INFOMATEK: Jurnal Informatika, Manajemen dan Teknologi
Volume 27 Nomor 1, Juni 2025

PENERAPAN METODE K-MEANS CLUSTERING UNTUK SEGMENTASI PERFORMA PEMBALAP F1 SEASON 2024

Mutia Sahira*, Adella Salsabila, Shofi Salsabila, Aulia Najibah Putri, Ken Ditha Tania Winda Kurnia Sari

Sistem Informasi, Universitas Sriwijaya, Indonesia

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 Silhouette Score menunjukkan bahwa segmentasi yang dihasilkan cukup baik dengan nilai sebesar 0.5735.

Kata kunci: *Clustering Analysis*; *Driver Performance*; *Formula 1*; *K-Means*; *Segmentation*

I. PENDAHULUAN

Formula One (F1) merupakan kejuaraan balap mobil internasional paling prestisius yang diselenggarakan oleh Fédération Internationale de l'Automobile (FIA), Formula 1 terdiri dari dua gelar Juara Dunia, satu untuk pembalap dan satu untuk konstruktor (FIA, 2012). Dalam praktiknya, performa pembalap tidak hanya diukur dari posisi akhir balapan, tetapi juga dari konsistensi catatan waktu dan kemampuan mencatatkan lap tercepat. Balapan ini melibatkan mobil-mobil berteknologi tinggi yang dirancang khusus untuk mencapai kecepatan maksimum, sering kali melebihi 300 km/jam, di berbagai sirkuit global seperti Monaco, Silverstone, dan Abu Dhabi (Formula 1, 2024). 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 (Martínez-Cevallos dkk., 2020). Penggunaan algoritma clustering dalam analisis data olahraga, khususnya Formula 1, memungkinkan pengelompokan pembalap berdasarkan

*) 09031282227086@student.unsri.ac.id

Diterima: 23 April 2025
Direvisi: 29 Mei 2025
Disetujui:25 Juni 2025
DOI: 10.23969/infomatek.v27i1.24297

karakteristik performa yang serupa, sehingga memudahkan identifikasi kekuatan dan kelemahan masing-masing individu (Gregorry & Nataliani, 2022).

Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan data performa dengan menghitung jarak setiap titik data ke pusat klaster menggunakan metode Euclidean Distance, yaitu metrik yang mengukur jarak linier antar dua titik dalam ruang numerik berdasarkan selisih catatan waktu dan lap tercepat. Pendekatan ini didukung oleh penelitian Aditya dkk., (2020), 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 Pamulang dkk., (2021), yang menggunakan Davies-Bouldin Index untuk menilai kualitas pengelompokan dalam K-Medoids, menegaskan bahwa validasi klaster meningkatkan keandalan hasil analisis.

Analisis prediktif dalam F1 telah berkembang pesat, sebagaimana ditunjukkan oleh Sicoie (2022), Machine Learning untuk memprediksi pemenang balapan dan klasemen kejuaraan. Hasil penelitian ini menunjukkan bahwa model seperti Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), dan Support Vector Regressor (SVR) dapat 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, Franssen, (2022) 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 (Suhanda dkk., 2020), 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 (Javed dkk., 2020). 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 menyiratkan terkait tentang segmentasi performa pembalap Formula 1 secara analitis, sekaligus

menawarkan perspektif baru dalam analisis data olahraga berbasis teknologi.

Pada penelitian sebelumnya yang berjudul *“Clustering of Football Players Based on Performance Data and Aggregated Clustering Validity Indexes"* (Akhanli & Hennig, 2022). 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”* (Susilo dkk., 2024), 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.

