JURNAL BUFFER INFORMATIKA
Volume 11 Nomor 2, Oktober 2025
p-ISSN : 2527-4856, e-ISSN : 2614-5413
https://journal.fkom.uniku.ac.id/index.php/buffer

Penerapan Metode K-Means *Clustering* untuk Segmentasi Performa Pembalap F1 Season 2024

Mutia Sahira[1)], Adella Salsabila[2)], Shofi Salsabila[3)], Aulia Najibah Putri[4)], Ken Ditha Tania[5)]
[6)]Winda Kurnia Sari

[1,2,3,4,5,6]Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang
Email: [1)]09031282227086@student.unsri.ac.id, [2)]09031282227057@student.unsri.ac.id,
[3)]09031282227089@student.unsri.ac.id, [4)]09031282227087@student.unsri.ac.id,
[5)]ken_tania@unsri.ac.id, [6)]windakurniasari@unsri.ac.id

Abstrak

Performa pembalap Formula 1 tidak hanya ditentukan oleh hasil akhir balapan, tetapi juga oleh konsistensi catatan waktu dan lap tercepat. Penelitian ini menerapkan algoritma *K-Means clustering* untuk mengelompokkan pembalap berdasarkan performa mereka. Data yang digunakan mencakup hasil balapan resmi musim 2024 yang diterbitkan oleh FIA. Proses pengolahan data mencakup pengumpulan data, *preprocessing*, analisis eksploratori, penerapan algoritma *clustering*, serta evaluasi dan interpretasi hasil. Untuk menentukan jumlah cluster yang optimal, digunakan Metode Elbow dan skor Silhouette, yang menghasilkan empat kelompok pembalap dengan karakteristik performa yang berbeda. Hasil analisis menunjukkan bahwa metode ini berhasil mengidentifikasi pola performa yang relevan, memberikan wawasan bagi tim balap dalam menyusun strategi. Evaluasi menggunakan skor Silhouette menunjukkan bahwa segmentasi yang dihasilkan cukup baik dengan nilai sebesar 0.5735.

Kata Kunci : Formula 1, Clustering Analysis, K-Means, Segmentasi, Driver Performance.

Abstract

The performance of Formula 1 drivers is not solely determined by race results, but also by the consistency of lap times and fastest laps. This study applies the K-Means clustering algorithm to group drivers based on their performance. The data used comprises official race results from the 2024 season, as published by the FIA. The data processing stages include data collection, preprocessing, exploratory analysis, application of the clustering algorithm, and evaluation and interpretation of the results. To determine the optimal number of clusters, the Elbow Method and Silhouette Score were employed, resulting in four groups of drivers with distinct performance characteristics. The analysis findings indicate that this method successfully identifies relevant performance patterns, providing valuable insights for racing teams in formulating strategies. Evaluation using the Silhouette Score shows that the resulting segmentation is reasonably good, with a score of 0.5735.

Keywords: Formula 1, Clustering Analysis, K-Means, Segmentasi, Driver Performance.

1. PENDAHULUAN

Formula One (F1) merupakan ajang balap mobil internasional paling bergengsi yang diselenggarakan oleh *Fédération Internationale de l'Automobile* (FIA), dan terdiri dari dua gelar utama yang diperebutkan: Juara Dunia untuk pembalap dan konstruktor [1]. Dalam praktiknya, performa pembalap tidak hanya diukur dari posisi akhir balapan, tetapi juga dari konsistensi waktu putaran (*lap time*) dan pencapaian lap tercepat. Balapan F1 melibatkan kendaraan dengan teknologi tinggi yang memungkinkan performa ekstrem, sehingga

menghasilkan data performa kompleks yang dapat dianalisis untuk menggali pola kompetitif antar pembalap [2]. F1 tidak hanya menguji kemampuan pembalap dalam mengendalikan mobil di bawah tekanan, tetapi juga mengandalkan kerja sama tim untuk strategi pit stop, aerodinamika, dan pengembangan teknologi.

Pada musim 2024, performa pembalap dapat dianalisis melalui berbagai parameter, seperti catatan waktu dan lap tercepat, yang mencerminkan kemampuan individu dalam menghadapi tantangan sirkuit yang beragam. Segmentasi performa pembalap menjadi penting untuk memahami pola keunggulan kompetitif dan memberikan wawasan bagi tim dalam merancang strategi balapan [3]. Penggunaan algoritma *clustering* dalam analisis data olahraga, khususnya Formula 1, memungkinkan pengelompokan pembalap berdasarkan karakteristik performa yang serupa, sehingga memudahkan identifikasi kekuatan dan kelemahan masing-masing individu [4].

