2	92	87	88	85	1
3	87	84	86	84	0
4	83	84	86	82	0
5	88	82	88	81	0
6	87	81	81	78	0
7	87	84	80	80	0
8	80	83	88	80	0
9	82	84	85	80	0
10	84	82	88	82	0
11	82	75	82	76	0
12	88	82	84	79	0
13	92	86	92	87	0
14	82	82	88	84	0
15	84	75	82	75	0
16	74	75	78	75	0
17	87	84	90	80	0
18	78	89	90	81	0
19	82	79	81	78	0
20	88	80	85	81	0
21	88	92	80	90	1
22	88	77	78	87	0
23	89	89	91	80	0
24	87	85	75	80	0
25	81	80	83	82	0
26	76	84	77	84	0
27	80	89	87	81	0

Figure 8. Decision Tree Attributes

Modeling is done by dividing the labeled student dataset into training data and random states. The data used is MTK, BI, IPA, Religion, Class achievement and cluster, so the model results are in Figure 9:

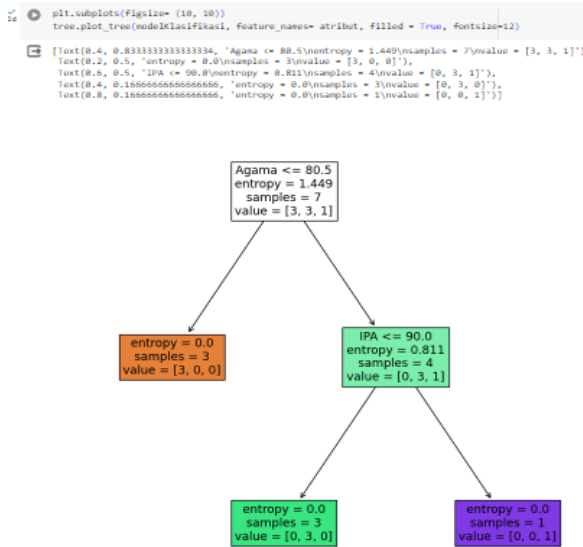


Figure 9. Decision Tree Algorithm Modeling

To make it easier to obtain information from the Decision Tree modeling that has been carried out, we will detail or explain the model that was successfully executed. This is the explanation in Figure 10:

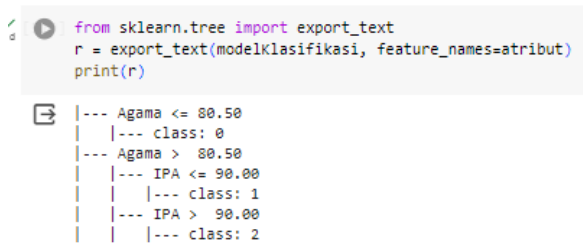


Figure 10. Explanation of the Decision Tree Model

### 3.5 Decision Tree Testing

In the context of Decision Tree modeling, the first step taken is making predictions, which will then be assessed for accuracy, confusion matrix, and classification report. All processes are using Google Colab. The following is a list of test results in Figure 11 :

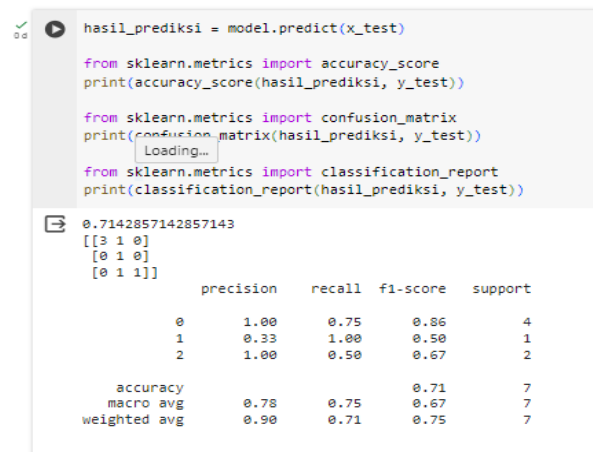


Figure 11. accuracy, confusion matrix, classification report

In the analysis of student achievement, it can be seen from the K-Means and Decision Tree algorithms that have been created. Modeling the dataset in K-Means which has been labeled as follows:

- A. Fairly Good, there are 11 students in the Fairly good category.
- B. Good, there are 15 students in the good category.
- C. Very Good, there are 2 students in the Very good category.

The results of determining the cluster are Fairly Good, Good, and Very Good by adding up the distance calculation values for each cluster member in the K-Means algorithm and then sorting them from the values of Fairly Good, Good, and Very Good.

Table 1. K-Means Algorithm Results

Algorithm	Cluster	Amount	Category
K-Means	0	11	Fairly Good
	1	15	Good
	2	2	Very Good

Decision Tree modeling is based on subject score IPA, subject score Religion, and cluster. The decision tree approach is capable of achieving a prediction accuracy score of 71%. In this research in determining students perform with the subject score IPA, subject score Religion, and cluster using the Decision Tree algorithm but with different attributes. Decision Tree modeling based on subject score IPA, subject score Religion and cluster. Decision tree approaches are able to a prediction accuracy score of approximately 71%. That in this research in determining student achievement with subject score IPA, subject score Religion, and cluster using the Decision Tree algorithm but with different attributes.

