

Penerapan Metode K-Means Clustering Untuk Seleksi Atlet Taekwondo Porprov

Maulina Tria Audina Gultom^{*}, Raissa Amanda Putri

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,*}maulinatriaaudina@gmail.com, ²raissa.ap@uinsu.ac.id

Email Penulis Korespondensi: maulinatriaaudina@gmail.com

Abstrak—Dengan adanya peningkatan jumlah data atlet taekwondo pada Koni Sumut tentunya memiliki data yang berbeda-beda. Pada Koni Sumut proses seleksi atlet taekwondo masih menggunakan microsoft excel. Proses seleksi menggunakan microsoft excel ini masih tidak tepat karena resiko terjadinya eror pada penginputan data dan butuh waktu lama untuk membandingkan data sebelumnya dengan data hasilnya. Tujuan penulisan tugas akhir ini yaitu untuk menerapkan metode K-Means Clustering untuk data seleksi atlet taekwondo yang mengikuti ajang Porprov kedalam beberapa cluster. Maka dari itu penulis tertarik mengambil topik menerapkan metode K-Means Clustering untuk seleksi atlet taekwondo Porprov pada Koni Sumut. Berdasarkan Metode Elbow dapat dilihat jumlah cluster optimal untuk metode K-Means adalah 3 Cluster. Cluster 0 kategori layak dengan jumlah atlet paling sedikit berjumlah 213 atlet. Cluster 1 kategori tidak layak yang berjumlah 57 atlet. Cluster 2 kategori sangat layak dengan jumlah atlet terbanyak berjumlah 216 atlet. Hal ini dikarenakan pada jumlah cluster 3 yang terdiri dari tidak layak, layak, dan sangat layak. Sesuai dengan hasil dari dilakukannya penelitian ini, yakni dengan menerapkan metode K-Means Clustering Untuk Seleksi Atlet Taekwondo Porprov Pada Koni Sumut berhasil diterapkan. Kesimpulan yang dapat ditarik dari penelitian ini yaitu dengan menggunakan metode K-Means Clustering untuk seleksi atlet taekwondo porprov berhasil menghasilkan tiga kelompok (cluster) berdasarkan karakteristik yang ada pada data tersebut. Hasil dari dilakukannya penelitian ini juga dapat dijadikan sebagai landasan dalam mengembangkan aplikasi serta berbagai penelitian yang dilakukan kedepannya dengan melakukan perbandingan dalam penggunaan metode K-Means Clustering dalam upaya mengelompokkan data atlet yang terdapat pada KONI SUMUT.

Kata Kunci: Clustering; Data Mining; Atlet; Python; Jupyter Notebook

Abstract—With the increase in the number of data on taekwondo athletes in North Sumatra Koni, of course the data will be different. In Koni North Sumatra, the selection process for taekwondo athletes still uses Microsoft Excel. The selection process using Microsoft Excel is still not appropriate because of the risk of errors in data input and it takes a long time to compare previous data with the resulting data. The purpose of writing this final assignment is to apply the K-Means Clustering method for selection data for taekwondo athletes who took part in the Porprov event into several clusters. Therefore, the author is interested in taking the topic of applying the K-Means Clustering method for the selection of Porprov taekwondo athletes at Koni North Sumatra. Based on the Elbow Method, it can be seen that the optimal number of clusters for the K-Means method is 3 Clusters. Cluster 0 feasible category with the least number of athletes totaling 213 athletes. Cluster 1 category is not feasible which amounted to 57 athletes. Cluster 2 category is very feasible with the highest number of athletes totaling 216 athletes. This is because the number of clusters is 3 which consists of not feasible, feasible and very feasible. In accordance with the results of this research, namely by applying the K-Means Clustering method for Selection of Provincial Taekwondo Athletes in Koni North Sumatra, it was successfully implemented. The conclusion that can be drawn from this research is that using the K-Means Clustering method for selecting provincial taekwondo athletes succeeded in producing three groups (clusters) based on the characteristics contained in the data. The results of this research can also be used as a basis for developing applications as well as various research carried out in the future by making comparisons in the use of the K-Means Clustering method in an effort to group athlete data contained in KONI SUMUT.

Keywords: Clustering; Data Mining; Athlete; Python; Jupyter Notebooks

1. PENDAHULUAN

Saat ini penggunaan teknologi sangat berpengaruh dalam hal apapun. Perkembangan teknologi ini begitu cepat dan tidak terbandung. Pada saat ini semakin mudah memperoleh informasi dan memproses penginputan data dan semakin cepat jumlah data bertambah [1]. Perkembangan informasi dan teknologi dengan menggunakan data yang besar telah menjadi suatu pengetahuan baru yang dikenal dengan teknologi data mining. Data mining sudah banyak digunakan oleh peneliti saat ini [2]. Salah satu metode data mining yang sering digunakan adalah *K-Means Clustering* [3].

Berhubungan dengan penginputan data dan penggunaan sistem yang terjadi di Koni Sumut adanya peningkatan jumlah data atlet taekwondo pada Koni Sumut tentunya memiliki data yang berbeda-beda. Pada Koni Sumut proses seleksi atlet taekwondo ini masih menggunakan Microsoft Excel. Proses seleksi menggunakan Microsoft Excel ini masih tidak berjalan dengan tepat karena resiko terjadinya eror pada penginputan data dan butuh waktu lama untuk membandingkan data sebelumnya dengan data hasilnya. Maka dari itu penulis tertarik mengambil topik penerapan metode *K-Means Clustering* untuk seleksi atlet taekwondo Porprov pada Koni Sumut. Pengelompokkan umur, berat badan, sikap kuda-kuda, tendangan, dan pukulan atlet taekwondo merupakan salah satu tujuan untuk mempermudah Koni Sumut dalam menyeleksi atlet taekwondo yang akan mengikuti pertandingan Porprov.

Penyeleksian atlet taekwondo ini metode *clustering* cukup sering diterapkan, dalam beberapa penelitian terdahulu dalam proses pengelompokkan data. Seperti penelitian yang dilakukan Saragih, *et al.*, pada tahun 2018 yang telah sukses diterapkan, Proses seleksi calon peserta yang berhak mewakili daerah dinilai belum cukup efektif dan efisien. Ditemukan beberapa tindakan kecurangan dalam proses pemberian nilai dengan cara melakukan manipulasi terhadap nilai tersebut. Maka solusi yang dapat diterapkan adalah dengan menggunakan metode *K-Means Clustering* untuk mengelompokkan data tersebut. Hal itu dilakukan dengan menetapkan jumlah *cluster*, menghitung jarak terdekat terhadap pusat *cluster*.

Sesuai hasil dari pengujian tersebut menampilkan bahwa total *cluster* 0 terdapat 35 items, *cluster* 1 terdapat 28 items, dan *cluster* 2 terdapat 10 items, dengan total sejumlah 36 items [4]. Hal yang membedakan penelitian terdahulu dengan penelitian ini adalah dalam penerapan objek, proses pengumpulan, serta proses mengelompokkan data penelitiannya.