Penelitian oleh Niko dkk., (2023) mengindikasikan bahwa penggunaan algoritma K-Means efektif untuk mengelompokkan data kuantitatif demi memberikan wawasan yang dapat diimplementasikan. Dalam penelitian ini, data produk ritel dikelompokkan berdasarkan ketersediaan dan inventaris di gudang menjadi tiga klaster: ketersediaan rendah (Low), sedang (Medium), dan tinggi (High). Proses validasi menggunakan metode validasi silang k-fold dan perbandingan antara perhitungan manual serta alat RapidMiner, yang menunjukkan konsistensi hasil. Penemuan ini relevan untuk analisis performa pembalap Formula 1 musim 2024, karena menegaskan bahwa K-Means dapat digunakan untuk mengelompokkan data berbasis metrik kuantitatif seperti catatan waktu dan lap tercepat, dengan potensi validasi yang kuat untuk memastikan keandalan segmentasi. Pendekatan ini mendukung tujuan penelitian untuk mengidentifikasi pola performa pembalap yang dapat diinterpretasikan secara praktis oleh tim balap.

Pada penelitian lainnya yang berjudul "Clustering analysis across different speeds reveals two distinct running techniques with no differences in running economy" (Rivadulla dkk., 2024), menunjukan bahwa pelari dapat dikelompokan 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 (Zhao dkk., 2022) 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" (Nagle, 2022), 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" (Afiasari dkk., 2023) 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 (Sholeh & Aeni, 2023) 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.

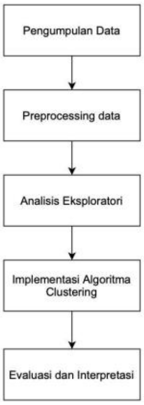
Penelitian ini bertujuan untuk mengelompokkan performa pembalap Formula 1 musim 2024 berdasarkan data kuantitatif seperti waktu lap, jumlah kemenangan, dan kecepatan maksimum dengan menggunakan algoritma K-Means dan pendekatan Euclidean Distance. Untuk memperoleh hasil pengelompokan yang optimal, penelitian 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. Selain itu, penelitian ini turut mengembangkan model prediktif berbasis Machine Learning dengan memanfaatkan data historis balapan, faktor cuaca, serta karakteristik trek, sehingga dapat menghasilkan prediksi hasil balapan yang lebih akurat. 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.

II. METODOLOGI

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.

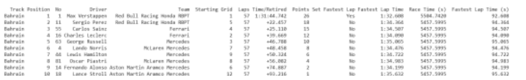
2.1 Metode Pengolahan Data

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 bersumber dari hasil balapan Formula 1 musim 2024.



Track	Position	No	Driver	Team	Starting Grid	Laps	Time/Retired	Points	Set Fastest Lap	Fastest Lap Time	Race Time (s)	Fastest Lap Time (s)
Bahrain	1	1	Max Verstappen	Red Bull Racing Honda RBPT	1	57	1:31:44.742	26	Yes	1:32.608	5504.7420	92.608
Bahrain	2	11	Sergio Perez	Red Bull Racing Honda RBPT	5	57	+22.457	18	No	1:34.364	5457.5995	94.364
Bahrain	3	55	Carlos Sainz	Ferrari	4	57	+25.110	15	No	1:34.507	5457.5995	94.507
Bahrain	4	16	Charles Leclerc	Ferrari	2	57	+39.669	12	No	1:34.090	5457.5995	94.090
Bahrain	5	63	George Russell	Mercedes	3	57	+46.788	10	No	1:35.065	5457.5995	95.065
Bahrain	6	4	Lando Norris	McLaren Mercedes	7	57	+48.458	8	No	1:34.476	5457.5995	94.476
Bahrain	7	44	Lewis Hamilton	Mercedes	9	57	+50.324	6	No	1:34.722	5457.5995	94.722
Bahrain	8	81	Oscar Piastri	McLaren Mercedes	8	57	+56.082	4	No	1:34.983	5457.5995	94.983
Bahrain	9	14	Fernando Alonso	Aston Martin Aramco Mercedes	6	57	+74.887	2	No	1:34.199	5457.5995	94.199
Bahrain	10	18	Lance Stroll	Aston Martin Aramco Mercedes	12	57	+93.216	1	No	1:35.632	5457.5995	95.632

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.