Penelitian ini menggunakan algoritma *K-Means* dengan metrik *Euclidean Distance* untuk mengukur kedekatan antar data numerik seperti selisih waktu putaran dan kecepatan maksimum, yang kemudian digunakan untuk mengelompokkan pembalap ke dalam klaster performa yang serupa. Pendekatan ini didukung oleh penelitian [5], yang berhasil mengelompokkan data nilai ujian nasional SMP di Indonesia tahun 2018/2019 menjadi tiga klaster menggunakan *K-Means* dengan *Euclidean Distance*, menunjukkan efektivitas metode ini untuk segmentasi data kuantitatif seperti performa pembalap F1.

Untuk menentukan jumlah cluster optimal, penelitian ini memanfaatkan *Elbow Method*, yang mengidentifikasi titik di mana penambahan klaster tidak lagi signifikan meningkatkan variansi data, dan *Silhouette Score*, yang mengukur seberapa baik data terpisah dalam klaster berdasarkan kedekatan waktu lap, memastikan segmentasi yang akurat untuk strategi balapan. Pentingnya evaluasi klaster ini sejalan dengan studi [6], yang menggunakan Davies-Bouldin Index untuk menilai kualitas pengelompokan dalam K-Medoids, menegaskan bahwa validasi klaster meningkatkan keandalan hasil analisis.

Meskipun fokus utama penelitian ini adalah segmentasi performa menggunakan *K-Means*, pendekatan-pendekatan lain seperti Random Forest dan DNN telah digunakan dalam studi sebelumnya untuk memprediksi hasil balapan, namun belum banyak yang mengeksplorasi *clustering* untuk pemetaan karakteristik performa. Sebagaimana ditunjukkan oleh penelitian terdahulu [7], menunjukkan bahwa algoritma Machine Learning seperti Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), dan Support Vector Regressor (SVR) mampu memprediksi pemenang balapan dan klasemen akhir dengan tingkat akurasi tinggi. Hasil penelitian ini menghasilkan prediksi yang memiliki korelasi tinggi dengan data aktual, dengan nilai Spearman ρ mencapai 0.903. Dengan memanfaatkan data historis dari Ergast API serta informasi cuaca dan karakteristik trek, model ini menunjukkan bahwa faktor utama yang mempengaruhi hasil balapan adalah tim pembalap, usia pembalap, hasil kualifikasi, dan performa musim sebelumnya. Selain itu, [8] meneliti efektivitas *Deep Neural Network* (DNN) dan Radial Basis *Function Neural Network* (RBFNN) dalam memprediksi hasil balapan Formula 1 pada musim 2021. Dengan menggunakan *dataset* dari Ergast API dan Visual *Crossing* API. Dengan demikian, penelitian ini berkontribusi pada pengembangan model prediksi balapan yang lebih akurat dan dapat digunakan dalam pengambilan keputusan strategis oleh tim Formula 1.

Segmentasi adalah proses pengelompokan suatu populasi ke dalam beberapa kelompok berdasarkan karakteristik yang berbeda satu sama lain [9], yang dalam konteks ini dapat diterapkan untuk mengelompokkan pembalap Formula 1 berdasarkan data kuantitatif seperti jumlah kemenangan dan lap tercepat. Algoritma *clustering*, seperti *K-Means*, telah terbukti mampu mengelompokkan data secara efektif berdasarkan kesamaan karakteristik, sebagaimana dibuktikan dalam penelitian yang menggunakan metode CRISP-DM untuk analisis akademik [10]. Pendekatan ini dapat diadaptasi ke dalam domain olahraga untuk mengolah data

performa musim 2024, yang mencakup variabel-variabel penting seperti konsistensi waktu putaran dan kecepatan maksimum. Dengan demikian, penelitian ini mengisi celah penelitian dengan menawarkan pendekatan segmentasi performa berbasis data kuantitatif secara eksploratif melalui algoritma *K-Means*, sekaligus memberikan perspektif baru dalam penerapan analitik data pada olahraga balap profesional.

