

Implementation of the K-Means Clustering Method for Selecting Pencak Silat Athletes at the North Sumatra KONI

Abdi Alamsyah¹, Rahmat Kurniawan R²

^{1,2}Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara Medan

Article Info

Article history:

Received 12 01, 2024

Revised 12 13, 2024

Accepted 12 26, 2024

Keywords:

Athlete Selection

Classification

Data Clustering

K-Means Clustering

Physical Ability

ABSTRACT

This article discusses the application of the K-Means clustering method in the selection of pencak silat athletes at the North Sumatra Koni. The study aims to improve the efficiency of the selection process, which was previously carried out manually, by classifying athletes based on attributes such as age, weight, and physical ability. K-Means clustering is used to group athletes into three categories: very suitable, suitable, and unsuitable. The methodology includes determining the number of clusters, calculating centroids, and analyzing Euclidean distances for data clustering. The results of the study showed that this method was effective in classifying 394 athletes in the "unsuitable" category, 354 athletes in the "suitable" category, and 376 athletes in the "very suitable" category. This study is expected to support the athlete selection process more systematically and efficiently. Recommendations for further research include the use of additional criteria and exploration of other clustering methods for more optimal results.

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



Corresponding Author:

Abdi Alamsyah

Program Studi Ilmu Komputer

Universitas Islam Negeri Sumatera Utara

Sumatra Utara, Indonesia

Email: abdi.alam.syah.30.agus@gmail.com

© The Author(s) 2024

1. Introduction

In the era of globalization, the rapid development of increasingly sophisticated technology has made data one of the most important components of the digital age. All data-related activities are recorded and stored in data storage media or databases, allowing this information to be processed and turned into valuable insight[1]. Currently, the field of data science is growing rapidly, driven by the increasing demands of human society. In this context, data science plays a crucial role, particularly in terms of data processing [2].

Athlete selection is a critical process that determines a team's quality and performance in competitions. The goal of this process is to identify individuals with the greatest potential, thereby increasing the chances of success in the competition [3]. With the increasing number of athletes and data to be analyzed, traditional manual selection methods are becoming inefficient. Therefore, it is important to adopt a more systematic and data-driven approach [4].

The selection of Pencak Silat athletes at KONI North Sumatra faces challenges due to the manual and time-consuming nature of the current process. This study aims to develop a more efficient and objective selection system by applying K-Means Clustering to group athletes based on physical and technical attributes, such as age, weight, physical condition, and skills in techniques like kicking, punching, throwing, squeezing, and cutting. By using this method, KONI North Sumatra can more effectively identify the most

suitable athletes for each competition category. The data used includes demographic information (age, weight), physical data (results of physical tests), and technical data (assessment of Pencak Silat techniques). The primary data source is the KONI North Sumatra athlete database, with data collection conducted through three methods: documentation from medical records and physical test results, direct observation during training or matches, and questionnaires for subjective assessments by coaches and judges regarding athlete abilities.

The K-Means clustering method was chosen for its ability to efficiently group large and complex datasets [5]. This method allows athletes to be categorized based on various attributes [6]. In this study, the criteria used to assess the effectiveness of the K-Means clustering method include classification accuracy, which measures how precisely the athlete groupings match the specified categories: very suitable, suitable, and unsuitable. This accuracy is calculated by comparing the results of the K-Means clustering classification with previous manual assessments [7]. Additionally, cluster stability is observed to ensure consistency in clustering results across iterations. This stability indicates that the method can produce similar groups even if the data varies. Computation time is another important factor, as the time required for clustering must be efficient, particularly in the context of athlete selection, which demands quick decisions. Interpretability is also evaluated, focusing on how easily the clustering results can be understood and used in decision-making. Clear and easily interpretable results will support the selection process. Finally, resource utilization is considered, with an emphasis on the efficiency of hardware and software in implementing the method [8].

In several previous studies, clustering methods have been commonly used in the selection of Pencak Silat athletes to group data. One such study, conducted by Sulistiyawati & Supriyanto in 2021, successfully implemented clustering for determining superior class students, resulting in effective data grouping classification. The application of the K-Means Clustering Algorithm led to the identification of superior classes, with an information system implementation forming two clusters for each class. The results from the 6-class cluster at SMPN 2 Jati Agung, South Lampung Regency, based on a total of 192 data points, showed that 96 students were classified as superior, while the other 96 were not. The user acceptance level for this application was 97.56%, indicating very good feedback [9]. The key difference between this study and previous research lies in the application of objects, the data collection process, and the method used for grouping the research data.

