

ANALISIS CLUSTERING PRESTASI ATLET PADA BERBAGAI CABANG OLAHRAGA MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS: DINAS PEMUDA DAN OLAHRAGA KABUPATEN CIREBON)

Puput Putriana, Nana Suarna, Willy Prihartono

Teknik Informatika, STMIK IKMI Cirebon

Jln Perjuangan No 10 B Karyamulya Kesambi Kota Cirebon Jawa Barat Indonesia

puputputriana2002@gmail.com

ABSTRAK

Atlet, sebagai individu yang berkomitmen tinggi di berbagai cabang olahraga, mencari pencapaian tertinggi dalam bidangnya. Namun, relevansi dan generalisasi hasil analisis prestasi atlet terhadap seluruh populasi atlet Kabupaten Cirebon belum sepenuhnya dipahami. Penelitian ini fokus pada analisis *Clustering* prestasi atlet menggunakan algoritma *K-Means*. Metode ini digunakan untuk mengelompokkan atlet berprestasi di Kabupaten Cirebon berdasarkan atribut seperti Umur, Prestasi, Jenis Kelamin, dan cabang olahraga. Pra-pemrosesan data melibatkan pembersihan dan penanganan data yang tidak valid atau hilang. Hasil analisis *Clustering* memberikan wawasan tentang pola prestasi, mendukung pengembangan bakat olahraga, dan perencanaan pelatihan yang efektif. Tujuan penelitian ini adalah memberikan wawasan yang lebih mendalam tentang pola prestasi atlet, memfasilitasi pengembangan bakat olahraga, dan merancang program pelatihan yang lebih efektif. Dengan mengidentifikasi klaster atlet berdasarkan karakteristik tertentu, penelitian ini dapat memberikan dasar yang kuat untuk pengambilan keputusan terkait pengembangan olahraga di Kabupaten Cirebon. Dampak tambahan melibatkan pengembangan strategi yang lebih efektif untuk mendukung para atlet dan meningkatkan prestasi olahraga secara keseluruhan di wilayah tersebut.

Kata kunci : analisis clustering, atlet, cabang olahraga, prestasi olahraga, pengembangan bakat olahraga

1. PENDAHULUAN

Atlet seringkali terlibat dalam latihan intensif sepanjang tahun atau memilih fokus pada satu cabang olahraga sejak usia muda. Meskipun pendekatan ini dianggap dapat mengoptimalkan pengembangan keterampilan olahraga, banyak literatur menyatakan bahwa spesialisasi olahraga dini dapat meningkatkan risiko cedera mental dan fisik tanpa memberikan manfaat yang diharapkan. Penelitian ini menggunakan metode deskriptif untuk membahas isu spesialisasi olahraga dini, merinci prinsip aktivitas fisik yang sesuai dengan perkembangan, dan menyediakan panduan serta sumber daya bagi praktisi dan keluarga atlet. Meski terdapat berbagai dorongan untuk mencapai prestasi dan komitmen dalam satu cabang olahraga, keputusan tersebut perlu dipertimbangkan dengan hati-hati karena dapat menimbulkan dampak merugikan [1].

Penelitian ini mendalami permasalahan *Clustering* atlet berprestasi di Kabupaten Cirebon dengan tujuan memahami dan mengklasifikasikan potensi atletik di daerah ini. Keberhasilan pengembangan olahraga di Kabupaten Cirebon sangat dipengaruhi oleh kurangnya informasi terkait klasifikasi atlet berprestasi. Data atlet dari berbagai cabang olahraga memperlihatkan tingkat prestasi yang beragam, sehingga analisis *Clustering* perlu memperhitungkan skala yang sesuai. Atribut yang digunakan untuk mengukur prestasi atlet dapat mencakup jumlah medali, skor rata-rata, atau parameter olahraga tertentu. Tantangan dalam revolusi industri, khususnya *Big Data*, menjadi aspek penting

dalam penelitian ini karena data olahraga cenderung memiliki volume besar dan kompleksitas tinggi yang memerlukan pendekatan pemrosesan khusus.

Penelitian sebelumnya oleh Fiqri Azzahra berjudul "Pengelompokan Data Atlet Pusat Pendidikan Latihan Pelajar Di Jawa Barat Menggunakan Algoritma *K-Means*" telah menggunakan metode *Clustering* untuk menganalisis data atlet di Jawa Barat dan menemukan bahwa faktor-faktor keberhasilan atlet memengaruhi prestasi individu mereka. Studi ini melanjutkan penggunaan metode *Clustering* pada data atlet dari berbagai cabang olahraga, menunjukkan bahwa pendekatan ini dapat membantu mengidentifikasi kelompok atlet dengan karakteristik serupa. Temuan ini membuktikan bahwa *Clustering* efektif dalam mengelompokkan atlet berdasarkan prestasi olahraga dan karakteristik fisik [2].

