

PENERAPAN METODE K-MEANS CLUSTERING UNTUK SEGMENTASI PERFORMA PEMBALAP FORMULA 1 SEASON 2025

Muhammad Wahyu Hikmalsyah, Muhammad Dzaky Hasyim,
Gabriel Sebastian Santoso, Fathoni, Ken Ditha Tania, Allsela Meiriza

Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih KM. 32, Indralaya, Indonesia
Wahyuhikmalsyah3@gmail.com

ABSTRAK

Performa pembalap Formula 1 merupakan hasil interaksi kompleks antara kemampuan teknis, konsistensi waktu lap, dan efisiensi dalam memperbaiki posisi saat balapan. Penelitian ini menerapkan algoritma *K-Means Clustering* untuk melakukan segmentasi performa pembalap Formula 1 musim 2025 secara objektif dan terukur. Data yang digunakan bersumber dari hasil balapan resmi yang diterbitkan oleh FIA, yang mencakup metrik perolehan poin, waktu lap rata-rata, hingga perubahan posisi (*position gain*). Tahapan penelitian meliputi pengumpulan data, *preprocessing*, analisis eksploratori, penentuan jumlah kluster optimal melalui metode *Elbow* dan *Silhouette Score*, hingga evaluasi hasil pengelompokan. Hasil penelitian menunjukkan bahwa penggunaan lima kluster ($K=5$) memberikan gambaran paling representatif terhadap dinamika kompetisi musim 2025, mulai dari kelompok elit dengan performa dominan hingga kelompok pendatang baru dengan karakteristik teknis tertentu. Evaluasi menggunakan *Silhouette Score* menghasilkan nilai sebesar 0,4678, yang menunjukkan bahwa segmentasi yang dihasilkan cukup baik dan akurat dalam mengidentifikasi pola performa antar-kluster. Temuan ini memberikan wawasan strategis bagi tim balap untuk memahami peta persaingan sebagai landasan pengembangan teknis dan strategi balapan di masa depan.

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

1. PENDAHULUAN

Formula 1 (F1) telah bertransformasi menjadi ekosistem olahraga yang sangat bergantung pada integrasi data analitik dan teknologi tingkat tinggi untuk menentukan keunggulan kompetitif. Xu (2023) mengidentifikasi adanya transisi regulasi Power Unit Formula 1 dari tahun 2023 menuju 2026 yang berfokus pada efisiensi termodinamika. Peningkatan kapasitas MGU-K hingga 350 kW serta penggunaan bahan bakar berkelanjutan dinilai mampu meningkatkan efisiensi termal secara signifikan, namun juga meningkatkan kompleksitas pengendalian kendaraan akibat lonjakan torsi pada kecepatan rendah. Kondisi ini menuntut pembalap untuk memiliki kemampuan adaptasi tinggi terhadap teknologi hibrida sebagai faktor penting dalam menentukan performa di lintasan [1].

Dalam konteks tersebut, analisis performa pembalap tidak lagi hanya didasarkan pada hasil akhir balapan, tetapi juga pada data kuantitatif seperti waktu lap, konsistensi, serta manajemen ban. Pendekatan berbasis data ini memungkinkan identifikasi pola performa yang lebih objektif melalui teknik pengelompokan (*clustering*). Seiring dengan persiapan menuju regulasi besar tahun 2026 yang berfokus pada efisiensi energi dan keberlanjutan, analisis performa pembalap pada musim 2025 menjadi sangat penting sebagai dasar pengambilan keputusan strategis bagi tim balap [2], [3].

Di era big data, Formula 1 menghasilkan data telemetri dalam jumlah besar yang memerlukan teknik analisis khusus untuk mengolahnya. Salah satu metode yang efektif adalah *K-Means Clustering*, yang mampu

mengelompokkan pembalap berdasarkan kemiripan karakteristik performa, bukan sekadar peringkat klasemen. Dengan mengintegrasikan variabel seperti durasi pit stop, strategi ban, dan konsistensi waktu lap, analisis dapat memberikan gambaran yang lebih mendalam mengenai pola performa pembalap sepanjang musim [3], [4].