Pada penelitian sebelumnya yang berjudul *"Clustering of Football Players Based on Performance Data and Aggregated Clustering Validity Indexes"* [11]. Penelitian ini menganalisis data performa dari musim 2014-15 di delapan liga utama Eropa. Mereka menggunakan ukuran disimilaritas yang disesuaikan dan indeks validitas *clustering* untuk membentuk kelompok, yang dapat diadaptasi untuk mengelompokkan pembalap F1 berdasar waktu lap dan lap tercepat.

Penelitian sebelumnya yang berjudul *"Implementasi Data Mining dalam Pengelompokan Data Pembelian Menggunakan Algoritma K-Means Pada PT. Otomotif 1"* [12], menunjukkan bahwa *K-Means* mampu memberikan wawasan yang berharga mengenai strategi pemasaran dan manajemen inventaris dengan interpretasi pengelompokan yang jelas mengenai pola pembelian konsumen. Pendekatan ini tentunya relevan dengan penelitian Formula 1 karena menunjukkan fleksibilitas *K-Means* dalam mengelompokkan data berbasis kinerja kuantitatif, seperti catatan waktu dan putaran tercepat, yang dapat disesuaikan untuk mengidentifikasi pola keahlian pembalap. Selain itu, penggunaan metrik jarak *Euclidean* dalam penelitian ini memperkuat fondasi teknis untuk analisis data olahraga yang membutuhkan pengukuran kesamaan antar entitas.

Pada penelitian lainnya yang berjudul "*Clustering analysis across different speeds reveals two distinct running techniques with no differences in running economy*" [13], menunjukkan bahwa pelari dapat dikelompokkan berdasarkan teknik lari mereka pada kecepatan yang berbeda, tanpa perbedaan dalam efisiensi ekonomi lari. Ini menunjukkan bahwa *clustering* dapat membantu memahami variasi teknik performa atlet, yang relevan untuk pembalap F1.

Penelitian lain yaitu oleh [14] ini menggunakan algoritma *clustering* inkremental dinamis untuk menganalisis kemampuan lari pelatih sepak bola berdasarkan data performa fisik yang terus diperbarui. Pendekatan ini dapat relevan untuk F1 musim 2024, di mana data waktu lap dapat diperbarui secara *real-time* selama musim berlangsung untuk membentuk kelompok performa yang dinamis.

Pada penelitian lain yang berjudul "Racing Your Rival: Cluster Analysis of Formula 1 Drivers" [15], pendekatan *clustering* digunakan untuk mengelompokkan pembalap berdasarkan karakteristik performa mereka, seperti posisi akhir rata-rata, strategi kualifikasi, serta hubungan antar-teammate. Hasil penelitian menunjukkan bahwa konsistensi performa lebih penting dibandingkan kemenangan insidental, dan tim yang memiliki kombinasi pembalap dengan keunggulan yang saling melengkapi lebih cenderung sukses.

Penelitian lainnya yang berjudul "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode *K-Means*" [16] membahas bagaimana mengelompokan data transaksi menggunakan *clustering K-Means*. Hasilnya terdapat pola tersembunyi dalam data saat metode *clustering* diimplementasikan, yang relevan dengan penelitian ini karena dapat digunakan untuk mengelompokkan pembalap berdasarkan pola performa mereka.

Penelitian oleh Sholeh dan Aeni [17] menunjukkan bahwa metode *Elbow* efektif dalam menentukan jumlah *cluster* optimal berdasarkan perubahan *Within-Cluster Sum of Squares* (WCSS), sementara *Silhouette Score* menilai kedekatan antar data dalam klaster. Penerapan kedua metode ini dalam segmentasi pembalap Formula 1 dapat memastikan jumlah klaster yang dipilih mencerminkan perbedaan signifikan dalam catatan waktu dan lap tercepat selama musim 2024.

Penelitian ini bertujuan untuk mengelompokkan performa pembalap Formula 1 musim 2024 berdasarkan data kuantitatif seperti waktu lap,