Teknik *data mining* diharapkan memberikan kontribusi dalam pengolahan data atlet berdasarkan usia, dengan tujuan mendapatkan wawasan baru terkait pola pengelolaan data tersebut. Penelitian ini akan fokus pada data Atlet Berprestasi di Kabupaten Cirebon yang diperoleh dari Dinas Pemuda dan Olahraga setempat. Analisis akan dilakukan terkait pola pengelompokan data atlet, dengan memperhatikan atribut seperti Nama, jenis kelamin, usia, dan prestasi. Metode *Clustering* menggunakan algoritma *K-Means* akan digunakan untuk mengidentifikasi kelompok data atlet berdasarkan atribut yang dipertimbangkan. Proses *Clustering* bertujuan untuk mengelompokkan data sehingga

kesamaan di dalam kelompok tinggi, sedangkan perbedaan antar kelompok rendah [2].

Proses *Clustering* dengan algoritma *K-Means* membagi objek data menjadi kelompok atau *Cluster* berdasarkan kesamaan. Algoritma ini, termasuk dalam *unsupervised learning*, efektif dalam mengklasifikasikan data ke dalam beberapa kelompok melalui metode partisi [3].

Penelitian ini bertujuan mengembangkan solusi berbasis teknologi informasi untuk memahami karakteristik dan potensi atlet di Kabupaten Cirebon. Dengan menerapkan algoritma *K-Means* pada data atlet dari berbagai cabang olahraga, penelitian ini akan menghasilkan *Clustering* yang informatif. Signifikansi penelitian ini terletak pada kontribusinya terhadap pengembangan olahraga dan bakat atlet. Analisis data ini memberikan wawasan berharga bagi Dispora Kabupaten Cirebon dalam perencanaan program pelatihan yang lebih efektif. Metode utama melibatkan penerapan algoritma *K-Means Clustering* untuk mengelompokkan atlet berdasarkan prestasi olahraga dan karakteristik demografis. Evaluasi hasil *Clustering* menggunakan metrik relevan akan memberikan pemahaman yang lebih dalam tentang atlet, mendukung pengembangan olahraga di wilayah tersebut.

Analisis *Clustering* menggunakan algoritma *K-Means* memberikan wawasan mendalam tentang pola prestasi olahraga dan karakteristik atlet. Temuan ini dapat menjadi landasan bagi Dispora Kabupaten Cirebon untuk merancang program pelatihan yang lebih terfokus. Selain itu, penelitian ini berpotensi menjadi kontribusi berharga dalam bidang Informatika dan pengolahan data olahraga. Implikasinya mencakup peluang untuk penelitian lebih lanjut tentang pemanfaatan teknologi informasi dalam pengembangan bakat olahraga serta referensi bagi studi komparatif di berbagai wilayah atau cabang olahraga. Penggunaan teknologi informasi dan analisis data diharapkan dapat meningkatkan efisiensi dan efektivitas program olahraga di Kabupaten Cirebon, menciptakan atlet berkualitas dengan potensi prestasi tinggi, dan berkontribusi pada pengembangan olahraga secara keseluruhan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sejumlah penelitian dalam bidang olahraga telah dilakukan untuk menggali berbagai aspek yang memengaruhi prestasi atlet. [4] dalam penelitiannya yang berjudul “Analisis Tingkat Motivasi Berprestasi Atlet Futsal di Akademi SGF Kota Bengkulu” menganalisis tingkat motivasi berprestasi atlet futsal di Akademi SGF Kota Bengkulu menyoroti permasalahan di akademi tersebut, seperti kurangnya target prestasi yang jelas, rendahnya tanggung jawab atlet terhadap program latihan, dan kurangnya disiplin.

Penelitian lainnya yakni “Analisis Tingkat Kepercayaan Diri Saat Bertanding Atlet Pencak Silat Perguruan Satria Sejati” [1] berfokus pada analisis

tingkat kepercayaan diri atlet pencak silat di Perguruan Datria Sejati, mengaplikasikan metode deskriptif kuantitatif untuk menghasilkan data deskriptif berupa angka hasil dari responden. Penelitian [5] menggambarkan kesehatan mental seorang atlet sepakbola, menekankan pentingnya pemahaman psikologi atlet sebelum dan selama pertandingan.

“*Clustering* Data Menggunakan Metode *K-Means* untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler” [6] membahas *clustering* data menggunakan metode *K-Means* untuk merekomendasikan pembelajaran akademik bagi siswa aktif dalam ekstrakurikuler. Penelitian lainnya, “Analisis Kesiapan Mental Pada Atlet Bola Tangan Kota Surabaya Menjelang Porprov Jawa Timur 2022” [7] menganalisis kesiapan mental atlet bola tangan Kota Surabaya menjelang Propov Jawa Timur 2022 dengan pendekatan deskriptif kualitatif.

[8] membuat penelitian “Peran Orang Tua Dalam Mendukung Prestasi Olahraga Atlet Karate Lebah Bukit Wira Samarinda” yang mengeksplorasi peran orang tua dalam mendukung prestasi atlet karate Lebah Bukit Wira Samarinda melalui penelitian deskriptif kuantitatif dengan metode survey. Penelitian “Analisis Antropometri dalam Mengidentifikasi Bibit Atlet Berbakat Cabang Olahraga Sepatu Roda” [9] menggunakan analisis antropometri untuk mengidentifikasi bibit atlet berbakat cabang olahraga sepatu roda.