Data mining is a process that utilizes statistical, mathematical, artificial intelligence, and machine learning techniques to extract valuable information and derive knowledge from large databases [10]. Clustering involves grouping a set of data or objects into clusters (groups) such that the data within each cluster are as similar as possible, while being distinct from objects in other clusters [11]. The patterns of information generated from the data mining process may differ from the initial hypothesis. If this occurs, the results can be used as feedback to refine and improve the data mining process [12].

Clustering in data mining is useful for identifying distribution patterns within a dataset that aid in the data analysis process [13]. Object similarity is typically determined by the proximity of attribute values that describe data objects, which are often represented as points in a multidimensional space [14]. K-Means is a data mining method that groups data into clusters, ensuring that data with similar characteristics are placed in the same cluster, while data with differing characteristics are assigned to different clusters [15]. The goal of clustering is to minimize the objective function defined in the process, which generally seeks to reduce variation within each cluster and maximize variation between clusters [16].

In ancient times, not all regions in Indonesia used the term 'pencak silat' to refer to martial arts activities. Pencak refers to self-defense movements performed as a form of dance and rhythm, following specific rules (customs of politeness), and can be used for demonstrations [17]. Silat represents the essence of pencak, where fighting or self-defense is more than just a performance. Therefore, the term 'pencak silat' literally means 'fighting with art' [18].

The outcome of this study is the classification of athletes into three clusters—very suitable, suitable, and unsuitable—based on attributes such as age, weight, physique, kicks, punches, throws, srekelan, and cuts. These data were collected from pencak silat athletes at Koni North Sumatra from September 2023 to January 2024, totaling 1,150 entries. The purpose of this research is to apply the K-Means Clustering method for selecting pencak silat athletes at Koni North Sumatra who are participating in competitions, categorizing them into various clusters.

2. Research Methods

2.1 *K-Means Clustering Method*

Proses K-Means Clustering pada penelitian ini terdiri dari tahapan - tahapan yang mungkin dilakukan sebagai berikut :

1. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk.
2. Tentukan k *centroid* (titik pusat klaster) awal secara random, kemudian untuk menghitung *centroid* klaster ke-i berikutnya, digunakan rumus sebagai berikut :

$$D(ij) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Keterangan:

D(ij) = Jarak data ke i ke pusat *cluster* j [1]

X_{ki} = Data ke i atribut data ke j [2]

X_{kj} = Titik pusat ke j pada atribut ke k [3]

3. Hitung jarak setiap data ke masing-masing *centroid*.

$$d = \sum_{i=1}^n (x_i - y_i)^2 \quad (2)$$

Keterangan:

x_i = Objek pengamatan ke I [1]

y_i = *Centroid* ke I [2]

n = Banyaknya obyek yang menjadi anggota klaster [3]

4. Setiap data memilih *centroid* yang terdekat. Tentukan posisi *centroid* yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama.
5. Kembali ke langkah-3 jika posisi *centroid* baru dengan *centroid* yang lama tidak sama.

Misalnya saja pada penelitian saya, teori jarak *Euclidean* yang diungkapkan sebagai berikut akan digunakan untuk menghitung jarak antara *objek* dengan data atlet pencak silat menghasilkan nilai rata-rata *centroid* 1 sebagai berikut :

$$d(9) = \sqrt{(18 - 27)^2 + (61 - 59)^2 + (39 - 42)^2 + (26 - 24)^2 + (48 - 64)^2 + (17 - 22)^2 + (21 - 28)^2 + (30 - 18)^2}$$

$$d(9) = 23,91652$$

Dari hasil diatas jarak data atlet pencak silat ke-9 dengan *cluster* pertama adalah 23,91652. Kemudian jarak dari data atlet pencak silat ke-18 ke pusat *cluster* kedua :

$$d(18) = \sqrt{(27 - 21)^2 + (59 - 69)^2 + (42 - 50)^2 + (24 - 34)^2 + (64 - 59)^2 + (22 - 22)^2 + (28 - 22)^2 + (18 - 21)^2}$$

$$d(18) = 19,23538$$

Dari hasil diatas jarak data atlet ke-18 dengan pusat *cluster* kedua adalah 19,23538.