Penelitian menggunakan *data mining* merupakan tahap untuk menentukan model maupun pola yang baru dan lebih lengkap dan lebih mudah untuk dipahami dengan jumlah data yang besar [5]. Pola tersebut nantinya akan mampu menyajikan analisis data yang lebih praktis dan bermanfaat dan akan dilakukan identifikasi lebih lanjut terhadap data tersebut. Dengan menerapkan alat pendukung keputusan yang berbeda dari sebelumnya [6]. Proses data mining bermula dari dilakukannya pemilihan data yang berasal dari sumber dan target yang telah ditetapkan, yang dilanjutkan dengan tahap *pre-processing* guna memaksimalkan kualitas dari data yang digunakan sehingga proses interpretasi dan evaluasi yang dilakukan pada tahap selanjutnya dapat dilakukan lebih baik lagi [7]. proses pengkodean KDD adalah proses transformasi jenis atau model data yang diambil dari database [8]. Teknik, metode, atau algoritma proses pengumpulan data yang memiliki beberapa variasi [9]. Tahapan terakhir adalah evaluasi [10].

Beberapa proses *data mining* yang dilakukan selanjutnya *Clustering* atau pengelompokan adalah konsep *data mining* yang diterapkan dalam proses Analisa data dan melakukan pemecahan terhadap masalah yang mempunyai kaitan dalam proses pengelompokan datanya dan dijadikan dalam suatu bentuk himpunan bagian [11]. Pada konsep *Clustering*, tujuannya adalah untuk membagi kasus-kasus menjadi sebuah kelompok [12]. Konsep ini melakukan pengelompokan data menjadi beberapa bagian yang sesuai dengan kriteria data yang digunakan [13]. *K-Means* adalah satu diantara beberapa algoritma partisi dikarenakan *K-Means* mempunyai dasar untuk menentukan total kelompok awal sesuai dengan rata-rata nilainya [14]. Menentukan K-Method secara acak dengan menggunakan rumus *Euclidean* [15].

Sejarah taekwondo ini berkembang di negara Korea sejak tahun 37M [16]. Belajar taekwondo bukan hanya berbicara tentang bagaimana cara memukul, menangkis dan menendang saja. Tetapi belajar taekwondo kita melatih tingkat kewaspadaan, kemampuan menganalisa situasi dan menyusun strategi saat keadaan tidak memungkinkan [17]. Tujuan mempelajari olahraga taekwondo ialah untuk kebugaran jasmani, sebagai beladiri, dan yang paling penting ialah untuk prestasi dalam ajang pertandingan [18]. *Python* adalah bahasa pemrograman yang bersifat open source [19]. *Jupyter Notebook* adalah aplikasi web gratis yang digunakan untuk membuat dan berbagi dokumen yang berisi kode, proses, analisis, visualisasi, dan kemampuan pemrosesan data lainnya [20].

Hasil akhir penelitian ini ialah mengelompokkan *cluster* sangat layak, layak dan tidak layak dengan karakteristik berat badan, sikap kuda-kuda, tendangan, dan pukulan atlet taekwondo yang di dapat dari Koni Sumut pada bulan Juli-November tahun 2022 sebanyak 500 data. Tujuan dari pelaksanaan dan penulisan tugas akhir ini yaitu untuk menerapkan metode *K-Means Clustering* untuk data seleksi atlet taekwondo yang mengikuti ajang Porprov kedalam beberapa *cluster*.

2. METODOLOGI PENELITIAN

2.1 Metode K-Means Clustering

Dalam penelitian ini terdapat prosedur pembentukan *K-Means Clustering* terdapat langkah-langkah yang dapat dilakukan, antara lain :

- a. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk.
- b. 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

X_{ki} = Data ke i atribut data ke j

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

- c. 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

y_i = *Centroid* ke i

n = Banyaknya obyek yang menjadi anggota klaster

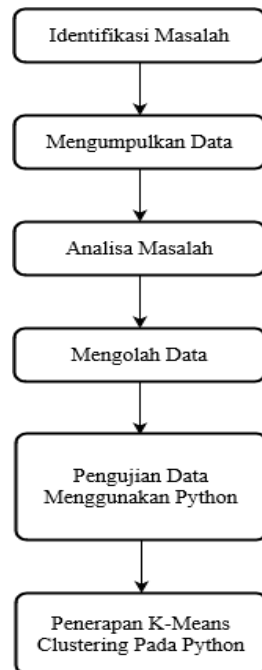
- d. Setiap data memilih *centroid* yang terdekat.
Tentukan posisi *centroid* yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama.
- e. Kembali ke langkah-3 jika posisi *centroid* baru dengan *centroid* yang lama tidak sama.

2.1 Kerangka Penelitian

Adapun penjelasan dari kerangka penelitian sebagai berikut:

- a. Identifikasi masalah, ialah masalah penelitian yang telah dibuat pada jurnal harus dirumuskan dengan jelas.

- b. Mengumpulkan data dengan menggunakan metode studi literatur dan dokumentasi. Studi literatur ini adalah untuk mengumpulkan dan memperoleh informasi yang sesuai dengan cara meneliti, membaca dan mencatat literatur dari beberapa jurnal tentang topik-topik yang sesuai dengan permasalahan pada jurnal ini. Dokumentasi adalah pengumpulan data atau dokumen berbentuk Excel (.xlsx), sampai data terkumpul menjadi satu dan siap untuk kepentingan penelitian.
- c. Menganalisa masalah yang ada pada Koni Sumut dan mencari referensi-referensi teori yang relevan dengan permasalahan yang ada pada penelitian ini.
- d. Mengolah data, data yang diolah pada penelitian ini adalah data seleksi atlet taekwondo dengan menggunakan metode *K-Means Clustering*. Dilakukan k dari jumlah *cluster* baru yang ingin dibentuk. Kemudian tentukan titik pusat awal *cluster* secara acak. Kemudian hitung jarak setiap data ke pusat *cluster* ke *Centroid* terdekat. Lakukan sampai hasil perhitungan *Centroid* tidak berubah maka iterasi dihentikan. Kemudian data di proses dengan *Python*. Karena data akan dimasukkan kedalam suatu library pada *Python*, maka data dirubah ke dalam format “csv” (Comma Delimited). Karena software yang digunakan hanya dapat membaca file format “csv”.
- e. Pengujian data, data yang telah diolah dapat diuji dengan proses-proses yang telah dilakukan. Seperti proses mengolah data, dari pengumpulan data, pra-proses data, transformasi data, proses data mining, dan evaluasi data miningnya dengan metode *K-Means Clustering* pada *Python*.
- f. Penerapan algoritma, sehubungan dengan pengujian yang dilakukan peneliti mencoba untuk melakukan pengelompokkan data dengan menggunakan *K-Means Clustering* dalam menentukan layak dan tidak layak atlet taekwondo untuk dipertandingkan di ajang Porprov. Peneliti menggunakan *Python* untuk mengelompokkan dan menguji data yang sudah ada. Maka dengan hasil pengelompokkan data atlet tersebut, KONI SUMUT dapat dengan mudah menentukan atlet yang ada di *cluster 0*, *cluster 1* dan *cluster 2*.