2.2. Atlet

Atlet atau disebut dengan olahragawan merupakan seseorang yang pandai atau mahir dalam suatu cabang olahraga yang dikuasai dan memerlukan kekuatan, ketangkasan dan kecepatan. Pencarian bibit atlet atau *athlete talent identification* merupakan suatu kegiatan yang bertujuan untuk mencari potensi atlet berdasarkan bakat yang dimilikinya dalam suatu cabang olahraga, dan dilakukan sedini mungkin, guna dapat mengembangkan potensi tersebut secara maksimal [9].

2.3. Data Mining

Data mining adalah teknik untuk menemukan pola, tren, dan korelasi dalam dataset besar dengan menggunakan pengenalan pola, statistik, dan matematika. Tujuannya adalah mengidentifikasi pola yang dapat mendukung pengambilan keputusan di masa depan [10].

2.4. Clustering

Clustering adalah proses pengelompokan data ke dalam dua atau lebih *Cluster* sehingga data dalam satu *Cluster* lebih serupa satu sama lain daripada dengan data di *Cluster* lainnya. Metode ini merupakan bagian dari pembelajaran tanpa pengawasan (*unsupervised learning*) dan tidak memerlukan label pada setiap *Cluster* [6]. Dengan kata lain, *Clustering* adalah pengelompokan menggunakan teknik *unsupervised*

learning tanpa fase pembelajaran dan tanpa penggunaan label pada setiap *Cluster*. Metode *Clustering* merupakan suatu pendekatan untuk mengelompokkan data ke dalam kelas atau *Cluster* berdasarkan kemiripan atribut di antara data dalam *Cluster* tersebut.

2.5. K-Means

Algoritma *K-Means* adalah metode klasterisasi yang mengelompokkan data berdasarkan *centroid* terdekat. Tujuannya adalah memaksimalkan kesamaan dalam satu *Cluster* dan meminimalkan kesamaan antar *Cluster*. Prosesnya melibatkan penentuan jumlah klaster, inisialisasi *centroid*, pengklasifikasian data, dan pembaruan nilai *centroid* hingga konvergensi [11].

3. METODE PENELITIAN

3.1. Metode Penelitian

Dalam penelitian ini, digunakan metode kuantitatif dengan algoritma *K-Means Clustering*

untuk mengelompokkan data atlet berprestasi dari berbagai cabang secara kuantitatif. Pendekatan ini memungkinkan identifikasi pola dan hubungan antar data numerik, sehingga memberikan hasil yang lebih terukur dan statistik terkait dengan kluster atlet berprestasi.



Gambar 1. Metode Penelitian

Gambar 1 merupakan visualisasi metode penelitian yang dimulai dari studi permasalahan, pengumpulan data, pemilihan data, analisis data mining, interpretasi dan kesimpulan, serta validasi dan reliabilitas.

Tabel 1. Metode penelitian

No	Tahapan	Kegiatan/ Aktifitas	Capaian Indikator Tahapan
1.	Studi Permasalahan	Proses yang melibatkan analisis mendalam terhadap aktivitas atau program tertentu untuk mengevaluasi sejauh mana tujuan telah tercapai dan untuk mengidentifikasi masalah atau hambatan yang mungkin timbul dalam mencapai tujuan tersebut.	Tercapainya data yang akan dijadikan sampel untuk menghasilkan informasi.
	Pengumpulan Data	Pengambilan data dari lembaga terkait, atau penggunaan <i>dataset</i> yang sudah tersedia.	Sumber Data Identifikasi sumber data yang relevan, seperti data atlet data demografi, dan cabang olahraga.
	Pemilihan Data	Pembersihan Data	Melakukan pembersihan data untuk mengatasi data yang tidak lengkap, data yang hilang, atau data yang tidak valid.
		Integrasi Data	Mengintegrasikan data dari berbagai sumber yang relevan untuk analisis. Transformasi Data: Melakukan transformasi data jika diperlukan, seperti normalisasi atau standarisasi atribut.
	Analisis Data Mining	<i>Selection</i>	Pemilihan data (<i>data target</i>) yang akan menjadi sampel untuk proses selanjutnya.
2.		<i>Pre-processing</i>	Melakukan serangkaian proses untuk melengkapi data dan menjaga konsistensi data.
3.		<i>Transformation</i>	Mengubah representasi data untuk mempermudah dan memperbaiki agar sesuai dengan Teknik data mining yang akan dipergunakan
4.	Interpretasi dan Kesimpulan	Interpretasi hasil	Menginterpretasikan hasil analisis data mining untuk memahami faktor-faktor yang berkontribusi terhadap tingkat prestasi dan mengevaluasi pengaruh variabel-variabel yang diteliti.
5.	Validitas dan Reliabilitas	Validitas: Reliabilitas:	Memastikan keabsahan dan validitas data yang digunakan serta metode analisis yang digunakan dalam penelitian. Memastikan konsistensi dan keandalan hasil penelitian dengan mengadopsi metode yang tepat dan melakukan pengujian ulang (<i>retest</i>) jika diperlukan.

Tabel 1 merupakan penjelasan detail mengenai kegiatan dari setiap tahapan-tahapan metode penelitian beserta capaian indikatornya.

3.2. Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Dinas Pemuda dan Olahraga Kabupaten Cirebon. Keberlanjutan penggunaan data dari instansi tersebut menjamin realibilitas dan validitas yang dapat dipertanggungjawabkan. Data mencakup atribut seperti tanggal pembuatan, nama peserta, usia, cabang olahraga, dan prestasi. Setelah pengumpulan data, dilakukan analisis dan langkah-langkah pemrosesan data sesuai dengan Teknik Analisis Data KDD untuk memastikan kaidah pengolahan yang tepat.