2.2 Tahap Perencanaan

Penulis penelitian ini membuat *flowchart* atau plot berikut untuk berbagai tahapan penelitian :

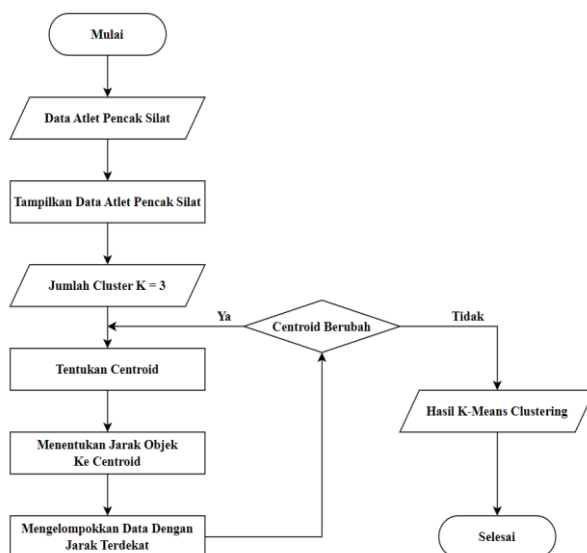


Figure 2.1 *Flowchart* Penelitian

Berikut penjelasan *flowchart* diatas :

1. Tersedia rincian data atlet pencak silat dari seleksi yang dilakukan koni sumatera utara, termasuk informasi tentang umur, fisik, tendangan, pukulan, bantingan, srekelan, dan guntingan.
2. Menampilkan data atlet dari seleksi yang dilakukan koni sumatera utara.

3. Menggunakan metode K-Means untuk menganalisis data atlet pencak silat yang bertanding. Langkah pertama dalam analisis adalah mencari tahu berapa banyak cluster yang ada (k). Tiga cluster yaitu tidak pantas, pantas, dan sangat pantas yang telah diidentifikasi untuk jurnal ini.
4. Setelah jumlah cluster ditentukan selanjutnya menentukan centroid awal.
5. Rumus Euclidean digunakan untuk menghitung jarak setiap titik data yang ada dari setiap centroid setelah nilai centroid pertama ditetapkan. Setelah jarak setiap titik data diukur dari centroid, titik-titik data tersebut dikelompokkan menurut seberapa dekat jarak titik tersebut ke centroid.
6. Selanjutnya, susun datanya sesuai dengan jarak centroid terdekat yang telah temukan menggunakan rumus Euclidean.
7. Pada saat menentukan centroid, maka:
 - a. Jika centroid baru berubah, maka proses dilanjut ke langkah berikutnya.
 - b. Jika centroid baru tidak berubah, maka proses perhitungan selesai.
8. Selanjutnya, setelah mendapatkan centroid maka dapat hasil K-Means Clustering dari pengelompokan data dengan jarak centroid tersebut.

3. Results and Discussion

3.1. Penerapan K-Means Clustering

Pemilihan sampel dataset yang akan digunakan merupakan tahap awal dalam penerapan teknik pengelompokan data. Kumpulan data sampel terdiri dari 8 variabel (umur, fisik, berat badan, tendangan, pukulan, bantingan, srekelan, dan guntingan) dan data 18 atlet pencak silat. Informasi mengenai 18 peserta pencak silat akan dibagi menjadi tiga kelompok tidak pantas (C3), pantas (C2), dan sangat pantas (C1).