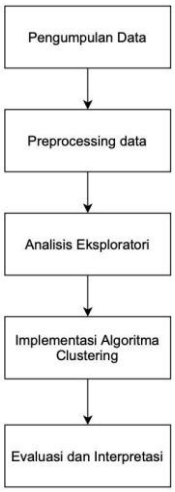
jumlah kemenangan, dan kecepatan maksimum dengan menggunakan algoritma *K-Means* dan pendekatan *Euclidean Distance*. Untuk memperoleh hasil pengelompokan yang optimal, penelitian ini juga memanfaatkan *Elbow Method* dan *Silhouette Score* dalam menentukan jumlah klaster yang paling representatif. Evaluasi terhadap kualitas klaster dilakukan dengan menggunakan Davies-Bouldin Index guna memastikan keandalan segmentasi data yang dihasilkan. Secara keseluruhan, penelitian ini bertujuan untuk memberikan pendekatan analitis berbasis teknologi dalam mengolah dan menganalisis performa pembalap F1, serta menawarkan perspektif baru dalam pengambilan keputusan strategis oleh tim balap.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan berbasis data untuk menganalisis dan mengelompokkan performa para pembalap Formula 1 pada musim 2024 dengan menggunakan teknik pembelajaran tanpa pengawasan (*unsupervised learning)*, khususnya algoritma pengelompokan *K-Means*. Dengan menggunakan data kuantitatif dari hasil balapan, penelitian ini bertujuan untuk memberikan wawasan sistematis tentang bagaimana pola kinerja pembalap berdasarkan waktu dan lap tercepat.

2.1. Metode Pengolahan Data

Untuk mengelompokan data agar kesamaan karakteristiknya efisien, maka dipilih metode algoritma *K-Means*, sebagaimana didukung oleh penelitian sebelumnya dalam analisis olahraga. *Dataset* yang digunakan mencakup hasil balapan resmi musim 2024 yang diterbitkan oleh FIA. Tahapan dalam metode pengolahan data seperti pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

1. Pengumpulan Data
   Data yang digunakan dalam penelitian ini diperoleh dari hasil balapan Formula 1 musim 2024. *Dataset* tersebut diunduh dari repositori GitHub milik toUpperCase78, yang berisi hasil balapan sepanjang musim 2024 dalam format CSV. *Dataset* dapat diakses melalui tautan berikut: https://github.com/toUpperCase78/formula1-datasets/blob/master/Formula1_2024season_raceResults.csv.

	Track	Position	No	Driver	Team	Starting Grid	Laps	Time/Retired	Points	Set Fastest Lap	Fastest Lap Time
2	Bahrain	1	1	Max Verstappen	Red Bull Racing Honda RBPT	1	57	1:31:44.742	26	Yes	1:32.608
3	Bahrain	2	11	Sergio Perez	Red Bull Racing Honda RBPT	5	57	+22.457	18	No	1:34.364
4	Bahrain	3	55	Carlos Sainz	Ferrari	4	57	+25.110	15	No	1:34.507
5	Bahrain	4	16	Charles Leclerc	Ferrari	2	57	+39.669	12	No	1:34.090
6	Bahrain	5	63	George Russell	Mercedes	3	57	+46.788	10	No	1:35.065
7	Bahrain	6	4	Lando Norris	McLaren Mercedes	7	57	+48.458	8	No	1:34.476
8	Bahrain	7	44	Lewis Hamilton	Mercedes	9	57	+50.324	6	No	1:34.722
9	Bahrain	8	81	Oscar Piastri	McLaren Mercedes	8	57	+56.082	4	No	1:34.983
10	Bahrain	9	14	Fernando Alonso	Aston Martin Aramco Mercedes	6	57	+74.887	2	No	1:34.199
11	Bahrain	10	18	Lance Stroll	Aston Martin Aramco Mercedes	12	57	+93.216	1	No	1:35.632
12	Bahrain	11	24	Guanyu Zhou	Kick Sauber Ferrari	17	56	+1 lap	0	No	1:35.458
13	Bahrain	12	20	Kevin Magnussen	Haas Ferrari	15	56	+1 lap	0	No	1:35.570
14	Bahrain	13	3	Daniel Ricciardo	RB Honda RBPT	14	56	+1 lap	0	No	1:35.163
15	Bahrain	14	22	Yuki Tsunoda	RB Honda RBPT	11	56	+1 lap	0	No	1:35.833
16	Bahrain	15	23	Alexander Albon	Williams Mercedes	13	56	+1 lap	0	No	1:35.723
17	Bahrain	16	27	Nico Hulkenberg	Haas Ferrari	10	56	+1 lap	0	No	1:34.834
18	Bahrain	17	31	Esteban Ocon	Alpine Renault	19	56	+1 lap	0	No	1:36.226
19	Bahrain	18	10	Pierre Gasly	Alpine Renault	20	56	+1 lap	0	No	1:34.805
20	Bahrain	19	77	Valtteri Bottas	Kick Sauber Ferrari	16	56	+1 lap	0	No	1:36.202
21	Bahrain	20	2	Logan Sargeant	Williams Mercedes	18	55	+2 laps	0	No	1:34.735