3.3. Populasi dan Sampel

Penelitian ini berfokus pada data Atlet berprestasi di Kabupaten Cirebon dari Dinas Pemuda dan Olahraga selama 5 tahun terakhir (2018-2022), dengan jumlah populasi sebanyak 398 data. Sampel awal terdiri dari 7 atribut, termasuk tanggal pembuatan, nama peserta, jenis kelamin, umur, cabang olahraga, dan prestasi. Namun, setelah diolah, hanya 4 atribut yang dipertahankan, yaitu umur, jenis kelamin, cabang olahraga, dan prestasi. Sumber data penelitian berasal dari Dinas Pemuda Olahraga Kabupaten Cirebon.

3.4. Teknik Pengumpulan Data

Data dikumpulkan menggunakan teknik pengumpulan data sekunder yang bersumber dari Dinas Pemuda dan Olahraga Kabupaten Cirebon. Pengambilan data dilakukan secara langsung dengan kerjasama federasi olahraga untuk mendapatkan informasi resmi dan terkini mengenai prestasi atlet dalam berbagai kompetisi. Proses verifikasi dan validasi diimplementasikan untuk memastikan keakuratan dan keabsahan informasi.

3.5. Teknik Analisis Data

Pada penelitian ini, menggunakan metode *Knowledge Discovery in Database (KDD)*. Berikut merupakan tahapan-tahapan dan proses yang ada pada *Knowledge Discovery in Database*:

a. Selection

Tahapan awal adalah *selection*, di mana data relevan dipilih dari *database* Atlet Berprestasi di Kabupaten Cirebon tahun 2018-2022 sebanyak 391 data, mengikuti prinsip bahwa tidak semua data diperlukan dalam proses *data mining*.

b. Pre-processing

Langkah selanjutnya adalah *pre-processing*, di mana data yang telah dikumpulkan dibersihkan dari kesalahan atau ketidaklengkapan, termasuk penanganan data yang kosong, ganda, atau hilang untuk memastikan kesesuaian dengan format yang diinginkan.

c. Transformation

Transformasi dilakukan untuk mengubah representasi data, mempermudah, dan memperbaiki agar sesuai dengan teknik *data mining* yang akan digunakan.

d. Data mining

Tahap utama dari proses *data mining*, di mana jumlah *Cluster* harus ditentukan sebelumnya untuk mencapai nilai *k* yang optimal. Proses ini menggunakan *RapidMiner* sebagai alat, dengan langkah-langkah, antara lain, menentukan jumlah *Cluster*, menetapkan nilai *centroid* atau pusat *Cluster*, dan mengevaluasi jarak antara setiap titik data dan *centroid* terdekat menggunakan model *euclidean* (Muningsih E, 2021).

$$dij = \frac{\sqrt{(x1i - x1j)^2 + (x2i - x2j)^2 + \dots + (xki - xkj)^2}}{(1)}$$

Keterangan:

Dij = Jarak data ke – i ke titik pusat (*Cluster*j).

Xki = Data ke -i *attribut* data ke -k.

Xkj = Data ke -j *attribut* data ke -j.

e. Interpretation/evaluation

Pada tahap *Interpretation / Evaluation*, data yang sudah diproses akan diukur menggunakan operator kinerja untuk menghitung *Davies Bouldin Index (DBI)*, dengan tujuan untuk memperoleh nilai *k* yang optimal.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Tahapan yang pertama yaitu *data selection*. Pada tahap ini memasukkan operator *Read Excel* yang berfungsi untuk membaca data *excel* Prestasi Atlet di Kabupaten Cirebon.



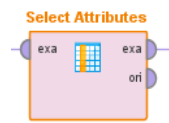
Gambar 2. Operator Read Excel

Gambar 2 di atas menunjukkan operator *Read Excel* yang digunakan untuk mengimpor data. Peneliti menggunakan parameter *default*.

	A	B	C	D	E	F	G	H	I
1	No	Tgl Dibuat	Nama	Umur	Jenis Kel.	Cabor	Prestasi		
2	1.000		Chelly Ca.	21 Tahun	Perempu.	BILLIARD	MEDALI E.		
3	2.000		Rahmat	59 Tahun	Laki-Laki	BILLIARD	MEDALI E.		
4	3.000		Yoniasari	27 Tahun	Perempu.	BILLIARD	MEDALI E.		
5	4.000		Ahmad R.	20 Tahun	Laki-Laki	VOLI	MEDALI P.		
6	5.000		AHMAD Y.	20 Tahun	Laki-Laki	MENEMB.	MEDALI P.		
7	6.000		Audias An.	20 Tahun	Perempu.	JU-JITSU	MEDALI P.		
8	7.000		Cherita	21 Tahun	Perempu.	JU-JITSU	MEDALI P.		
9	8.000		Desri	22 Tahun	Laki-Laki	RUGBY	MEDALI P.		
10	9.000		Dwiti Sor.	15 Tahun	Laki-Laki	SENAH	MEDALI P.		
11	10.000		EKA AND.	17 Tahun	Laki-Laki	Persabua.	MEDALI P.		
12	11.000		EKO RID.	22 Tahun	Laki-Laki	SENAH	MEDALI P.		
13	12.000		Elma Ry.	22 Tahun	Perempu.	RUGBY	MEDALI P.		