Namun, tantangan utama dalam penerapan *K-Means* adalah menentukan jumlah kluster (K) yang optimal. Oleh karena itu, digunakan metode *Elbow Method* dan *Silhouette Score* untuk memastikan hasil pengelompokan yang representatif dan objektif [5]. Validasi tambahan menggunakan *Davies-Bouldin Index* (DBI) juga diperlukan untuk mengukur kualitas kluster yang terbentuk [6].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan performa pembalap Formula 1 musim 2025 menggunakan metode *K-Means Clustering* berdasarkan variabel kuantitatif seperti waktu lap, jumlah kemenangan, dan kecepatan maksimum. Selain itu, penelitian ini juga mengintegrasikan pendekatan machine learning untuk mendukung analisis prediktif terhadap hasil balapan. Diharapkan hasil penelitian ini dapat memberikan wawasan strategis yang lebih mendalam bagi tim balap dalam menghadapi dinamika kompetisi Formula 1 yang semakin kompleks.

Metodologi penelitian ini dirancang secara sistematis dengan menggunakan pendekatan kuantitatif berbasis pembelajaran tanpa pengawasan (*unsupervised learning*). Tahapan dimulai dengan pengumpulan data sekunder dari statistik resmi FIA yang mencakup seluruh populasi pembalap reguler

musim 2025 (N=20) dari 24 seri balapan. Data tersebut kemudian melalui tahap *preprocessing*, yang meliputi konversi waktu ke detik, penanganan nilai hilang dengan imputasi median, serta normalisasi menggunakan *StandardScaler* guna menyetarakan bobot setiap variabel. Setelah data siap, dilakukan analisis eksploratori untuk mengidentifikasi pola awal sebelum menerapkan algoritma *K-Means*. Penentuan jumlah kluster optimal dilakukan melalui integrasi *Elbow Method* dan *Silhouette Score* dengan perhitungan jarak *Euclidean Distance*. Tahap akhir melibatkan evaluasi kualitas kluster dan interpretasi hasil secara mendalam untuk mendeskripsikan karakteristik teknis unik dari setiap kelompok pembalap yang terbentuk.

2. TINJAUAN PUSTAKA

2.1. K-Means Clustering

K-Means Clustering merupakan salah satu algoritma dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan karakteristik. Algoritma ini bekerja dengan meminimalkan variansi dalam satu kluster (*within-cluster sum of squares*) sehingga setiap anggota dalam kluster memiliki tingkat kemiripan yang tinggi [7].

Metode ini banyak digunakan karena efisien dalam menangani data numerik dalam jumlah besar dan mampu mengidentifikasi pola tersembunyi dalam dataset [8].

2.2. Segmentasi Data dalam Analisis Performa

Segmentasi merupakan proses pengelompokan data berdasarkan kesamaan karakteristik tertentu. Dalam konteks olahraga, segmentasi digunakan untuk mengelompokkan atlet berdasarkan performa mereka menggunakan indikator kuantitatif seperti kecepatan, konsistensi, dan efisiensi performa [9].

Pada Formula 1, variabel seperti waktu lap, posisi finis, dan kecepatan maksimum dapat digunakan sebagai dasar segmentasi untuk memahami pola performa pembalap [10].

2.3. Penerapan Clustering dalam Olahraga

Beberapa penelitian menunjukkan bahwa teknik clustering efektif dalam menganalisis performa atlet. Sahira et al. (2025) menggunakan K-Means untuk mengelompokkan pembalap Formula 1 berdasarkan statistik balapan dan berhasil mengidentifikasi kategori performa tinggi, sedang, dan rendah [10].

Selain itu, penelitian oleh Guan dan Xue (2025) menunjukkan bahwa clustering dapat membantu dalam proses seleksi dan pelatihan atlet dengan mengelompokkan berdasarkan indikator performa multidimensi [11].

Penelitian lain oleh Rivadulla et al. (2026) menunjukkan bahwa clustering mampu mengidentifikasi pola teknik berlari yang berbeda meskipun memiliki efisiensi energi yang sama. Hal ini membuktikan bahwa metode clustering mampu

mengungkap pola performa yang tidak terlihat melalui analisis konvensional [12].

2.4. Validasi Jumlah Kluster

Penentuan jumlah kluster (K) merupakan tahap penting dalam K-Means Clustering. Metode *Elbow* digunakan untuk menentukan titik optimal berdasarkan penurunan nilai *Sum of Squared Errors* (SSE), sedangkan *Silhouette Score* digunakan untuk mengukur kualitas pemisahan antar kluster [13], [14].

Nilai *silhouette* yang tinggi menunjukkan bahwa data telah dikelompokkan dengan baik.