Gambar 2. Kumpulan data

2. *Preprocessing* Data
   Preprocessing data bertujuan meningkatkan kualitas data dengan menghilangkan nilai hilang, outlier, serta menormalisasi skala variabel. Selain itu, dilakukan penanganan ketidakpastian untuk memastikan data siap dianalisis

3. Analisis Eksploratori
   Melakukan eksplorasi awal terhadap data untuk mengidentifikasi pola, tren, dan hubungan antar variabel.
4. Implementasi Algoritma *Clustering*
   Memanfaatkan algoritma *K-Means* untuk membagi pembalap ke dalam kelompok berdasarkan parameter performa yang telah ditentukan sebelumnya.
5. Evaluasi dan Interpretasi Menginterpretasikan hasil *clustering* dalam konteks performa pembalap F1 dan mengidentifikasi karakteristik unik setiap *cluster*.

2.2. Algoritma *K-Means Clustering*

Setelah fase eksplorasi selesai, proses pengelompokan dilakukan dengan algoritma *K-Means*. Jumlah klaster terbaik ditentukan lewat gabungan *Elbow Method* dan *Silhouette Score*, yang membantu menemukan titik ideal dengan mempertimbangkan keseimbangan antara jumlah klaster dan kualitas pemisahan. Setelah itu, algoritma *K-Means* diaplikasikan dengan mengukur jarak antara setiap pembalap dan pusat klaster menggunakan *Euclidean Distance*.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Keterangan :

d(x, y) = jarak objek antara nilai data dan nilai pusat *cluster*

xi = nilai data dari dimensi ke-i

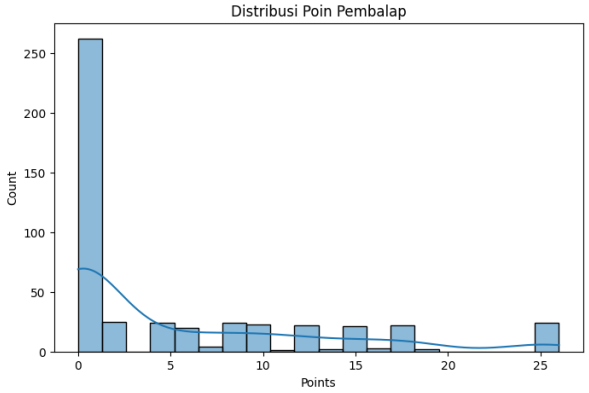
yi = nilai pusat *cluster* dari dimensi ke-i

n = banyaknya dimensi atau atribut data

3. HASIL DAN PEMBAHASAN

3.1. Eksplorasi Data

Tahap eksplorasi data dilakukan untuk memahami struktur *dataset* sebelum pemrosesan lebih lanjut. Langkah awal meliputi menampilkan beberapa baris pertama *dataset*, memeriksa tipe data dan jumlah entri dengan *df.info*(), serta mendeteksi missing values menggunakan *df.isnull*().*sum*(). Selain itu, untuk memahami distribusi performa pembalap, dilakukan visualisasi distribusi poin menggunakan histogram dengan *sns.histplot*(), dilengkapi *Kernel Density Estimation* (KDE) untuk mengidentifikasi pola penyebaran data. Analisis ini membantu dalam menentukan strategi *preprocessing* yang sesuai sebelum menerapkan algoritma *clustering*.



Gambar 3. Distribusi Poin Pembalap

Eksplorasi data menunjukkan bahwa terdapat variasi yang cukup signifikan dalam performa pembalap. Beberapa pembalap mencatatkan waktu tercepat yang jauh lebih konsisten dibandingkan yang lain, sementara ada pula yang memiliki catatan waktu lebih lambat namun tetap meraih poin yang tinggi.

3.2. *Preprocessing* Data

Sebelum menerapkan algoritma *clustering*, dilakukan *preprocessing* untuk memastikan data dalam format yang sesuai dan bebas dari inkonsistensi. Langkah-langkah *preprocessing* yang diterapkan dalam penelitian ini meliputi:

1. Konversi Waktu ke Detik

Dalam *dataset*, waktu balapan dan lap tercepat disajikan dalam format mm:ss.sss atau hh:mm:ss.sss, yang tidak dapat langsung digunakan dalam perhitungan numerik. Oleh karena itu, dilakukan konversi waktu ke dalam satuan detik menggunakan fungsi *time_to_seconds*(). Fungsi ini memproses nilai waktu dengan memisahkan string berdasarkan tanda titik dua (:) dan mengkonversinya ke dalam satuan detik.