Gambar 3. Output data Read Excel

Gambar 3 merupakan tampilan *output* berupa *preview data* yang telah diimpor.



Gambar 4. *Select attributes*

Tahap selanjutnya yaitu memasukkan operator *Select Attributes* yang berfungsi untuk memilih atau menyeleksi atribut mana yang akan digunakan untuk dianalisa yang tertera pada Gambar 4. Parameter pada operator *Select Attributes* hanya mengganti *attribute filter type a subset*.

Tabel 2. Hasil operator *Select Attributes*

Atribut tidak yang digunakan	Atribut Yang di gunakan
Id	Umur
Tanggal dibuat	Jenis Kelamin
Nama	Cabor
	Prestasi

Tabel 2 di atas merupakan hasil penggunaan operator *Select Attributes*. Dari total 7 atribut (id, Tanggal Dibuat, Nama Peserta, Umur, Jenis Kelamin, Cabang Olahraga, dan Prestasi), dipilih 4 atribut (Umur, Jenis Kelamin, Cabang Olahraga, dan Prestasi) berdasarkan relevansi informasi terkait dengan analisis yang diinginkan. Umur dan jenis kelamin dianggap faktor demografis penting, sementara cabang olahraga dan prestasi memberikan informasi spesifik terkait partisipasi dan pencapaian peserta dalam kegiatan tertentu. Kombinasi keempat atribut ini diharapkan dapat memberikan wawasan mendalam tentang profil peserta dan kinerja mereka dalam konteks tertentu.



Gambar 5. Model proses langkah *Selection* di *RapidMiner*

Gambar 5 menggambarkan model proses pada fase *Selection* dalam platform *RapidMiner*. Pada tahap ini, proses seleksi data dan fitur dilakukan untuk menyaring variabel yang paling relevan dan signifikan dalam analisis data.

4.2. Data Pre-Processing/Cleaning

Pada tahap *Pre-Processing/Cleaning*, langkah-langkah dilakukan untuk membersihkan dan menyiapkan data sebelum analisis. Tujuannya adalah memastikan data memiliki kualitas baik, akurat, dan sesuai dengan kebutuhan analisis. Proses ini melibatkan penanganan data kosong, penghapusan duplikat, serta penanganan nilai yang hilang agar sesuai dengan format yang diinginkan. Dengan menggunakan operator *Replace Missing Values*, kita dapat mengatasi masalah data yang tidak lengkap

dengan menggantikan nilai hilang menggunakan metode seperti rata-rata, median, atau modus. Hal ini memastikan data yang digunakan untuk analisis atau pemodelan tidak memiliki nilai yang hilang, sehingga memungkinkan hasil yang lebih tepat dan akurat.



Gambar 6. *Replace Missing Values*

Gambar 6 menunjukkan operator *Replace Missing Values* dengan parameter *default*. Setelah dilakukan pembersihan data oleh operator tersebut, hasilnya tidak menunjukkan perubahan data; jumlahnya tetap 389 data dari tahun 2018 sampai tahun 2022. Hal ini disebabkan oleh penggunaan data yang sudah lengkap, tanpa adanya data yang kosong, ganda, atau hilang.

4.3. Data Transformation

Dalam tahap transformasi data, atribut terpilih disesuaikan untuk digunakan dalam algoritma *K-Means*. Menggunakan operator *Nominal to Numerical*, atribut non-numerik diubah menjadi tipe numerik, memungkinkan algoritme *K-Means* untuk memproses data tersebut. Proses transformasi ini esensial karena *K-Means* hanya dapat bekerja dengan data numerik. Operator tersebut memfasilitasi pengubahan atribut non-numerik menjadi representasi numerik, seperti pengkodean *one-hot* atau konversi label ke nilai numerik sesuai skema yang ditentukan. Transformasi ini memungkinkan analisis klusterisasi pada data yang telah diubah ke dalam format numerik.

4.4. Data Mining

Tahap selanjutnya memasukkan operator *Clustering* yang berfungsi untuk menentukan *Cluster* atau pengelompokan data. Penelitian ini menggunakan algoritme *K-Means* dan menggunakan 4 *Cluster* max perulangan 10. Parameter pada operator *Clustering K-Means* dimana diketahui $K = 4$, $max\ run = 10$ kali putaran, tipe *measure = Numerical Measure* yang *Euclidean Distance* dan $max\ optimal\ steps = 100$. Hasil operator *Clustering K-Means* diperoleh 4 *Cluster* yaitu *Cluster 0* sebanyak 75 *items*, *Cluster 1* sebanyak 15 *items*, *Cluster 2* sebanyak 187 *items*, *Cluster 3* sebanyak 112 *items* dengan total *items* sebanyak 389 *items*.

Cluster Model

```
Cluster 0: 75 items
Cluster 1: 15 items
Cluster 2: 187 items
Cluster 3: 112 items
Total number of items: 389
```

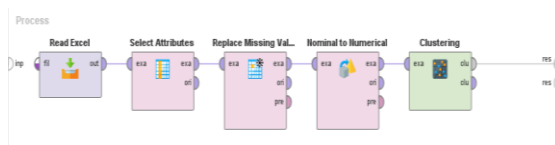
Gambar 7. Cluster model

Gambar 7 menyajikan *output cluster model* hasil dari operator *Clustering K-Means*.