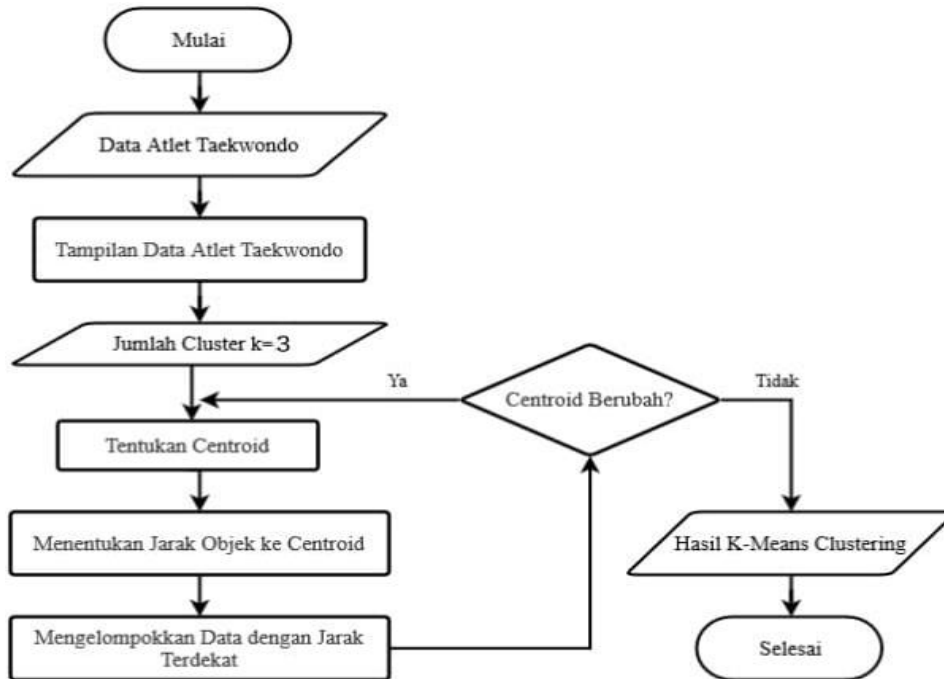
Gambar 1. Kerangka Penelitian

2.2 Tahap Perancangan

Adapun penjelasan dari Flowchart sebagai berikut:

- a. Data atlet taekwondo pada ajang pertandingan Porprov sebanyak 500 data dengan kriteria umur, berat badan, sikap kuda-kuda, tendangan, dan pukulan atlet.
- b. Menampilkan data atlet pada ajang pertandingan Porprov.
- c. Untuk menganalisis data pada ajang pertandingan Porprov dengan *K-Means* dimulai dengan penentuan jumlah *cluster* (*k*). Penentuan jumlah *cluster* pada jurnal ini ada 3 *cluster* yaitu tidak layak, layak dan sangat layak.
- d. Setelah jumlah *cluster* ditentukan selanjutnya menentukan *Centroid* awal.
- e. Setelah nilai *Centroid* awal ditentukan maka dihitung jarak setiap data yang ada terhadap terhadap masing-masing *Centroid* menggunakan rumus Euclidean.
- f. Kemudian mengelompokkan data dengan jarak terdekat *Centroid* yang telah dihitung menggunakan rumus Euclidean.
- g. Pada saat menentukan *Centroid*, maka:
 1. Jika *Centroid* baru berubah, maka proses dilanjut ke langkah berikutnya.
 2. Jika *Centroid* baru tidak berubah, maka proses perhitungan selesai.

- h. Kemudian setelah mendapatkan *Centroid* maka dapat hasil *K-Means Clustering* dari pengelompokan data dengan jarak *Centroid* tersebut.



Gambar 2. Flowchart Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Penerapan *K-Means Clustering*

Dalam menggunakan teknik pengelompokan data, langkah awal ialah menentukan sampel dataset yang akan digunakan. Sampel dataset yang digunakan sebanyak 15 data atlet taekwondo dan 5 atribut yaitu umur, berat badan, kuda-kuda, tendangan dan pukulan. Dari 15 data atlet taekwondo ini akan dikelompokkan menjadi 3 *cluster* yaitu tidak layak (C1), layak (C2), dan sangat layak (C3).

Tabel 1. Data Atlet Taekwondo

No	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan
1	22	63	16	27	47
2	22	50	10	20	40
3	21	60	20	30	48
4	19	65	17	31	50
5	15	56	11	23	56
6	18	58	15	28	55
7	17	58	13	24	50
8	16	70	18	30	60
9	20	68	12	23	65
10	15	76	16	34	64
11	17	80	20	26	68
12	15	80	18	27	50
13	15	62	13	20	55
14	15	74	21	30	60
15	16	67	16	31	61

Setelah menentukan sampel dataset, maka perlu menentukan *Centroid* pusat awal secara random. Pada penelitian ini *Centroid* pusat awalnya menggunakan *Centroid* 1 yang diambil dari data ke 1. *Centroid* 2 yang diambil dari data ke 3. Dan *Centroid* 3 yang diambil dari data ke 13 pada tabel dibawah ini.

Tabel 2. *Centroid* Awal

<i>Centroid</i> Awal	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan
<i>Centroid</i> 1	22	63	16	27	47

Centroid 2	21	60	20	30	48
Centroid 3	15	62	13	20	55

Setelah menentukan Centroid pusat awal, selanjutnya dilakukan perhitungan dengan rumus dibawah ini:

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

Iterasi 1:

- Menghitung jarak Centroid terdekat pada Centroid 1 dengan nilai atribut (22, 63, 16, 27, 47)

$$C1 = \sqrt{(22 - 22)^2 + (63 - 63)^2 + (16 - 16)^2 + (27 - 27)^2 + (47 - 47)^2} = 0$$

$$C2 = \sqrt{(22 - 21)^2 + (63 - 60)^2 + (16 - 20)^2 + (27 - 30)^2 + (47 - 48)^2} = 6$$

$$C3 = \sqrt{(22 - 15)^2 + (63 - 62)^2 + (16 - 13)^2 + (27 - 20)^2 + (47 - 55)^2} = 13,11$$