Table 3.1 Data Atlet Pencak Silat

No	Umur	Fisik	Berat Badan	Tendangan	Pukulan	Bantingan	Srekelan	Guntingan
1	18	61	39	26	48	17	21	30
2	33	56	35	21	39	25	15	20
3	39	66	49	31	46	14	27	15
4	20	67	36	32	51	11	18	21
5	13	59	38	24	55	25	21	20
6	23	64	39	29	53	13	24	23
7	22	60	39	25	51	16	18	31
8	24	66	36	31	59	12	25	18
9	27	59	42	24	64	22	28	18
10	28	60	52	35	65	12	25	17
11	22	62	38	27	68	24	23	27
12	30	60	51	25	41	22	27	32
13	26	63	39	28	49	20	30	20
14	27	56	35	21	54	21	18	15
15	26	66	51	31	59	20	30	33
16	26	67	41	32	62	19	20	17
17	30	69	45	34	67	21	28	24
18	21	69	50	34	59	22	22	21

Centroid pusat pertama harus ditentukan secara acak setelah menentukan kumpulan data. Centroid pusat pertama pada penelitian ini adalah Centroid 1 yang berasal dari Data 1. Centroid kedua yang diambil dari data 9. dan Centroid 3 yang berasal dari data 1

Table 3.2 Centroid

Centroid	Umur	Fisik	Berat Badan	Tendangan	Pukulan	Bantingan	Srekelan	Guntingan
Awal								

<i>Centroid 1</i>	18	61	39	26	48	17	21	30
<i>Centroid 2</i>	27	59	42	24	64	22	28	18
<i>Centroid 3</i>	21	69	50	34	59	22	22	21

Setelah centroid pusat pertama yang pertama telah ditetapkan, maka menghasilkan hasil perhitungan Iterasi 1 yang ditunjukkan pada Table di bawah ini :

Table 3.3 Hasil Iterasi 1

No	Umur	Fisik	Berat Badan	Tendangan	Pukulan	Bantingan	Srekelan	Guntingan	C1	C2	C3	Cluster
1	18	61	39	26	48	17	21	30	0	23,9	22,0	1
2	33	56	35	21	39	25	15	20	23,9	30,1	34,1	1
3	39	66	49	31	46	14	27	15	29,4	26,2	25,2	3
4	20	67	36	32	51	11	18	21	14,8	24,7	20,1	1
5	13	59	38	24	55	25	21	20	15,7	18,8	20,8	1
6	23	64	39	29	53	13	24	23	11,9	17,8	17,3	1
7	22	60	39	25	51	16	18	31	6,1	22,5	22,3	1
8	24	66	36	31	59	12	25	18	20,0	16,6	18,4	2
9	27	59	42	24	64	22	28	18	23,9	0	19,2	2
10	28	60	52	35	65	12	25	17	29,1	18,2	17,2	3
11	22	62	38	27	68	24	23	27	21,9	13,6	19,1	2
12	30	60	51	25	41	22	27	32	20,0	28,6	26,7	1
13	26	63	39	28	49	20	30	20	16,2	16,7	19,6	1
14	27	56	35	21	54	21	18	15	20,8	16,6	26,0	2
15	26	66	51	31	59	20	30	33	21,8	20,9	16	3
16	26	67	41	32	62	19	20	17	22,5	14,4	12,3	3
17	30	69	45	34	67	21	28	24	27,7	16,2	14,6	3
18	21	69	50	34	59	22	22	21	22,0	19,2	0	3

Berdasarkan perhitungan menggunakan rumus clustering saat ini, pengelompokan berikut ditentukan berdasarkan jarak terdekat ke pusat centroid terdekat :

Cluster Lama (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)

Cluster Baru (1 1 3 1 1 1 1 1 2 2 3 2 1 1 2 3 3 3 3)

Pengelompokan data telah berubah, dan penghitungan akan dilanjutkan pada iterasi berikutnya.

Adapun 3 centroid baru yang akan di iterasikan kembali sebagai berikut:

C1 = (23,125, 61,25, 39,5, 26,25, 48,375, 18,625, 21,75, 24,625)

C2 = (25, 60,75, 37,75, 25,75, 61,25, 19,75, 23,5, 19,5)

C3 = (28,33, 66,167, 48, 32,83, 59,667, 18, 25,333, 21,167)

Dari perhitungan jarak centroid terdekat, maka diperoleh hasil perhitungan Iterasi 2 pada Table dibawah ini.