Row No.	id	label	Umur	Jenis kelamin	Cabor	Prestasi
1	1	cluster_0	0	0	0	0
2	2	cluster_0	1	1	0	0
3	3	cluster_0	2	0	0	0
4	4	cluster_0	3	1	1	1
5	5	cluster_0	4	1	2	1
6	6	cluster_0	3	0	3	1
7	7	cluster_0	0	0	3	1
8	8	cluster_0	5	1	4	1
9	9	cluster_0	6	1	5	1
10	10	cluster_0	7	1	6	1
11	11	cluster_0	5	1	5	1
12	12	cluster_0	5	0	4	1
13	13	cluster_0	8	0	5	1

Gambar 8. Hasil operator *clustering* menggunakan *K-Means*

Gambar 8 memvisualisasikan data hasil penerapan operator *clustering* dengan menggunakan algoritma *K-Means*. Pada tahap ini, algoritma *K-Means* digunakan untuk mengelompokkan data menjadi kluster yang memiliki kesamaan karakteristik. Visualisasi ini memberikan gambaran tentang pola kluster yang terbentuk dalam dataset, memungkinkan pemahaman lebih mendalam terkait struktur dan distribusi data setelah proses klusterisasi.



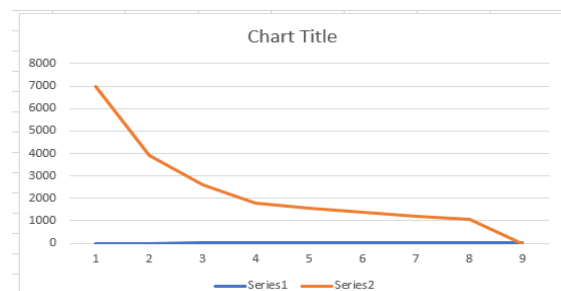
Gambar 9. Model proses langkah *Data Mining* di *RapidMiner*

Gambar 9 memperlihatkan rangkaian model dalam proses *data mining* setelah pengelompokan menggunakan algoritma *clustering K-Means*. Langkah berikutnya adalah melakukan validasi terhadap kluster yang dihasilkan, dengan salah satu metode umum menggunakan performa jarak kluster. Evaluasi ini dilakukan melalui model *performance* dengan menghitung *Davies Bouldin Index* (DBI), sebuah metrik yang mengukur kualitas kluster berdasarkan jarak antar kluster. Operator ini bertujuan untuk mengidentifikasi nilai *k* yang paling optimal dalam klusterisasi. Dengan menguji DBI pada berbagai nilai *k*, operator membantu mengevaluasi performa kluster yang dihasilkan oleh model *K-Means*, memberikan pemahaman tentang seberapa baik pembagian kluster. Nilai DBI mendekati 0 menunjukkan pembagian kluster dengan jarak antar kluster yang optimal, menandakan nilai *k* yang dipilih mungkin lebih baik merepresentasikan struktur data.



Gambar 10. Hasil operator *Performance*

Gambar 10 menampilkan *output* nilai *Davies Bouldin Index* dari penggunaan operator *Performance* dalam penelitian ini. Hasil evaluasi menunjukkan nilai terdekat dengan 0 adalah 0.196, dicapai dengan nilai *k* = 4 dan maksimal 10 perulangan.



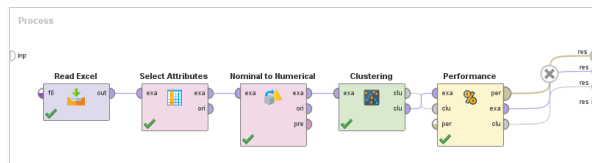
Gambar 11. Metode *Elbow*

Gambar 11 menggambarkan jumlah kluster menggunakan metode *Elbow*. Dari grafik, terlihat bahwa beberapa nilai *K* mengalami penurunan yang signifikan, dan setelahnya penurunan menjadi lebih perlahan hingga mencapai stabilitas. Pada grafik ini, penurunan yang drastis membentuk siku pada titik *K* = 4, menandakan bahwa nilai *K* yang ideal adalah 4.

Tabel 3. Hasil DBI

No	Cluster	Davies Bouldi Index
1	K2	0.209
2	K3	0.220
3	K4	0.196
4	K5	0.230
5	K6	0.229
6	K7	0.254
7	K8	0.282
8	K9	0.305
9	K10	0.268

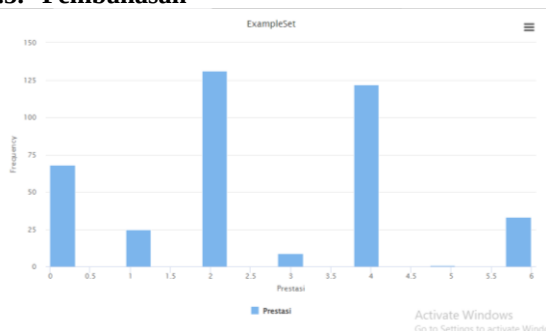
Tabel 3 menampilkan hasil *Davies Bouldin Index* (DBI) dari 4 percobaan untuk menemukan nilai *k* yang optimal. Nilai *k* yang optimal diidentifikasi berdasarkan DBI yang mendekati 0. Dari 4 percobaan tersebut, nilai *k* yang paling optimal adalah *k* = 4 dengan DBI sebesar 0.196.