2. Penanganan Nilai Hilang

Beberapa kolom dalam *dataset* memiliki nilai hilang (*missing values*) yang dapat mengganggu analisis.

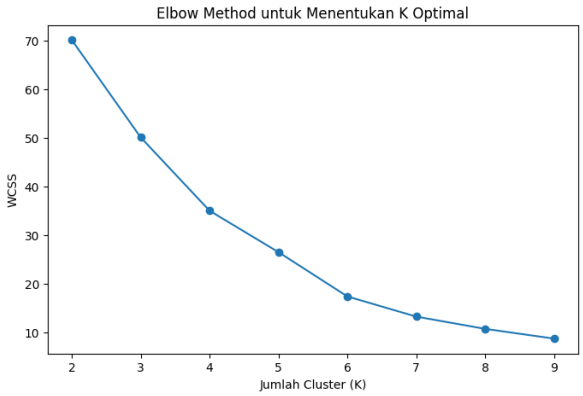
Untuk mengatasi hal ini, nilai yang hilang pada waktu balapan dan lap tercepat diisi dengan median dari masing-masing kolom. Pendekatan ini dipilih untuk mengurangi distorsi data akibat outlier.

3. Normalisasi Data

Karena *dataset* memiliki variabel dengan skala yang berbeda, dilakukan normalisasi menggunakan metode *StandardScaler*. Teknik ini mentransformasikan setiap fitur agar memiliki distribusi dengan mean = 0 dan standar deviasi = 1, sehingga semua fitur memiliki bobot yang setara dalam proses *clustering*.

3.3. Menentukan Jumlah Klaster (K)

Untuk menentukan jumlah *cluster* yang optimal dalam *K-Means*, digunakan metode Elbow dan *Silhouette Score*. Hasil analisis menunjukkan bahwa elbow pada grafik WCSS terjadi di K = 4 atau 5, dan berdasarkan perhitungan *Silhouette Score*, nilai tertinggi diperoleh pada K = 4. Oleh karena itu, jumlah klaster yang digunakan dalam penelitian ini adalah 4 klaster.



Gambar 7. *Elbow Method* untuk Menentukan K Optimal

3.4. Hasil *Clustering*

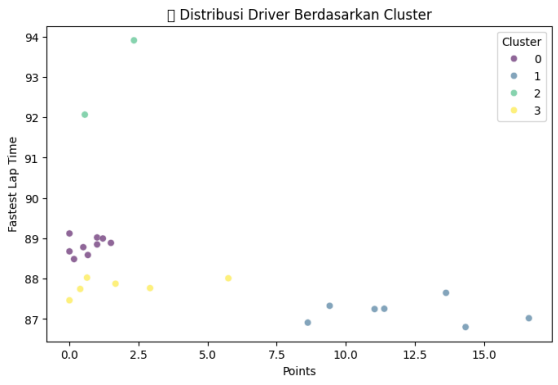
Berdasarkan hasil penerapan algoritma *K-Means* dengan K = 4, pembalap dikelompokkan ke dalam empat *cluster* berdasarkan karakteristik performa mereka. Hasil *clustering* dapat dilihat pada tabel berikut:

Tabel 1. Data Hasil *Clustering*

No	***Driver***	***Cluster***
1	Alexander Albon	0
2	Carlos Sainz	1
3	Charles Leclerc	1
4	Daniel Ricciardo	3
5	Esteban Ocon	0
6	Fernando Alonso	3
7	Franco Colapinto	2
8	George Russell	1
9	Guanyu Zhou	0
10	Jack Doohan	0
11	Kevin Magnussen	3
12	Lance Stroll	0
13	Lando Norris	1
14	Lewis Hamilton	1
15	Liam Lawson	0
16	Logan Sargeant	3
17	Max Verstappen	1
18	Nico Hulkenberg	0
19	Oliver Bearman	2
20	Oscar Piastri	1
21	Pierre Gasly	3
22	Sergio Perez	3
23	Valtteri Bottas	0
24	Yuki Tsunoda	0

3.5 Interpretasi Hasil *Clustering*

Berdasarkan hasil *clustering*, para pembalap terbagi ke dalam empat kelompok berdasarkan performa mereka, terutama dari segi jumlah poin yang diperoleh dan waktu lap tercepat. Visualisasi hasil *clustering* dapat dilihat pada gambar di bawah ini.