In previous research, the Decision Tree algorithm was not suitable to use. The results of the research are different in several ways from the results of this research, in that the Decision Tree has an influence in determining student achievement, namely in terms of methods and based on the many attributes used. Previous research used parental income and distance from home to school [1], and the majority only used the K-Means algorithm.

### 4. CONCLUSION

The research that has been carried out in applying the clustering method using K-Means and the Decision Tree algorithm classification method using a student dataset containing number, gender, MTK, BI, IPA, Religion, class achievement, and school achievement. In the research user, judging from the science and religion scores, the application in determining high achieving students is suitable. Fairly Good, Good, and Very Good are three separate categories that are thought to be related to the K-Means method. Next, the

accuracy of the Decision Tree's prediction algorithm is 71%. To determine student achievement with MTK, BI, IPA, Religion, Class achievement and School Achievement, it was found that influence in determining student achievement.

Accuracy in determining student achievement by looking at subject score IPA and subject score Religion, by selecting several academic scores, class achievements, and school achievements in determining student achievement. For further research, it is recommended to add other attributes to determine student achievement and increase the amount of data in the dataset.

## 5. REFERENCES

- [1] M. B. Abdul Majid, Y. M. Cani, and U. Enri, "Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 355, Dec. 2022, doi: 10.30865/json.v4i2.5299.
- [2] H. Yuwafi, F. Marisa, and I. Darma Wijaya, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN SANTRI BERPRESTASI DI PP.MANAARULHUDA DENGAN METODE CLUSTERING ALGORITMA K-MEANS," 2019. doi: DOI: <http://dx.doi.org/10.53567/spirit.v11i1.115>.
- [3] T. Widiyanti, S. Shofiah Hilabi, A. Hananto, Tukino, and E. Novalia, "Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi," *Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi*, vol. 5, 2023, doi: <https://doi.org/10.37034/jidt.v5i1.255>.
- [4] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *Jurnal Sistim Informasi dan Teknologi*, pp. 28–33, Mar. 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [5] J. Hutagalung, Y. Hendro Syahputra, Z. Pertiwi Tanjung, S. Triguna Dharma, and J. I. Pintu Air, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *Hal AH Nasution*, vol. 9, no. 1, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [6] M. Sholeh and K. Aeni, "PERBANDINGAN EVALUASI METODE DAVIES BOULDIN, ELBOW DAN SILHOUETTE PADA MODEL CLUSTERING DENGAN MENGGUNAKAN ALGORITMA K MEANS," 2023. doi: <http://dx.doi.org/10.30998/string.v8i1.16388>.
- [7] M. Sinan, J. Leng, K. Shah, and T. Abdeljawad, "Advances in numerical simulation with a clustering method based on K-means algorithm and Adams Bashforth scheme for fractional order laser chaotic system," *Alexandria Engineering Journal*, vol. 75, pp. 165–179, Jul. 2023, doi: 10.1016/j.aej.2023.05.080.
- [8] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [9] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," vol. 15, no. 2.
- [10] Z. Guo, Y. Shi, F. Huang, X. Fan, and J. Huang, "Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management," *Geoscience Frontiers*, vol. 12, no. 6, Nov. 2021, doi: 10.1016/j.gsf.2021.101249.
- [11] S. Kurniawan, A. M. Siregar, and H. Y. Novita, "Penerapan Algoritma K-Means dan Fuzzy C-Means Dalam Mengelompokkan Prestasi Siswa Berdasarkan Nilai Akademik," vol. IV, no. 1, 2023.
- [12] E. Fernando, P. Mudjiraharjo, and M. Aswin, "IMPLEMENTASI PENDEKATAN COLLABORATIVE FILTERING DAN K-MEANS CLUSTERING PADA SISTEM REKOMENDASI MATA KULIAH," *Jurnal Informatika dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 5, no. 2, 2022, doi: 10.33387/jiko.
- [13] M. A. Ayub, "ANALISIS TOPIK EKONOMI DENGAN ALGORITMA K-MEANS PADA MEDIA ONLINE ERA PANDEMI COVID-19 DI SULAWESI TENGGARA," *Jurnal Informatika dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 4, no. 2, 2021, doi: 10.33387/jiko.
- [14] R. Anggraini, E. Haerani, J. Jasril, and I. Afrianty, "Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 6, p. 1840, Dec. 2022, doi: 10.30865/jurikom.v9i6.5145.
- [15] F. Salsabila and S. Maulida Intani, "IMPLEMENTASI ALGORITMA K-MEANS DAN C4.5 DALAM MENENTUKAN TINGKAT PENYEBARAN COVID-19 DI INDONESIA," *Kata Kunci : Algoritma C4*, vol. 7, no. 1, 2021.
- [16] A. H. Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," vol. 7, no. 2, 2021, doi: <https://doi.org/10.35329/jiik.v7i2.203>.
- [17] E. Laber, L. Murtinho, and F. Oliveira, "Shallow decision trees for explainable k-means clustering," 2022. doi: <https://doi.org/10.1016/j.patcog.2022.109239>.