- Menghitung jarak Centroid terdekat pada Centroid 2 dengan nilai atribut (21, 60, 20, 30, 48)

$$C1 = \sqrt{(21 - 22)^2 + (60 - 63)^2 + (20 - 16)^2 + (30 - 27)^2 + (48 - 47)^2} = 6$$

$$C2 = \sqrt{(21 - 21)^2 + (60 - 60)^2 + (20 - 20)^2 + (30 - 30)^2 + (48 - 48)^2} = 0$$

$$C3 = \sqrt{(21 - 15)^2 + (60 - 62)^2 + (20 - 13)^2 + (30 - 20)^2 + (48 - 55)^2} = 15,42$$

- Menghitung jarak Centroid terdekat pada Centroid 3 dengan nilai atribut (15, 62, 13, 20, 55)

$$C1 = \sqrt{(15 - 22)^2 + (62 - 63)^2 + (13 - 16)^2 + (20 - 27)^2 + (55 - 47)^2} = 13,11$$

$$C2 = \sqrt{(15 - 21)^2 + (62 - 60)^2 + (13 - 20)^2 + (20 - 30)^2 + (55 - 48)^2} = 15,42$$

$$C3 = \sqrt{(15 - 15)^2 + (62 - 62)^2 + (13 - 13)^2 + (20 - 20)^2 + (55 - 55)^2} = 0$$

Dari perhitungan jarak Centroid terdekat diatas maka diperoleh hasil perhitungan Iterasi 1 pada tabel dibawah ini.

Tabel 3. Hasil Iterasi 1

No	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan	C1	C2	C3	Cluster
1	22	63	16	27	47	0	6	13,11	1
2	22	50	10	20	40	17,40	19,10	20,66	1
3	21	60	20	30	48	6	0	15,42	2
4	19	65	17	31	50	6,244	6,55	13,67	1
5	15	56	11	23	56	14,83	15,68	7,07	3
6	18	58	15	28	55	10,34	9,53	9,64	2
7	17	58	13	24	50	8,77	10,44	7,81	3
8	16	70	18	30	60	16,34	16,52	14,66	3
9	20	68	12	23	65	19,62	21,61	13,07	3
10	15	76	16	34	64	23,57	24,08	21,95	3
11	17	80	20	26	68	27,78	28,84	24,12	3
12	15	80	18	27	50	18,73	21,28	20,56	1
13	15	62	13	20	55	13,11	15,42	0	3
14	15	74	21	30	60	19,31	19,41	18,24	3
15	16	67	16	31	61	16,24	16,12	13,85	3

Dengan dilakukannya perhitungan dengan menerapkan rumus clustering, selanjutnya pengelompokkan berdasarkan jarak minimum ke Centroid yang paling dekat, yakni:

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

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

Terjadi perubahan pengelompokkan data yang kemudian akan dilanjutkan perhitungan iterasi berikutnya.

Iterasi 2

Menentukan Centroid baru ke-1 berdasarkan dengan nilai setiap atribut yang ada pada cluster 1.

$$C1 = \left(\frac{22+22+19+15}{4}\right) = 19,5$$

$$= \left(\frac{63+50+65+80}{4}\right) = 64,5$$

$$= \left(\frac{16+10+17+18}{4}\right) = 15,25$$

$$= \left(\frac{27+20+31+27}{4}\right) = 26,25$$

$$= \left(\frac{47+40+50+50}{4}\right) = 46,75$$

$$C1 = (19,5, 64,5, 15,25, 26,25, 46,75)$$

Menentukan Centroid baru ke-2 berdasarkan dengan nilai setiap atribut yang ada pada cluster 2.

$$C2 = \left(\frac{21+18}{2}\right) = 19,5$$

$$= \left(\frac{60+58}{2}\right) = 59$$

$$= \left(\frac{20+15}{2}\right) = 17,5$$

$$= \left(\frac{30+28}{2}\right) = 29$$

$$= \left(\frac{48+55}{2}\right) = 51,5$$

C2 = (19,5, 59, 17,5, 29, 51,5)

Menentukan *Centroid* baru ke-3 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 3.

$$C3 = \left(\frac{15+17+16+20+15+17+15+15+16}{9}\right) = 16,22$$

$$= \left(\frac{56+58+70+68+76+80+62+74+67}{9}\right) = 67,88$$

$$= \left(\frac{11+13+18+12+16+20+13+21+16}{9}\right) = 15,55$$

$$= \left(\frac{23+24+30+23+34+26+20+30+31}{9}\right) = 26,77$$

$$= \left(\frac{56+50+60+65+64+68+55+60+61}{9}\right) = 59,88$$

C3 = (16,22, 67,88, 15,55, 26,77, 59,88)

Adapun 3 *Centroid* baru yang akan di iterasikan kembali sebagai berikut:

C1 = (19,5, 64,5, 15,25, 26,25, 46,75)

C2 = (19,5, 59, 17,5, 29, 51,5)

C3 = (16,22, 67,88, 15,55, 26,77, 59,88)

1. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 1 dengan nilai atribut (22, 63, 16, 27, 47)

C1 = (19,5, 64,5, 15,25, 26,25, 46,75)

$$= \sqrt{(22 - 19,5)^2 + (63 - 64,5)^2 + (16 - 15,25)^2 + (27 - 26,25)^2 + (47 - 46,75)^2} = 3,11$$

C2 = (19,5, 59, 17,5, 29, 51,5)

$$= \sqrt{(22 - 19,5)^2 + (63 - 59)^2 + (16 - 17,5)^2 + (27 - 29)^2 + (47 - 51,5)^2} = 6,98$$

C3 = (16,22, 67,88, 15,55, 26,77, 59,88)

$$= \sqrt{(22 - 16,22)^2 + (63 - 67,88)^2 + (16 - 15,55)^2 + (27 - 26,77)^2 + (47 - 59,88)^2} = 14,95$$

2. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 2 dengan nilai atribut (21, 60, 20, 30, 48)

C1 = (19,5, 64,5, 15,25, 26,25, 46,75)

$$= \sqrt{(21 - 19,5)^2 + (60 - 64,5)^2 + (20 - 15,25)^2 + (30 - 26,25)^2 + (48 - 46,75)^2} = 7,79$$

C2 = (19,5, 59, 17,5, 29, 51,5)

$$= \sqrt{(21 - 19,5)^2 + (60 - 59)^2 + (20 - 17,5)^2 + (30 - 29)^2 + (48 - 51,5)^2} = 4,76$$

C3 = (16,22, 67,88, 15,55, 26,77, 59,88)

$$= \sqrt{(21 - 16,22)^2 + (60 - 67,88)^2 + (20 - 15,55)^2 + (30 - 26,77)^2 + (48 - 59,88)^2} = 16,01$$

3. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 3 dengan nilai atribut (15, 62, 13, 20, 55)

C1 = (19,5, 64,5, 15,25, 26,25, 46,75)

$$= \sqrt{(15 - 19,5)^2 + (62 - 64,5)^2 + (13 - 15,25)^2 + (20 - 26,25)^2 + (55 - 46,75)^2} = 11,77$$

C2 = (19,5, 59, 17,5, 29, 51,5)

$$= \sqrt{(15 - 19,5)^2 + (62 - 59)^2 + (13 - 17,5)^2 + (20 - 29)^2 + (55 - 51,5)^2} = 11,94$$

C3 = (16,22, 67,88, 15,55, 26,77, 59,88)

$$= \sqrt{(15 - 16,22)^2 + (62 - 67,88)^2 + (13 - 15,55)^2 + (20 - 26,77)^2 + (55 - 59,88)^2} = 10,60$$

Dari perhitungan jarak *Centroid* terdekat diatas maka diperoleh hasil perhitungan Iterasi 2 pada tabel dibawah ini.