Table 3.4 Hasil Iterasi 2

No	Umur	Fisik	Berat Badan	Tendangan	Pukulan	Bantingan	Srekelan	Guntingan	C1	C2	C3	Cluster
1	18	61	39	26	48	17	21	30	7,67	18,71	22,25	1
2	33	56	35	21	39	25	15	20	19,19	26,67	31,91	1
3	39	66	49	31	46	14	27	15	23,12	25,99	19,01	3
4	20	67	36	32	51	11	18	21	13,43	17,89	19,81	1
5	13	59	38	24	55	25	21	20	14,87	14,94	23,56	1
6	23	64	39	29	53	13	24	23	8,72	12,36	14,26	1

7	22	60	39	25	51	16	18	31	8,55	17,12	21,21	1
8	24	66	36	31	59	12	25	18	16,41	11,34	14,58	2
9	27	59	42	24	64	22	28	18	19,23	7,95	14,80	2
10	28	60	52	35	65	12	25	17	25,43	19,51	11,86	3
11	22	62	38	27	68	24	23	27	20,63	11,50	18,34	2
12	30	60	51	25	41	22	27	32	18,17	28,02	24,40	1
13	26	63	39	28	49	20	30	20	10,31	14,32	16,14	1
14	27	56	35	21	54	21	18	15	15,30	12,70	23,38	2
15	26	66	51	31	59	20	30	33	20,94	21,47	13,56	3
16	26	67	41	32	62	19	20	17	17,98	10,45	10,39	3
17	30	69	45	34	67	21	28	24	24,28	16,99	9,95	3
18	21	69	50	34	59	22	22	21	19,29	17,79	9,73	3

Berdasarkan perhitungan menggunakan rumus *clustering* saat ini, pengelompokan berikut ditentukan berdasarkan jarak terdekat ke pusat *centroid* terdekat :

Cluster Lama (1 1 3 1 1 1 1 2 2 3 2 1 1 2 3 3 3 3)

Cluster Baru (1 1 3 1 1 1 1 2 2 3 2 1 1 2 3 3 3 3)

Tidak ada lagi perubahan pada cluster pada iterasi 1 dan 2 setelah perhitungan menggunakan rumus *clustering* saat ini, sehingga perhitungan iterasi dihentikan.

3.2. Penerapan K-Means Clustering Pada Python (Jupyter Notebook)

3.2.1 Import Data

Saat menggunakan bahasa pemrograman *Python* untuk analisis data, langkah awalnya adalah mengimpor data. *Library* Pandas, yang banyak digunakan dalam analisis data, digunakan untuk mengimpor data. Pastikan *Notebook Jupyter* telah menginstal Pandas. Setelah itu, *impor* perpustakaan dengan membuat kode yang sesuai dengan *Jupyter Notebook* sehingga *Jupyter Notebook* dapat mengaksesnya dengan benar. Data dapat diproses setelah *impor* selesai. Dataset atlet pencak silat dalam format CSV pada tahap *impor* data, dengan 1124 data dengan total 9 atribut yaitu nama, umur, fisik, berat badan, tendangan, pukulan, bantingan, srekelan, dan guntingan [19].

3.2.2 Trnasformais Data

Persyaratan proses pemodelan data dipenuhi oleh tahap transformasi data ini. Proses pengubahan data menjadi data numerik, klasifikasi, dan format lainnya disebut dengan transformasi data [20]. Berikut hasil transformasi datanya.

No.	Nama	Umur	Fisik	Berat Badan	Tendangan	Pukulan	Bantingan	Srekelan	Guntingan
1	BAYU SUMASTA	18.0	61	39	26	48	17	21	30
2	ABD KARIM	33.0	56	35	21	39	25	15	20
3	BAYU SWANDANA	39.0	66	49	31	46	14	27	15
4	SUWANDI	20.0	67	36	32	51	11	18	21
5	WAHYU RAMADHAN	13.0	59	38	24	55	25	21	20
...
1146	NABIL NAWAWI NST	23.0	49	40	34	64	16	16	26
1147	NADIYA SAPUTRI HRP	19.0	44	43	29	45	22	20	17
1148	RAHIMY JELITA DAULAY	21.0	42	55	27	72	16	26	24
1149	SHOLINA ADELA ZAHWA	19.0	43	48	28	69	22	18	19
1150	SUCI KHAYLA	21.0	39	36	24	70	22	25	17

1150 rows x 9 columns

Figure 3.1 Dataset Awal Atlet Pencak Silat

3.2.3 Pre-prorecessing Data

Proses pemilihan dan penyiapan data untuk digunakan sesuai dengan tujuan analisis data disebut dengan langkah seleksi data. Pilih variabel yang akan digunakan dalam proses pengelompokan selama tahap seleksi data [21]. Figure di bawah ini menampilkan hasil atribut pada seleksi data.