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

Keterangan :

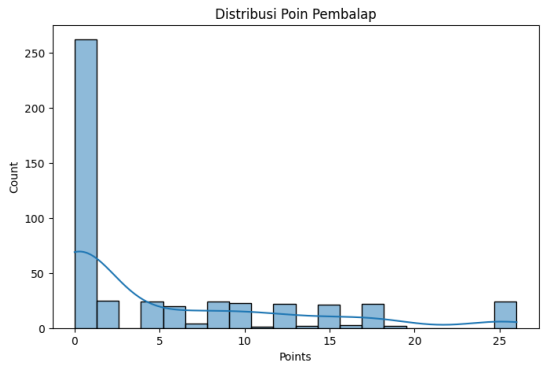
$d(x, y)$ = jarak objek antara nilai data dan nilai pusat *cluster*
xi = nilai data dari dimensi ke-i
yi = nilai pusat *cluster* dari dimensi ke-i
n = banyaknya dimensi atau atribut data

III. 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.

```
def time_to_seconds(time_str):
    if isinstance(time_str, str):
        parts = time_str.split(":")
        if len(parts) == 2:  # mm:ss.sss format
            return int(parts[0]) * 60 + float(parts[1])
        elif len(parts) == 3:  # hh:mm:ss.sss format
            return int(parts[0]) * 3600 + int(parts[1]) * 60 + float(parts[2])
    return np.nan

df["Race Time (s)"] = df["Time/Retired"].apply(time_to_seconds)
df["Fastest Lap Time (s)"] = df["Fastest Lap Time"].apply(time_to_seconds)
```

Gambar 4. Konversi Waktu ke 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.

Handling Missing Values

```
[8] df["Race Time (s)"] = df["Race Time (s)"].fillna(df["Race Time (s)"].median())
    df["Fastest Lap Time (s)"] = df["Fastest Lap Time (s)"].fillna(df["Fastest Lap Time (s)"].median())
```

Gambar 5. Penanganan Nilai Hilang

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*.

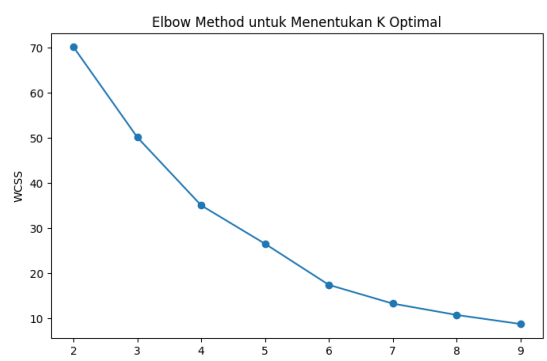
Normalisasi Data

```
[10] scaler = StandardScaler()
     df_scaled = scaler.fit_transform(df_clean.iloc[:, 1:])
```

Gambar 6. Normalisasi Data

3.3. Menentukan Jumlah Klaster (K)

Untuk menentukan jumlah *cluster* yang optimal dalam K-Means, digunakan metode Elbow dan Silhouette Score.



Gambar 7. Elbow Method untuk Menentukan K Optimal

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.

```
optimal_k = max(silhouette_scores, key=silhouette_scores.get)
print(f"Optimal K berdasarkan Silhouette Score: {optimal_k}")
```

```
Optimal K berdasarkan Silhouette Score: 4
```

Gambar 8. Optimal K berdasarkan Silhouette Score

3.4. Hasil Clustering

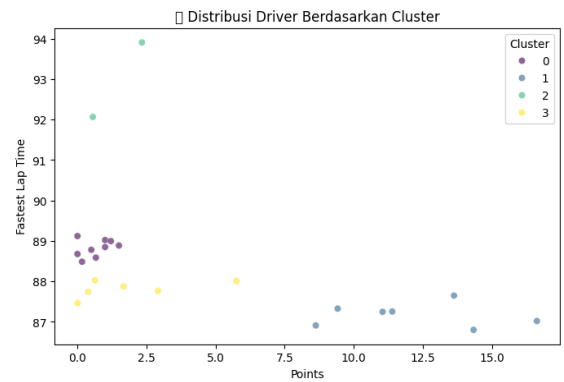
Berdasarkan hasil penerapan algoritma K-Means dengan K = 4, pembalap dikelompokkan ke dalam tiga *cluster* berdasarkan karakteristik performa mereka. Hasil *clustering* dapat dilihat pada tabel berikut:

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.

```
# 🎯 Evaluasi clustering pake Silhouette Score
sil_score = silhouette_score(df_cluster_scaled[num_cols], df_cluster_scaled["Cluster"])
print(f"🔥 Silhouette Score: {sil_score:.4f} (Semakin dekat ke 1, semakin baik)")

🔥 Silhouette Score: 0.5735 (Semakin dekat ke 1, semakin baik)
```

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*.