Gambar 9. Distribusi Driver Berdasarkan Cluster

1. *Cluster* 0
   Cluster ini terdiri dari pembalap dengan perolehan poin yang sangat rendah (0-2 poin).

Tabel 2. *Cluster* 0

No	*Driver*	*Cluster*
1	Alexander Albon	0
2	Esteban Ocon	0
3	Guanyu Zhou	0
4	Jack Doohan	0
5	Lance Stroll	0
6	Liam Lawson	0
7	Nico Hulkenberg	0
8	Valtteri Bottas	0
9	Yuki Tsunoda	0

2. *Cluster* 1
   Cluster 1 ini berisi pembalap dengan perolehan poin tertinggi (8-16 poin). Mereka juga memiliki waktu lap tercepat di bawah 87,5 detik.

Tabel 3. *Cluster* 1

No	*Driver*	*Cluster*
1	Carlos Sainz	1
2	Charles Leclerc	1
3	George Russell	1
4	Lando Norris	1
5	Lewis Hamilton	1
6	Max Verstappen	1
7	Oscar Piastri	1

3. *Cluster* 2
   Pembalap dalam *cluster* ini memiliki jumlah poin sekitar 2-3, tetapi waktu lap mereka lebih lambat dibanding pembalap lain (sekitar 92-94 detik).

Tabel 4. *Cluster* 2

No	*Driver*	*Cluster*
1	Franco Colapinto	2
2	Oliver Bearman	2

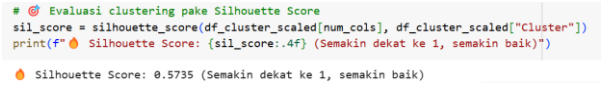
4. *Cluster* 3
   Cluster ini terdiri dari pembalap yang memperoleh 1-5 poin dan memiliki catatan waktu lap tercepat sekitar 88-89 detik.

Tabel 5. *Cluster* 3

No	*Driver*	*Cluster*
1	Daniel Ricciardo	3
2	Fernando Alonso	3
3	Kevin Magnussen	3
4	Logan Sargeant	3
5	Pierre Gasly	3
6	Sergio Perez	3

3.6 Evaluasi

Untuk menilai kualitas pengelompokan yang telah dilakukan, *Silhouette Score* dipakai sebagai ukuran utama. *Silhouette Score* menghitung sejauh mana setiap data dalam suatu klaster terpisah dari data di klaster lainnya.



Gambar 10. *Silhouette Score*

Hasil evaluasi menunjukkan bahwa *Silhouette Score* sebesar 0.5735, yang berarti *clustering* yang dilakukan cukup baik. Nilai ini menunjukkan bahwa sebagian besar pembalap berada dalam *cluster* yang sesuai, namun ada beberapa yang masih berada di batas antar *cluster*.

4. KESIMPULAN

Setelah perhitungan dengan menggunakan algoritma *K-Mean* dilakukan hasilnya, penulis memperoleh hasil *clustering* pada performa pembalap Formula 1 musim 2024 berdasarkan catatan waktu dan lap tercepat . Data hasil balapan resmi dari FIA diolah melalui tahap ***preprocessing***, analisis eksploratori, dan *clustering*, dengan jumlah cluster optimal sebanyak empat ditentukan melalui kombinasi *Elbow Method* dan *Silhouette Score*. *Euclidean Distance* digunakan untuk menghitung jarak antar data ke centroid klaster, menghasilkan empat kelompok pembalap dengan karakteristik

performa yang berbeda: cluster 0 (poin sangat rendah), cluster 1 (poin tertinggi dengan waktu lap tercepat), cluster 2 (poin rendah dengan waktu lambat), dan cluster 3 (poin sedang dengan waktu kompetitif). Penilaian *Silhouette Score* menunjukkan nilai 0.5735, hal ini menandakan kualitas segmentasi cukup baik dengan pemisahan klaster yang memadai. Hasil ini memberikan wawasan strategis bagi tim F1 dalam memahami pola keunggulan kompetitif pembalap, seperti konsistensi dan kecepatan, serta menawarkan pendekatan baru dalam analisis data olahraga berbasis teknologi *clustering*.