	Umur	Fisik	Berat Badan	Tendangan	Pukulan	Bantingan	Srekelan	Guntingan
0	18	61	39	26	48	17	21	30
1	33	56	35	21	39	25	15	20
2	39	66	49	31	46	14	27	15
3	20	67	36	32	51	11	18	21
4	13	59	38	24	55	25	21	20
...
1119	23	49	40	34	64	16	16	26
1120	19	44	43	29	45	22	20	17
1121	21	42	55	27	72	16	26	24
1122	19	43	48	28	69	22	18	19
1123	21	39	36	24	70	22	25	17

1124 rows x 8 columns

Figure 3.2 Hasil *pre-processing* Data Pencak Silat

3.2.4 Proses Clustering

Jupyter Notebook menggunakan bahasa pemrograman *Python 3* untuk memproses *clustering*. Mencari *K* optimal menggunakan *Metode Elbow* dengan keadaan acak: 0 merupakan langkah awal dalam proses *clustering* [21].

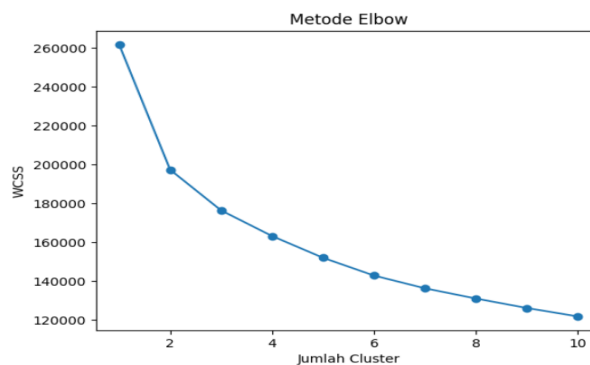


Figure 3.3 Grafik *K* Optimal Data Pencak Silat

Berdasarkan *output* grafik 3 *cluster* merupakan jumlah ideal untuk pendekatan *K-Means* yang menggunakan penilaian pendekatan *Elbow*, seperti terlihat dari keluaran grafik di atas. Pasalnya, grafik di atas menunjukkan sedikit lekukan pada *cluster* nomor 3. Warna merah, hijau, dan biru merupakan 3 *cluster* tersebut. Figure 4.4 menunjukkan hasil *cluster*.

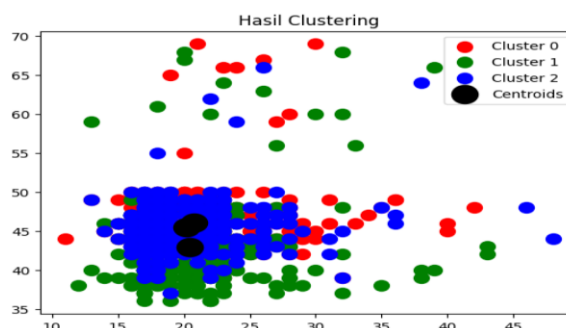


Figure 3.4 Hasil *Clustering* Data Pencak Silat

Centroid atau titik tengah yang muncul pada Figure kemudian ditentukan oleh hasil *clustering*, dan Figure di bawah menunjukkan jumlah anggota setiap *cluste*.

```

[[20.73857868 46.03045685 44.23857868 30.64974619 66.05076142 17.52284264
 21.99492386 19.8680203 ]
[20.47175141 42.9180791 45.45480226 27.18361582 50.24293785 17.57062147
 22.72881356 24.95762712]
[20.19148936 45.50531915 44.69680851 30.29255319 65.29255319 17.72340426
 22.83244681 30.6356383 ]]

```

Figure 3.5 *Centroid Cluster Data Pencak Silat*

Cluster	
0	394
1	354
2	376

Figure 3.6 Jumlah Atlet Pencak Silat Setiap *Cluster*

Terlihat dari perhitungan *Jupyter Notebook* di atas bahwa:

1. *Cluster 0* merupakan kategori tidak pantas dengan 394 atlet dan nilai *centroid* 20.73857868, 46.03045685, 44.23857868, 30.64974619, 66.05076142, 17.52284264, 21.99492386, dan 19.8680203. Umur >20 tahun, fisik 46 kali, berat badan >44 kg, 30 tendangan, 66 pukulan, 17 bantingan, 21 srekelan, dan 19 guntingan merupakan karakteristik dari *cluster 0*.
2. *Cluster 1* merupakan kategori pantas dengan jumlah 354 atlet dan nilai *centroid* 20.47175141, 42.9180791, 45.45480226, 27.18361582, 50.24293785, 17.57062147, 22.72881356, dan 24.95762712. Umur >20 tahun, fisik 42 kali, berat badan >45 kg, 27 tendangan, 50 pukulan, 17 bantingan, 22 srekelan, dan 24 guntingan merupakan karakteristik dari *cluster 1*.
3. *Cluster 2* merupakan kategori sangat pantas dengan jumlah 376 atlet dan nilai *centroid* 20.19148936, 45.50531915, 44.69680851, 30.29255319, 65.29255319, 17.72340426, 22.83244681, dan 30.6356383. Umur >20 tahun, fisik 45 kali, berat badan >44 kg, 30 tendangan, 65 pukulan, 18 bantingan, 23 srekelan, dan 31 guntingan merupakan karakteristik dari *cluster 2*.

4. Conclusion

Based on the research findings, the implementation of the K-Means Clustering method was successfully applied to the selection of pencak silat athletes at the North Sumatra KONI. The study results show that by using the K-Means Clustering method, three groups (clusters) were successfully created based on data characteristics. Cluster 0 represents the 'inappropriate' category, consisting of 394 athletes. The characteristics of this cluster include: age > 20 years, physical test score 46, weight > 44 kg, 30 kicks, 66 punches, 17 throws, 21 srekelan, and 19 cuts. Cluster 1 represents the 'suitable' category, with 354 athletes. The characteristics of this cluster are: age > 20 years, physical test score 42, weight > 45 kg, 27 kicks, 50 punches, 17 throws, 22 srekelan, and 24 cuts. Cluster 2 represents the 'very suitable' category, with 376 athletes. The characteristics of this cluster include: age > 20 years, physical test score 45, weight > 44 kg, 30 kicks, 65 punches, 18 throws, 23 srekelan, and 31 cuts.

References

- [1] Andini, Y., Hardinata, J. T., Purba, Y. P., Studi, P., Informasi, S., Utara, S., & Apriori, M. (2022). Penerapan Data Mining Terhadap Tata Letak Buku. *Jurnal Technology Informatics & Computer System*, *XI*(1), 9–15.
- [2] Debora Mait, C., Armando Watuseke, J., David Gibrael Saerang, P., Reynaldo Joshua, S., & Sam Ratulangi, U. (2022). Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan. *Jurnal Media Infotama*, *18*(2), 344–353.
- [3] Qasthari, M. W., & Kurniawan, R. (2024). Penggunaan Algoritma K-Means Clustering untuk Mengelompokkan Pemain Berdasarkan Gaya Bermain Pada Battle Royale Call of Duty Mobile. *Future Academia*, *2*(3), 280-292. <https://doi.org/10.61579/future.v2i3.177>
- [4] Dinata, R. K., Safwandi, S., Hasdina, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, *5*(1), 10. <https://doi.org/10.19184/isj.v5i1.17071>