Gambar 12. Model proses semua langkah di RapidMiner

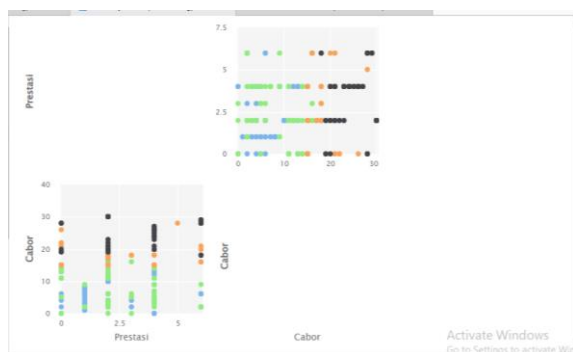
Gambar 12 menampilkan keseluruhan model proses dalam RapidMiner, mencakup langkah-langkah dari pengolahan awal data hingga evaluasi clustering dengan K-Means.

4.5. Pembahasan



Gambar 13. Diagram data prestasi atlet berdasarkan cluster

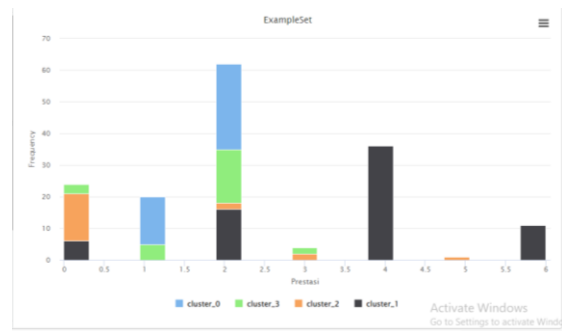
Gambar 13 menggambarkan jumlah prestasi atlet berdasarkan kluster hasil analisis. Hasil penelitian ini memberikan insight untuk menganalisis distribusi jumlah prestasi atlet di Kabupaten Cirebon berdasarkan kluster. Ditemukan bahwa jumlah prestasi tertinggi terdapat pada Cluster 2 dengan total 131, diikuti oleh Cluster 1 dengan jumlah prestasi 122. Cluster 0 menunjukkan jumlah prestasi sedang sebanyak 68, sementara Cluster 3 memiliki jumlah prestasi terendah, yakni 9 atlet.



Gambar 14. Scatter plot

Gambar 14 menunjukkan hasil scatter plot persebaran prestasi atlet di Kabupaten Cirebon dengan metode K-means Clustering, memvisualisasikan prestasi berdasarkan cabang olahraga dan jumlah prestasi. Metode K-means Clustering digunakan untuk mengelompokkan titik-titik prestasi ke dalam beberapa Cluster berdasarkan jarak antar titik. Scatter plot ini menggambarkan pola persebaran prestasi atlet

berdasarkan cabang olahraga, Cluster prestasi, dan jumlah atlet berprestasi di Kabupaten Cirebon.



Gambar 15. Jumlah semua cluster

Gambar 15 menampilkan diagram jumlah seluruh kluster. Cluster 0 memiliki nilai tertinggi sebanyak 62, Cluster 1 dengan nilai tertinggi 36, Cluster 2 dengan nilai tertinggi 21, dan Cluster 3 dengan nilai tertinggi 35.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Umur	19.947	21.133	4.733	12.768
Jenis kelamin	0.483	0.809	0.804	0.079
Prestasi	2.013	2.933	2.481	2.088

Gambar 16. Hasil dari Centroid Table

Gambar 16 memperlihatkan hasil dari Tabel Centroid. Tabel ini menyajikan informasi mengenai posisi pusat kluster (centroid) yang dihasilkan oleh algoritma K-Means. Pemahaman terhadap centroid masing-masing kluster memberikan wawasan tambahan terkait karakteristik dan perbedaan antar kluster dalam analisis prestasi atlet di Kabupaten Cirebon.

Dari hasil pengolahan dataset, berikut adalah ringkasan hasil dari masing-masing cluster:

1) Cluster 0

- Medali Emas: Atlet berusia 17-18 tahun, mayoritas laki-laki, cabang olahraga: panahan, bola basket.
- Medali Perunggu: Rentang usia atlet 18-25 tahun, mayoritas laki-laki, cabang olahraga beragam seperti bola basket, gulat, menembak, sepeda, tarung derajat, taekwondo, judo, renang, takraw, tenis lapangan, senam, dan bulutangkis.

2) Cluster 1

- Medali Perunggu: Atlet berusia 26-33 tahun, mayoritas laki-laki, cabang olahraga: binaraga, angkat berat, panahan, tenis meja, dan tenis lapangan.
- Medali Perak: Rentang usia beragam sekitar 29-36 tahun, mayoritas laki-laki, cabang olahraga: renang, sepeda, dan billiard.

3) Cluster 2

- Medali Perunggu: Atlet berusia 37-49 tahun, mayoritas laki-laki, cabang olahraga: billiard, voli, menembak, ju-jitsu, rugby, senam, gulat, binaraga, taekwondo, tarung derajat, dan tenis lapangan.