DAFTAR PUSTAKA

[1] FIA, "2025 FIA Formula One World Championship," Federation Internationale de l'Automobile. Diakses: 22 April 2025. [Daring]. Tersedia pada: https://www.fia.com/events/fia-formula-one-world-championship/season-2025/2025-fia-formula-one-world-championship

[2] Formula 1, "F1 - The Official Home of Formula 1® Racing." Diakses: 27 Februari 2025. [Daring]. Tersedia pada: https://www.formula1.com/

[3] D. Martínez-Cevallos, A. Proaño-Grijalva, M. Alguacil, D. Duclos-Bastías, dan D. Parra-Camacho, "Segmentation of Participants in a Sports Event Using Cluster Analysis," *Sustainability*, vol. 12, no. 14, hlm. 5641, Jul 2020, doi: 10.3390/su12145641.

[4] F. Gregorry dan Y. Nataliani, "*Clustering* Performa Pemain Basket Berdasarkan Posisi dan Statistik Pemain Menggunakan Metode Fuzzy c-Means," *J. Transform.*, vol. 20, no. 1, Art. no. 1, Jul 2022, doi: 10.26623/transformatika.v20i1.5137.

[5] A. Aditya, I. Jovian, dan B. N. Sari, "Implementasi K-Means *Clustering* Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 1, Art. no. 1, Jan 2020, doi: 10.30865/mib.v4i1.1784.

[6] M. N. P. Pamulang, M. N. Aini, dan U. Enri, "Komparasi Distance Measure Pada K-Medoids *Clustering* untuk Pengelompokkan Penyakit ISPA," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, hlm. 99–107, Jun 2021, doi: 10.29408/edumatic.v5i1.3359.

[7] H. Sicoie, "Machine Learning framework for Formula 1 race winner and championship standings predictor," 2022.

[8] K. Franssen, "COMPARISON OF NEURAL NETWORK ARCHITECTURES IN RACE PREDICTION Predicting the racing outcomes of the 2021 Formula 1 season," Tilburg University, School of Humanities and Digital Sciences, 2022.

[9] Y. Suhanda, I. Kurniati, dan S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means *Clustering* Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *J. Teknol. Inform. Dan Komput.*, vol. 6, no. 2, hlm. 12–20, Sep 2020, doi: 10.37012/jtik.v6i2.299.

[10] A. Javed, B. S. Lee, dan D. M. Rizzo, "A benchmark study on time series *clustering*," *Mach. Learn. Appl.*, vol. 1, hlm. 100001, Sep 2020, doi: 10.1016/j.mlwa.2020.100001.

[11] S. Akhanli dan C. Hennig, "*Clustering* of football players based on performance data and aggregated *clustering* validity indexes," arXiv.org. Diakses: 27 Februari 2025. [Daring]. Tersedia pada: https://arxiv.org/abs/2204.09793v1

[12] D. D. Susilo, S. S. Hilabi, B. Priyatna, dan E. Novalia, "Implementasi Data Mining dalam Pengelompokan Data Pembelian Menggunakan Algoritma K-Means Pada PT.Otomotif 1," *Jutisi J. Ilm. Tek. Inform. Dan Sist. Inf.*, vol. 13, no. 1, hlm. 476, Apr 2024, doi: 10.35889/jutisi.v13i1.1836.

[13] A. R. Rivadulla, X. Chen, D. Cazzola, G. Trewartha, dan E. Preatoni, "*Clustering* analysis across different speeds reveals two distinct running techniques with no differences in running economy," *Sports Biomech.*, hlm. 1–24, Jul 2024, doi: 10.1080/14763141.2024.2372608.

[14] P. Zhao, F. Xue, dan X. Zhang, "Analysis of the Running Ability Mining Model of Football Trainers Based on Dynamic Incremental *Clustering* Algorithm," *Comput. Intell. Neurosci.*, vol. 2022, hlm. 3255886, Jun 2022, doi: 10.1155/2022/3255886.

[15] D. Nagle, "Racing Your Rival: Cluster Analysis of Formula 1 Drivers," Renée Crown University, 2022.

[16] N. Afiasari, N. Suarna, dan N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode K-Means," *J. Saintekom Sains Teknol. Komput. Dan Manaj.*, vol. 13, no. 1, Art. no. 1, Mar 2023, doi: 10.33020/saintekom.v13i1.402.

[17] M. Sholeh dan K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model *Clustering* dengan Menggunakan Algoritma K-Means," *STRING Satuan Tulisan Ris. Dan Inov. Teknol.*, vol. 8, no. 1, Art. no. 1, Agu 2023, doi: 10.30998/string.v8i1.16388.