Selain itu, *Davies-Bouldin Index* (DBI) digunakan sebagai metode evaluasi tambahan untuk mengukur rasio antara kohesi dalam kluster dan separasi antar kluster. Nilai DBI yang lebih rendah menunjukkan kualitas kluster yang lebih baik [15].

2.5. Machine Learning dalam Formula 1

Pemanfaatan machine learning dalam Formula 1 semakin berkembang, terutama dalam prediksi hasil balapan. Urdhwarshie (2025) menggunakan model TabNet untuk memprediksi posisi finis pembalap berdasarkan data historis dan variabel pra-balap. Hasil penelitian menunjukkan bahwa pendekatan berbasis data mampu meningkatkan akurasi prediksi secara signifikan [16].

Selain itu, analisis perilaku mengemudi menggunakan teknik clustering juga telah diterapkan untuk mengidentifikasi pola akselerasi, pengereman, dan kecepatan di tikungan [17].

Hal ini relevan dengan penelitian ini yang menggunakan parameter performa seperti kecepatan maksimum dan konsistensi waktu lap.

2.6. Penelitian Terdahulu

Beberapa penelitian terdahulu telah memberikan landasan yang kuat mengenai efektivitas penggunaan teknik analisis data dalam ekosistem Formula 1. Penelitian oleh Sahira et al. (2025) menjadi studi yang paling relevan karena menerapkan algoritma K-Means Clustering untuk mengelompokkan pembalap musim 2024 berdasarkan statistik balapan resmi, yang berhasil mengidentifikasi kategori performa tinggi, sedang, dan rendah [10].

Pendekatan berbasis data ini juga dikembangkan oleh Sasikumar et al. (2025) yang mengintegrasikan variabel kuantitatif ke dalam model pembelajaran mendalam untuk mendukung sistem pengambilan keputusan strategis terkait pit stop [18].

Sejalan dengan hal tersebut, Urdhwarshie (2025) memanfaatkan model machine learning (TabNet) untuk memprediksi hasil balapan berdasarkan data historis dan variabel pra-balap guna meningkatkan akurasi analisis performa [19].

Terakhir, studi oleh Krzysztoń dan Smółka (2026) menekankan pentingnya kualitas data telemetri sebagai instrumen utama dalam prediksi hasil balapan menggunakan teknik pembelajaran mesin [20].

Secara kolektif, literatur ini membuktikan bahwa integrasi algoritma pengelompokan dan pemodelan prediktif mampu memberikan wawasan strategis yang mendalam bagi tim balap dalam menghadapi kompleksitas kompetisi.

3. METODE PENELITIAN

Metodologi penelitian ini dirancang secara sistematis untuk memastikan bahwa proses segmentasi performa pembalap Formula 1 musim 2025 dilakukan dengan langkah-langkah yang valid dan dapat dipercaya. Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan algoritma K-Means *Clustering* sebagai metode untuk membagi data ke dalam beberapa kelompok. Selain itu, penelitian ini juga menerapkan pendekatan berbasis data dengan teknik pembelajaran tanpa pengawasan (*unsupervised learning*) guna menganalisis serta mengelompokkan performa pembalap. Dataset numerik hasil balapan diproses guna memetakan karakteristik pembalap berdasarkan metrik waktu tempuh serta catatan lap tercepat.

3.1. Metode Pengumpulan Data

Kajian ini mengadopsi Metode Sekunder sebagai instrumen fundamental dalam perolehan data guna menjamin kedalaman analisis. Pemanfaatan data statistik resmi hasil balapan Formula 1 musim 2025 yang diterbitkan oleh otoritas terkait, meliputi catatan performa akumulatif dari seri pembuka di Australia hingga seri final di Abu Dhabi.

3.2. Populasi dan Teknik Pengambilan Sampel

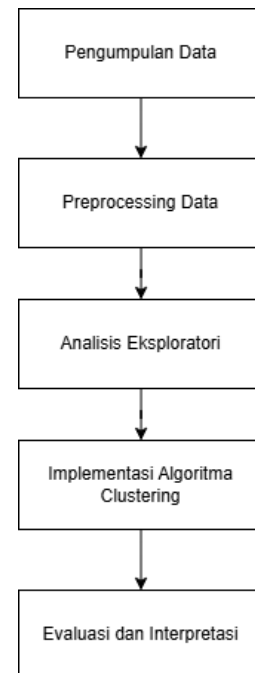
Penelitian ini tidak menggunakan metode pengambilan sampel secara acak, melainkan menerapkan teknik total sampling (sensus). Subjek penelitian mencakup seluruh populasi, yaitu semua pembalap reguler yang berkompetisi pada musim Formula 1 tahun 2025 dengan jumlah total $N = 20$ pembalap. Pendekatan sensus ini dipilih karena jumlah populasi yang relatif kecil dan terukur, sehingga memungkinkan peneliti untuk menganalisis seluruh anggota populasi secara menyeluruh tanpa risiko terjadinya kesalahan sampling.