Tabel 4. Hasil Iterasi 2

No	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan	C1	C2	C3	Cluster
1	22	63	16	27	47	3,11	6,98	14,95	1
2	22	50	10	20	40	18,12	18,88	28,73	1
3	21	60	20	30	48	7,79	4,76	16,01	2
4	19	65	17	31	50	6,05	6,53	11,56	1
5	15	56	11	23	56	14,37	11,30	13,89	2
6	18	58	15	28	55	10,75	4,76	11,25	2
7	17	58	13	24	50	8,31	7,39	14,50	2
8	16	70	18	30	60	15,48	14,37	4,56	3
9	20	68	12	23	65	19,14	18,15	8,20	3
10	15	76	16	34	64	22,59	22,19	11,68	3
11	17	80	20	26	68	26,84	27,10	15,27	3
12	15	80	18	27	50	16,70	21,62	15,87	3
13	15	62	13	20	55	11,77	11,94	10,60	3
14	15	74	21	30	60	18,25	18,18	8,88	3
15	16	67	16	31	61	15,64	13,14	4,48	3

Dengan dilakukannya perhitungan dengan menerapkan rumus *clustering*, selanjutnya pengelompokkan berdasarkan jarak minimum ke *Centroid* yang paling dekat, yakni:

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

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

Terjadi perubahan pengelompokkan data yang kemudian akan dilanjutkan perhitungan iterasi berikutnya.

Iterasi 3

Menentukan *Centroid* baru ke-1 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 1.

$$\begin{aligned} C1 &= \left(\frac{22+22+19}{3}\right) = 21 \\ &= \left(\frac{63+50+65}{3}\right) = 59,33 \\ &= \left(\frac{16+10+17}{3}\right) = 14,33 \\ &= \left(\frac{27+20+31}{3}\right) = 26 \\ &= \left(\frac{47+40+50}{3}\right) = 45,66 \end{aligned}$$

$$C1 = (21, 59,33, 14,33, 26, 45,66)$$

Menentukan *Centroid* baru ke-2 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 2.

$$\begin{aligned} C2 &= \left(\frac{21+15+18+17}{4}\right) = 17,75 \\ &= \left(\frac{60+56+58+58}{4}\right) = 58 \\ &= \left(\frac{20+11+15+13}{4}\right) = 14,75 \\ &= \left(\frac{30+23+28+24}{4}\right) = 26,25 \\ &= \left(\frac{48+56+55+50}{4}\right) = 52,25 \end{aligned}$$

$$C2 = (17,75, 58, 14,75, 26,25, 52,25)$$

Menentukan *Centroid* baru ke-3 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 3.

$$\begin{aligned} C3 &= \left(\frac{16+20+15+17+15+15+15+16}{8}\right) = 16,12 \\ &= \left(\frac{70+68+76+80+80+62+74+67}{8}\right) = 72,12 \\ &= \left(\frac{18+12+16+20+18+13+21+16}{8}\right) = 16,75 \\ &= \left(\frac{30+23+34+26+27+20+30+31}{8}\right) = 27,62 \\ &= \left(\frac{60+65+64+68+50+55+60+61}{8}\right) = 60,37 \end{aligned}$$

$$C3 = (16,12, 72,12, 16,75, 27,62, 60,37)$$

Adapun 3 *Centroid* baru yang akan di iterasikan kembali sebagai berikut:

$$C1 = (21, 59,33, 14,33, 26, 45,66)$$

$$C2 = (17,75, 58, 14,75, 26,25, 52,25)$$

$$C3 = (16,12, 72,12, 16,75, 27,62, 60,37)$$

1. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 1 dengan nilai atribut (22, 63, 16, 27, 47)

$$C1 = (21, 59,33, 14,33, 26, 45,66)$$

$$= \sqrt{(22 - 21)^2 + (63 - 59,33)^2 + (16 - 14,33)^2 + (27 - 26)^2 + (47 - 45,66)^2} = 4,47$$

$$C2 = (17,75, 58, 14,75, 26,25, 52,25)$$

$$= \sqrt{(22 - 17,75)^2 + (63 - 58)^2 + (16 - 14,75)^2 + (27 - 26,25)^2 + (47 - 52,25)^2} = 8,52$$

$$C3 = (16,12, 72,12, 16,75, 27,62, 60,37)$$

$$= \sqrt{(22 - 16,12)^2 + (63 - 72,12)^2 + (16 - 16,75)^2 + (27 - 27,62)^2 + (47 - 60,37)^2} = 17,25$$

2. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 2 dengan nilai atribut (21, 60, 20, 30, 48)

$$C1 = (21, 59,33, 14,33, 26, 45,66)$$

$$= \sqrt{(21 - 21)^2 + (60 - 59,33)^2 + (20 - 14,33)^2 + (30 - 26)^2 + (48 - 45,66)^2} = 7,34$$

$$C2 = (17,75, 58, 14,75, 26,25, 52,25)$$

$$= \sqrt{(21 - 17,75)^2 + (60 - 58)^2 + (20 - 14,75)^2 + (30 - 26,25)^2 + (48 - 52,25)^2} = 8,61$$

$$C3 = (16,12, 72,12, 16,75, 27,62, 60,37)$$

$$= \sqrt{(21 - 16,12)^2 + (60 - 72,12)^2 + (20 - 16,75)^2 + (30 - 27,62)^2 + (48 - 60,37)^2} = 18,44$$

3. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 3 dengan nilai atribut (15, 62, 13, 20, 55)

$$C1 = (21, 59,33, 14,33, 26, 45,66)$$

$$= \sqrt{(15 - 21)^2 + (62 - 59,33)^2 + (13 - 14,33)^2 + (20 - 26)^2 + (55 - 45,66)^2} = 12,96$$

$$C2 = (17,75, 58, 14,75, 26,25, 52,25)$$

$$= \sqrt{(15 - 17,75)^2 + (62 - 58)^2 + (13 - 14,75)^2 + (20 - 26,25)^2 + (55 - 52,25)^2} = 8,55$$

$$C3 = (16,12, 72,12, 16,75, 27,62, 60,37)$$

$$= \sqrt{(15 - 16,12)^2 + (62 - 72,12)^2 + (13 - 16,75)^2 + (20 - 27,62)^2 + (55 - 60,37)^2} = 14,31$$