- Medali Perak: Rentang usia 37-51 tahun, mayoritas laki-laki, cabang olahraga: wushu, ipsi, futsal, atletik, basket, bola tangan, taekwondo, rugby, takraw, e-sport, gulat.
 - Medali Emas: Mayoritas atlet berusia 37-47 tahun, mayoritas laki-laki, cabang olahraga: futsal, ipsi, atletik, basket, menembak, tarung derajat, panjat tebing, ju-jitsu, renang, petanque, muaythai, voli, gulat, senam, bola tangan, motor.
- 4) Cluster 3
- Medali Perunggu: Mayoritas atlet berusia 9-16 tahun, mayoritas laki-laki, cabang olahraga: binaraga, wushu, voli, menembak, bola tangan, judo, e-sport, takraw, rugby, panjat tebing, tenis lapangan, panahan, tarung derajat, dan angkat berat.
 - Medali Perak: Rentang usia 9-16 tahun, mayoritas laki-laki, cabang olahraga: wushu, tenis meja, ju-jitsu, ipsi, taekwondo, muaythai, petanque, dan renang.
 - Medali Emas: Mayoritas atlet berusia 9-14 tahun, mayoritas laki-laki, cabang olahraga: takraw, basket, senam, atletik, futsal, billiard, tenis meja, voli, panjat tebing, rugby, pencak silat, motor, tarung derajat, judo, muaythai, petanque, tenis lapangan, dan bulutangkis.
5. KESIMPULAN DAN SARAN
- Dari penelitian pengelompokan prestasi atlet berdasarkan cabang olahraga, disimpulkan bahwa penelitian ini optimal pada $k = 4$ (DBI: 0.196). Keempat klaster memiliki karakteristiknya sendiri, dengan medali emas dominan pada atlet pria di atas 37 tahun dalam berbagai cabang olahraga. Peneliti memberikan saran untuk melakukan analisis mendalam terhadap karakteristik klaster, mengeksplorasi faktor-faktor lain yang mungkin memengaruhi prestasi atlet seperti latar belakang dan jenis latihan, serta mengimplementasikan strategi pelatihan yang terfokus berdasarkan karakteristik masing-masing kelompok atlet.
- DAFTAR PUSTAKA
- [1] R. M. Agus and E. B. Fahrizqi, "ANALISIS TINGKAT KEPERCAYAAN DIRI SAAT BERTANDING ATLET PENCAK SILAT PERGURUAN SATRIA SEJATI," *Multilater. J. Pendidik. Jasm. dan Olahraga*, vol. 19, no. 2, 2021.
 - [2] F. A. Sulaeman, "Penerapan Algoritme K-Means Dalam Mengelompokan Atlet PPLP Keolahragaan Di Jawa Barat," *Algor*, vol. 2, 2023.
 - [3] Sartika and Juju, "PENGELOMPOKAN DATA MINING K-MEANS CLUSTERING," *J. Sci. Soc. Res.*, vol. 6, no. 2, pp. 345–349, 2023.
 - [4] Y. E. Nopiyanto, S. Raibowo, A. Prabowo, and B. Insanistyo, "Analisis Tingkat Motivasi Berprestasi Atlet Futsal di Akademi SGF Kota Bengkulu," vol. 5, pp. 138–146, 2023.
 - [5] Donie, A. N. Aida, E. Purnomo, E. Marheni, and Yendrizal, "Gambaran Kesehatan Mental Seorang Atlet Sepakbola untuk Menjadi Atlet Berprestasi," *J. Patriot*, vol. 5, pp. 147–154, 2023.
 - [6] M. A. Al Fauzie, Yuliadi, and J. A. Putra, "Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, 2023.
 - [7] M. D. Aulia and A. A. Hakim, "ANALISIS KESIAPAN MENTAL PADA ATLET BOLA TANGAN KOTA SURABAYA MENJELANG PORPROV JAWA TIMUR 2022," *Indones. J. Kinanthropology*, vol. 1, no. 2, 2021.
 - [8] H. Taufik and N. A. Mahardhika, "PERAN ORANG TUA DALAM MENDUKUNG PRESTASI OLAHRAGA ATLET KARATE LEBAH BUKIT WIRA SAMARINDA," *J. Olahraga ReKat (Rekreasi Masyarakat)*, vol. 2, no. 1, 2023.
 - [9] N. A. Ridwan, M. Y. Rizky, and W. Himawanto, "Analisis Antropometri dalam Mengidentifikasi Bibit Atlet Berbakat Cabang Olahraga Sepatu Roda," *SPRINTER J. Ilmu Olahraga*, vol. 4, no. 2, 2023.
 - [10] D. Jollyta, W. Ramdhan, and M. Zarlis, *Konsep Data Mining Dan Penerapan*. Deepublish, 2020. [Online]. Available: <https://www.google.com/books?hl=id&lr=&id=piMJEAAAQBAJ&oi=fnd&pg=PP1&dq=penelitian+Data+mining&ots=knvK8cWmdy&sig=SzMQdwZVh8QnCTMstrTlkXzFNxk>
 - [11] S. Ramadhani, D. Azzahra, and Z. Tomi, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022.