- [5] Ediyono, S., & Widodo, S. T. (2019). Memahami Makna Seni dalam Pencak Silat. *Panggung*, 29(3). <https://doi.org/10.26742/panggung.v29i3.1014>
- [6] Pane, P. P., Nasution, Y. R., & Furqan, M. (2024). Implementasi Data Mining dengan K-Means Clustering untuk Memprediksi Pengadaan Obat. *Journal of Computer System and Informatics (JoSYC)*, 5(2), 286-296. <https://doi.org/10.47065/josyc.v5i2.4920>
- [7] Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>
- [8] Syahrin, E., Santony, J., & Na'am, J. (2019). Pemodelan Penjualan Produk Herbal menggunakan Metode Monte Carlo. *Jurnal KomtekInfo*, 5(3), 33-41. <https://doi.org/10.35134/komtekinfo.v5i3.29>
- [9] Pilendia, D. (2020). Pemanfaatan Adobe Flash Sebagai Dasar Pengembangan Bahan Ajar Fisika: Studi Literatur. *Jurnal Tunas Pendidik.*, 2(2), 1-10.
- [10] I Komang Setia Buana. (2020). Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python. *Jurnal Sistem Dan Informatika (JSI)*, 14(2), 135–142. <https://doi.org/10.30864/jsi.v14i2.293>
- [11] Ramadhan, R. S., Kurniawan, R., & Hasibuan, M. S. (2024). Clustering Daerah Rawan Angin Puting Beliung pada Kabupaten di Sumatera Utara dengan Algoritma K-Means. *Jurnal Ilmiah KOMPUTASI*, 23(2), 247-252. <https://doi.org/10.32409/jikstik.23.2.3578>
- [12] Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>
- [13] Mustika, M., Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, G., Manuhutu, M. A., Ridwan, M., Hozairi, H., Wardhani, A. K., Alim, S., Romli, I., & Religia, Y. (2021). *Data Mining dan Aplikasinya*. Penerbit Widina.
- [14] Nofiyanti, E., & Oki Nur Haryanto, E. M. (2021). Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia. *Jurnal Ilmiah SINUS*, 19(2), 17. <https://doi.org/10.30646/sinus.v19i2.5>
- [15] Primanda, R. P., Alwi, A., & Mustikasari, D. (2021). Data Mining Seleksi Siswa Berprestasi Untuk Menentukan Kelas Unggulan Menggunakan Metode K-Means Clustering (Studi Kasus Di Mts Darul Fikri). *Komputek*, 5(1), 88. <https://doi.org/10.24269/Jkt.V5i1.686>
- [16] Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 62. <https://doi.org/10.26418/jp.v5i1.29611>
- [17] Suriani, L. (2020). Pengelompokan Data Kriminal pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal menggunakan Data Mining Algoritma K-Means Clustering. *Jurnal Sist. Komput. dan Inform.*, 1(2), 151. <https://doi.org/10.30865/json.v1i2.1955>
- [18] Nasution, M. Z., & Hasibuan, M. S. (2020). Pendekatan Initial Centroid Search untuk Meningkatkan Efisiensi Iterasi Klustering K-Means. *Techno.Com*, 19(4), 341-352. <https://doi.org/10.33633/tc.v19i4.3875>
- [19] Ratna, S. (2020). Pengolahan Citra Digital Dan Histogram Dengan Phyton Dan Text Editor Phycharm. *Technologia: Jurnal Ilmiah*, 11(3), 181. <https://doi.org/10.31602/tji.v11i3.3294>
- [20] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25. <https://doi.org/10.33365/jtk.v15i2.1162>
- [21] Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
- [22] Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20–28. <https://doi.org/10.58602/jaiti.v1i1.22>
- [23] Fathoroni, A., Fathonah, N. S., Andarsyah, R., & Riza, N. (2020). Sistem Pendukung Keputusan Penilaian Kinerja Dosen Menggunakan Metode 360 Degree Feedback. *Kreatif Industri Nusantara*.
- [24] Wira, B., Budiarto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *RAINSTEK: Jurnal Terapan Sains & Teknologi*, 1(3), 53–68. <https://doi.org/10.21067/jtst.v1i3.3046>
- [25] Kusnadi, Y., & Putri, M. S. (2021). Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah. *Jurnal Teknologi Informatika Dan Komputer*, 7(1), 17–24. <https://doi.org/10.37012/jtik.v7i1.498>
- [26] Gusman, A. P. (2019). Analisa Perancangan dan Implementasi Pemesanan Secara Online Berbasis Customer Relationship Management (CRM). *Majalah Ilmiah UPI YPTK*, 26(1), 7–13. <https://doi.org/10.35134/jmi.v26i1.17>
- [27] amtelahitu, T. M. (2020). Komparasi Algoritma Clustering dengan Dataset Penyebaran Covid-19 di Indonesia Periode Maret-Mei 2020. *Jurnal Teknologi Technoscintia*, 13(1), 27–34.