Dari perhitungan jarak *Centroid* terdekat diatas maka perhitungan Iterasi 3 menghasilkan nilai berikut:

Tabel 5. Hasil Iterasi 3

No	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan	C1	C2	C3	Cluster
1	22	63	16	27	47	4,47	8,52	17,25	1
2	22	50	10	20	40	13,22	17,13	32,29	1
3	21	60	20	30	48	7,34	8,61	18,44	1
4	19	65	17	31	50	9,32	9,12	13,34	2
5	15	56	11	23	56	13,19	7,08	18,29	2
6	18	58	15	28	55	10,11	3,27	15,33	2
7	17	58	13	24	50	6,50	3,70	18,30	2
8	16	70	18	30	60	19,33	15,22	3,44	3
9	20	68	12	23	65	21,54	16,90	9,86	3
10	15	76	16	34	64	26,77	23,04	8,403	3
11	17	80	20	26	68	31,20	27,57	11,58	3
12	15	80	18	27	50	22,27	22,53	13,14	3
13	15	62	13	20	55	12,96	8,55	14,31	2
14	15	74	21	30	60	22,73	19,41	5,35	3
15	16	67	16	31	61	18,61	13,59	6,21	3

Dengan dilakukannya perhitungan dengan menerapkan rumus *clustering*, selanjutnya pengelompokkan berdasarkan jarak minimum ke *Centroid* yang paling dekat, yakni:

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

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

Terjadi perubahan *cluster* yang kemudian akan dilanjutkan perhitungan iterasi berikutnya.

Iterasi 4

Menentukan *Centroid* baru ke-1 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 1.

$$C1 = \left(\frac{22+22+21}{3}\right) = 21,66$$

$$= \left(\frac{63+50+60}{3}\right) = 57,66$$

$$= \left(\frac{16+10+20}{3}\right) = 15,33$$

$$= \left(\frac{27+20+30}{3}\right) = 25,66$$

$$= \left(\frac{47+40+48}{3}\right) = 45$$

$$C1 = (21,66, 57,66, 15,33, 25,66, 45)$$

Menentukan *Centroid* baru ke-2 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 2.

$$C2 = \left(\frac{19+15+18+17+15}{5}\right) = 16,8$$

$$= \left(\frac{65+56+58+58+62}{5}\right) = 59,8$$

$$= \left(\frac{17+11+15+13+13}{5}\right) = 13,8$$

$$= \left(\frac{31+23+28+24+20}{5}\right) = 25,2$$

$$= \left(\frac{50+56+55+50+55}{5}\right) = 53,2$$

$$C2 = (16,8, 59,8, 13,8, 25,2, 53,2)$$

Menentukan *Centroid* baru ke-3 berdasarkan dengan nilai setiap atribut yang ada pada *cluster* 3.

$$C3 = \left(\frac{16+20+15+17+15+15+16}{7}\right) = 16,28$$

$$= \left(\frac{70+68+76+80+80+74+67}{7}\right) = 73,57$$

$$= \left(\frac{18+12+16+20+18+21+16}{7}\right) = 17,28$$

$$= \left(\frac{30+23+34+26+27+30+31}{7}\right) = 28,71$$

$$= \left(\frac{60+65+64+68+50+60+61}{7}\right) = 61,14$$

$$C3 = (16,28, 73,57, 17,28, 28,71, 61,14)$$

Adapun 3 *Centroid* baru yang akan di iterasikan kembali sebagai berikut:

$$C1 = (21,66, 57,66, 15,33, 25,66, 45)$$

$$C2 = (16,8, 59,8, 13,8, 25,2, 53,2)$$

$$C3 = (16,28, 73,57, 17,28, 28,71, 61,14)$$

1. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 1 dengan nilai atribut (22, 63, 16, 27, 47)

$$C1 = (21,66, 57,66, 15,33, 25,66, 45)$$

$$= \sqrt{(22 - 21,66)^2 + (63 - 57,66)^2 + (16 - 15,33)^2 + (27 - 25,66)^2 + (47 - 45)^2} = 5,89$$

$$C2 = (16,8, 59,8, 13,8, 25,2, 53,2)$$

$$= \sqrt{(22 - 16,8)^2 + (63 - 59,8)^2 + (16 - 13,8)^2 + (27 - 25,2)^2 + (47 - 53,2)^2} = 9,15$$

$$C3 = (16,28, 73,57, 17,28, 28,71, 61,14)$$

$$= \sqrt{(22 - 16,28)^2 + (63 - 73,57)^2 + (16 - 17,28)^2 + (27 - 28,71)^2 + (47 - 61,14)^2} = 18,68$$

2. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 2 dengan nilai atribut (21, 60, 20, 30, 48)

$$C1 = (21,66, 57,66, 15,33, 25,66, 45)$$

$$= \sqrt{(21 - 21,66)^2 + (60 - 57,66)^2 + (20 - 15,33)^2 + (30 - 25,66)^2 + (48 - 45)^2} = 7,44$$

$$C2 = (16,8, 59,8, 13,8, 25,2, 53,2)$$

$$= \sqrt{(21 - 16,8)^2 + (60 - 59,8)^2 + (20 - 13,8)^2 + (30 - 25,2)^2 + (48 - 53,2)^2} = 10,30$$

$$C3 = (16,28, 73,57, 17,28, 28,71, 61,14)$$

$$= \sqrt{(21 - 16,28)^2 + (60 - 73,57)^2 + (20 - 17,28)^2 + (30 - 28,71)^2 + (48 - 61,14)^2} = 19,70$$

3. Menghitung jarak terdekat *Centroid* yang baru pada *Centroid* 3 dengan nilai atribut (15, 62, 13, 20, 55)

$$C1 = (21,66, 57,66, 15,33, 25,66, 45)$$

$$= \sqrt{(15 - 21,66)^2 + (62 - 57,66)^2 + (13 - 15,33)^2 + (20 - 25,66)^2 + (55 - 45)^2} = 14,16$$

$$C2 = (16,8, 59,8, 13,8, 25,2, 53,2)$$

$$= \sqrt{(15 - 16,8)^2 + (62 - 59,8)^2 + (13 - 13,8)^2 + (20 - 25,2)^2 + (55 - 53,2)^2} = 6,24$$

$$C3 = (16,28, 73,57, 17,28, 28,71, 61,14)$$

$$= \sqrt{(15 - 16,28)^2 + (62 - 73,57)^2 + (13 - 17,28)^2 + (20 - 28,71)^2 + (55 - 61,14)^2} = 16,35$$

Dari perhitungan jarak *Centroid* terdekat diatas maka diperoleh hasil perhitungan Iterasi 4 pada tabel dibawah ini.