IV. 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

Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika*

Budidarma, *4*(1), Article 1. https://doi.org/10.30865/mib.v4i1.1784

Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means. *Jurnal Saintekom : Sains, Teknologi, Komputer Dan Manajemen*, *13*(1), Article 1. https://doi.org/10.33020/saintekom.v13i1.402

Akhanli, S., & Hennig, C. (2022, April 20). *Clustering of football players based on performance data and aggregated clustering validity indexes*. arXiv.Org. https://arxiv.org/abs/2204.09793v1

FIA. (2012, Maret 12). *2025 FIA Formula One World Championship*. Federation Internationale de l'Automobile. https://www.fia.com/events/fia-formula-one-world-championship/season-2025/2025-fia-formula-one-world-championship

Formula 1. (2024). *F1—The Official Home of Formula 1® Racing*. https://www.formula1.com/

Franssen, K. (2022). *comparison Of Neural Network Architectures In Race Prediction Predicting the racing outcomes of the 2021 Formula 1 season*. Tilburg University.

Gregorry, F., & Nataliani, Y. (2022). Clustering Performa Pemain Basket Berdasarkan Posisi dan Statistik Pemain Menggunakan Metode Fuzzy c-Means. *Jurnal Transformatika*, *20*(1), Article 1. https://doi.org/10.26623/transformatika.v20i1.5137

Javed, A., Lee, B. S., & Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications*, *1*, 100001. https://doi.org/10.1016/j.mlwa.2020.100001

Martínez-Cevallos, D., Proaño-Grijalva, A., Alguacil, M., Duclos-Bastías, D., & Parra-Camacho, D. (2020). Segmentation of Participants in a Sports Event Using Cluster Analysis. *Sustainability*, *12*(14), 5641. https://doi.org/10.3390/su12145641

Nagle, D. (2022). *Racing Your Rival: Cluster Analysis of Formula 1 Drivers*. Renée Crown University.

Niko, N. S., Rahman, A., Atmaja, D. M. U., & Basri, A. (2023). Klasterisasi Stok Produk Retail Untuk Menetukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means. *Bulletin of Information Technology (BIT)*, *4*(3), Article 3. https://doi.org/10.47065/bit.v4i3.736

Pamulang, M. N. P., Aini, M. N., & Enri, U. (2021). Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokkan Penyakit ISPA. *Edumatic: Jurnal Pendidikan Informatika*, *5*(1), 99–107. https://doi.org/10.29408/edumatic.v5i1.3359

Rivadulla, A. R., Chen, X., Cazzola, D., Trewartha, G., & Preatoni, E. (2024). Clustering analysis across different speeds reveals two distinct running techniques with no differences in running economy. *Sports Biomechanics*, 1–24. https://doi.org/10.1080/14763141.2024.2372608

Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, *8*(1), Article 1. https://doi.org/10.30998/string.v8i1.16388

Sicoie, H. (2022). *Machine Learning framework for Formula 1 race winner and championship standings predictor*.

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika dan Komputer*, *6*(2), 12–20. https://doi.org/10.37012/jtik.v6i2.299

Susilo, D. D., Hilabi, S. S., Priyatna, B., & Novalia, E. (2024). Implementasi Data Mining dalam Pengelompokan Data Pembelian Menggunakan Algoritma K-Means Pada PT.Otomotif 1. *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, *13*(1), 476. https://doi.org/10.35889/jutisi.v13i1.1836

Zhao, P., Xue, F., & Zhang, X. (2022). Analysis of the Running Ability Mining Model of Football Trainers Based on Dynamic Incremental Clustering Algorithm. *Computational Intelligence and Neuroscience*, *2022*, 3255886. https://doi.org/10.1155/2022/3255886