Tabel 6. Hasil Iterasi 4

No	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan	C1	C2	C3	Cluster
1	22	63	16	27	47	5,89	9,15	18,68	1
2	22	50	10	20	40	12,01	18,40	34,12	1
3	21	60	20	30	48	7,44	10,30	19,70	1
4	19	65	17	31	50	10,82	9,27	14,50	2
5	15	56	11	23	56	13,93	6,18	20,22	2
6	18	58	15	28	55	10,91	4,14	16,99	2
7	17	58	13	24	50	7,42	3,94	20,19	2
8	16	70	18	30	60	20,85	13,84	4,03	3
9	20	68	12	23	65	22,97	14,99	10,96	3
10	15	76	16	34	64	28,48	21,55	6,73	3
11	17	80	20	26	68	32,73	25,81	10,17	3
12	15	80	18	27	50	24,02	21,03	13,06	3
13	15	62	13	20	55	14,16	6,24	16,35	2
14	15	74	21	30	60	24,23	18,05	4,31	3
15	16	67	16	31	61	20,10	12,32	7,08	3

Dengan dilakukannya perhitungan dengan menerapkan rumus *clustering*, selanjutnya pengelompokkan berdasarkan jarak minimum ke *Centroid* yang paling dekat, yakni:

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

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

Dengan dilakukannya perhitungan dengan menerapkan rumus *clustering*, selanjutnya pada iterasi 3 dan iterasi 4 tidak ada lagi perubahan pada *clusternya* maka perhitungan iterasi diberhentikan.

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

a. Import Data

Mengimport data merupakan tahapan pertama dalam proses Analisa data dengan menerapkan bahasa pemrograman *Python*. Dalam proses mengimport data library yang digunakan adalah *Pandas*. Pastikan *Pandas* sudah terinstal pada *Jupyter Notebook*. Lalu mengimport perpustakaan dengan menulis baris kode yang sesuai dengan *Jupyter Notebook* agar dapat diakses dengan benar oleh *Jupyter Notebook*. Ketika import data selesai, data dapat di proses. Pada tahap import data dataset atlet taekwondo ini berformat CSV yang berjumlah 500 data terdiri dari 6 atribut yaitu nama, umur, berat badan, kuda-kuda, tendangan dan pukulan.

b. Transformasi Data

Tahapan transformasi data ini ialah tahapan data yang sesuai dengan kebutuhan pada proses pemodelan data. Transformasi data ialah proses perubahan data menjadi data numerik, klasifikasi, dan lainnya. Hasil data transformasi dapat dilihat dibawah ini.

No.	Nama	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan
1	DWI RAHMAD SIDDIQ	22.0	63	16	27	47
2	INDAH SARAH SINAGA	22.0	50	10	20	40
3	M WAHYU RIZKI KESUMA DAHARO	21.0	60	20	30	48
4	MUHAMAAD ROFIQ SIMATUPANG	19.0	65	17	31	50
5	NABILA ZAHIRA BILQIS	15.0	56	11	23	56
...
496	DYO SANDYKA	17.0	83	17	28	67
497	YAZID ARIFIN	20.0	82	15	26	66
498	AHMAD NABIL	15.0	69	23	27	71
499	MARCIA TIURMA BERTHA HUTAPEA	15.0	49	14	23	43
500	MUHRIM EFENDI RAMBE	23.0	80	15	29	65

500 rows × 6 columns

Gambar 3. Dataset Awal Atlet Taekwondo

c. Seleksi Data

Tahapan seleksi data ialah proses pemilihan data dan persiapan data yang akan digunakan sesuai dengan tujuan analisis datanya. Pada tahapan seleksi data pilih variabel yang digunakan dalam proses *clustering*. Hasil atribut setelah melakukan seleksi data ditunjukkan pada gambar dibawah ini.

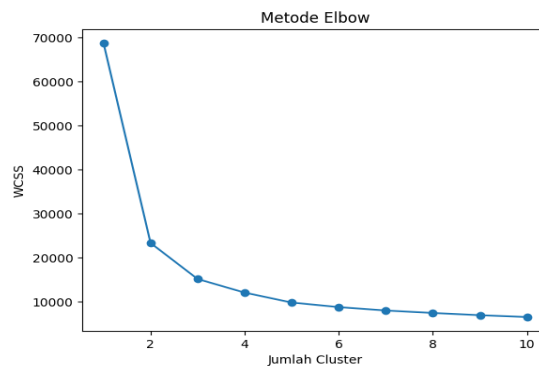
No	Umur	Berat Badan	Kuda-Kuda	Tendangan	Pukulan	
0	1	22	63	16	27	47
1	2	22	50	10	20	40
2	3	21	60	20	30	48
3	4	19	65	17	31	50
4	5	15	56	11	23	56
...
481	482	17	83	17	28	67
482	483	20	82	15	26	66
483	484	15	69	23	27	71
484	485	15	49	14	23	43
485	486	23	80	15	29	65

486 rows × 6 columns

Gambar 4. Hasil Seleksi Data

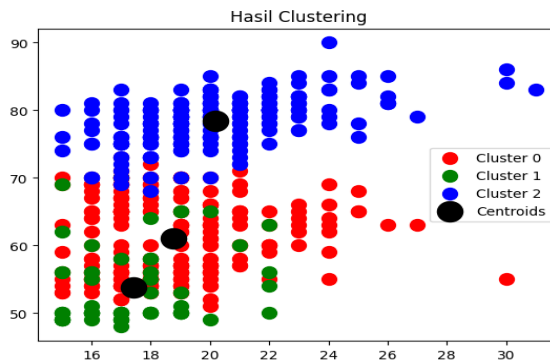
d. Proses Clustering

Proses *Clustering* dilakukan dengan menerapkan *Jupyter Notebook* dengan bahasa pemrograman *Python 3*. Proses *clustering* bermula dengan mencari K optimal menerapkan *Metode Elbow* dengan random state : 0.



Gambar 5. Grafik K Optimal

Berdasarkan output grafik di atas dapat dilihat jumlah *cluster* optimal untuk metode *K-Means* dengan menggunakan evaluasi *Metode Elbow* adalah 3 *Cluster*. Hal ini dikarenakan pada jumlah *cluster* 3, terdapat indikasi lekukan pada grafik di atas. 3 *cluster* yakni memiliki warna biru, hijau dan merah, yng dapat dilihat pada gambar 6.



Gambar 6. Hasil Clustering

Terlihat bahwa hasil clustering adalah *Centroid* atau titik pusat yang terlihat pada gambar disertai jumlah anggota setiap *cluster* nya terlihat pada gambar berikut :

```
[[18.77934272 61.05164319 17.81220657 30.15962441 67.20657277]
 [17.40350877 53.75438596 13.9122807 24.42105263 45.24561404]
 [20.18981481 78.47222222 17.68518519 28.14814815 60.60648148]]
```

Gambar 7. Centroid Cluster

Cluster	
0	213
1	57
2	216

Gambar 8. Jumlah Atlet Setiap Cluster

Berdasarkan perhitungan dengan menerapkan *Jupyter Notebook* di atas, dapat dilihat bahwa:

- Cluster 0* kategori layak dengan *Centroid* 18,77934272, 61,05164319, 17,81220657, 30,15962441, 67,20657177 dengan jumlah atlet paling sedikit berjumlah 213 atlet. Dapat diketahui bahwa *cluster 0* karakteristiknya umur >18 tahun, berat badan >61 kg, kuda-kuda 17 menit, tendangan 30 kali dan pukulan 67 kali.
- Cluster 1* kategori tidak layak dengan *Centroid* 17,40350877, 53,75438596, 13,9122807, 24,42105263, 45,24561404 yang berjumlah 57 atlet. Dapat diketahui bahwa *cluster 1* karakteristiknya umur <17 tahun, berat badan <53kg, kuda-kuda <13 menit, tendangan <24 kali dan pukulan <45 kali.
- Cluster 2* kategori sangat layak dengan *Centroid* 20,18981481, 78,47222222, 17,68518519, 28,14814815, 60,60648148 dengan jumlah atlet terbanyak berjumlah 216 atlet. Dapat diketahui bahwa *cluster 2* karakteristiknya umur >20 tahun, berat badan >78 kg, kuda-kuda >17 menit, tendangan >28 kali dan pukulan >60 kali.

4. KESIMPULAN

Sesuai dengan hasil dari dilakukannya penelitian ini, penerapan metode *K-Means Clustering* Untuk Seleksi Atlet Taekwondo Porprov Pada Koni Sumut berhasil diterapkan. Maka kesimpulan dari penelitian ini adalah penggunaan metode *K-Means Clustering* untuk seleksi atlet taekwondo porprov berhasil menghasilkan tiga kelompok (*cluster*) berdasarkan karakteristik yang ada pada data tersebut. *Cluster 0* kategori layak dengan jumlah atlet paling sedikit berjumlah 213 atlet. Dapat diketahui bahwa *cluster 0* karakteristiknya umur >18 tahun, berat badan >61 kg, kuda-kuda 17 menit, tendangan 30 kali dan pukulan 67 kali. *Cluster 1* kategori tidak layak yang berjumlah 57 atlet. Dapat diketahui bahwa *cluster 1* karakteristiknya umur <17 tahun, berat badan <53kg, kuda-kuda <13 menit, tendangan <24 kali dan pukulan <45 kali. *Cluster 2* kategori sangat layak dengan jumlah atlet terbanyak berjumlah 216 atlet. Dapat diketahui bahwa *cluster 2* karakteristiknya umur >20 tahun, berat badan >78 kg, kuda-kuda >17 menit, tendangan >28 kali dan pukulan >60 kali.

REFERENCES

- [1] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode *K-Means Clustering* Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- [2] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.,* vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [3] W. I. Rahayu, S. F. Pane, and ..., "Implementasi Data Mining Dengan Metode *K-Means Clustering* Untuk Menentukan Iklan Audio Berdasarkan User Behaviors Pada Aplikasi Audio Social Media Svara ..., " *J. Tek. ...,* vol. 10, no. 2, pp. 13–19, 2018, [Online]. Available: <https://ejournal.poltekpos.ac.id/index.php/informatika/article/view/437%0Ahttps://ejournal.poltekpos.ac.id/index.php/informatika/article/download/437/278>

- [4] A. T. R. Saragih, A. S. Sembiring, and M. Sayuthi, "Penerapan Metode *Clustering K-Means* untuk Proses Seleksi Calon Peserta Lomba MTQ," *Pelita Inform.*, vol. 17, no. April, pp. 117–122, 2018, [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/download/776/704>
- [5] V. Miralda, M. Zarlis, and E. Irawan, "Penerapan Metode *K-Means Clustering* Untuk Daging Ayam Buras," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 91–98, 2020, doi: 10.47065/bits.v2i2.493.
- [6] Y. F. S. Y. Damanik, S. Sumarno, I. Gunawan, D. Hartama, and I. O. Kirana, "Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma *K-Means*," *J. Ilmu Komput. dan Inform.*, vol. 1, no. 2, pp. 109–132, 2021, doi: 10.54082/jiki.13.
- [7] M. Iqbal, "Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode *K-Means Clustering*," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 97–104, 2019, doi: 10.33330/jurteksi.v5i2.352.
- [8] N. Damanik and M. Sigi, "Penerapan Data Mining Menggunakan Algoritma *K-Means Clustering* Pada Penerimaan Mahasiswa Baru Sebagai Metode Promosi," *J. Tek. Inform. Komput. Univers.*, vol. 4, no. 2, p. 158, 2021.
- [9] M. R. Nahjan, N. Heryana, and A. Voutama, "Implementasi Rapidminer Dengan Metode *Clustering K-Means* Untuk Analisa Penjualan Pada Toko Oj Cell," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 1–4, 2023.
- [10] Y. Andini *et al.*, "Penerapan Data Mining Terhadap Tata Letak Buku," vol. XI, no. 1, pp. 9–15, 2022.
- [11] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan *K-Means*," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.
- [12] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma *K-Means Clustering*," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019, doi: 10.34012/jusikom.v2i2.429.
- [13] T. M. Dista and F. F. Abdulloh, "*Clustering* Pengunjung Mall Menggunakan Metode *K-Means* dan Particle Swarm Optimization," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1339, 2022, doi: 10.30865/mib.v6i3.4172.
- [14] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan *K-Means Cluster* Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [15] S. A. Rahmah and J. Antares, "Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan *K-Means Clustering*," *INFORMATIKA*, vol. 13, no. 2, p. 25, 2022, doi: 10.36723/juri.v13i2.282.
- [16] U. Hasanah, I. Cholissodin, and M. A. Fauzi, "Penentuan Seleksi Atlet Taekwondo Menggunakan Algoritme Support Vector Machine (SVM)," ... *Teknol. Inf. dan Ilmu ...*, vol. 3, no. 4, pp. 3580–3588, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [17] C. A. Malasari, "Pengaruh Latihan Shuttle-Run dan Zig-Zag Run terhadap Kelincahan Atlet Taekwondo," *Gelangg. Olahraga J. Pendidik. Jasm. dan Olahraga*, vol. 3, no. 1, pp. 81–88, 2019, doi: 10.31539/jpjo.v3i1.828.
- [18] J. Usman and Argantos, "Jurnal Performa Olahraga," *J. Performa Olahraga*, vol. 5, no. 1, pp. 18–25, 2020.
- [19] I. N. M. Adiputra, "*Clustering* Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma *K-Means*," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, p. 99, 2022, doi: 10.23887/insert.v2i2.41673.
- [20] S. Agustian and S. Ramadhani, "Jurnal Computer Science and Information Technology (CoSciTech) menggunakan algoritma lexrank," vol. 3, no. 3, pp. 371–381, 2022.