Dari sisi kecukupan data, penelitian ini melibatkan 20 pembalap dengan dataset yang mencakup catatan performa kumulatif dari 24 seri balapan. Volume data tersebut dinilai cukup representatif untuk diolah menggunakan algoritma machine learning, karena mencerminkan performa pembalap secara komprehensif sepanjang musim kompetisi.

3.3. Alur Penelitian

Pengelompokan data dilakukan menggunakan algoritma K-Means guna menjamin efisiensi identifikasi kesamaan karakteristik. Prosedur ini merujuk pada praktik analisis performa olahraga dalam literatur terdahulu. Dataset penelitian bersumber dari catatan resmi hasil balapan musim

2025 yang dipublikasikan oleh FIA. Seluruh tahapan sistematis pengolahan data divisualisasikan secara mendalam melalui alur pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

3.4. Pengumpulan Data

Penelitian ini tidak menggunakan metode pengambilan sampel secara acak, melainkan menerapkan teknik total sampling (sensus). Subjek penelitian mencakup seluruh populasi, yaitu semua pembalap reguler yang berkompetisi pada musim Formula 1 tahun 2025 dengan jumlah total $N = 20$ pembalap. Pendekatan sensus ini dipilih karena jumlah populasi yang relatif kecil dan terukur, sehingga memungkinkan peneliti untuk menganalisis seluruh anggota populasi secara menyeluruh tanpa risiko terjadinya kesalahan sampling.

Dari sisi kecukupan data, penelitian ini melibatkan 20 pembalap dengan dataset yang mencakup catatan performa kumulatif dari 24 seri balapan. Volume data tersebut dinilai cukup representatif untuk diolah menggunakan algoritma machine learning, karena mencerminkan performa pembalap secara komprehensif sepanjang musim kompetisi.

Data primer bersumber dari statistik kumulatif hasil balapan Formula 1 sepanjang musim kompetisi 2025. Struktur dataset mentah yang digunakan sebagai dasar analisis disajikan secara visual pada Gambar 2.

	Track	Position	No	Driver	Team	Starting Grid	Laps	Time/Retired	Points	Set Fastest Lap	Fastest Lap Time
0	Australia	1	4	Lando Norris	McLaren Mercedes	1	57	1:42:04.304	25	Yes	1:23.167
1	Australia	2	1	Max Verstappen	Red Bull Racing/Honda RBPT	3	57	+0.895	18	No	1:23.081
2	Australia	3	43	George Russell	Mercedes	4	57	+8.481	15	No	1:25.045
3	Australia	4	12	Kimi Antonelli	Mercedes	16	57	+10.135	12	No	1:24.901
4	Australia	5	23	Alexander Albon	Williams Mercedes	6	57	+12.773	10	No	1:24.597

Gambar 2. Kumpulan data

3.5. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk meningkatkan integritas informasi melalui eliminasi nilai kosong serta data pencilan. Skala variabel diseragamkan melalui normalisasi guna menjamin kesiapan data sebelum tahap analisis statistik. Penanganan ketidakpastian data diterapkan secara ketat untuk menjaga reliabilitas hasil akhir penelitian.

3.6. Analisis Eksploratori

Analisis eksploratif dijalankan pada tahap awal guna mengidentifikasi kecenderungan pola serta korelasi antar variabel performa.

3.7. Implementasi Algoritma Clustering

Algoritma K-Means diaplikasikan untuk membagi populasi pembalap ke dalam partisi kelompok yang homogen. Segmentasi didasarkan pada matriks parameter performa yang telah ditetapkan dalam kerangka penelitian.

3.8. Evaluasi dan Interpretasi

Hasil pengelompokan diinterpretasikan secara mendalam berdasarkan konteks kompetisi balap mobil Formula 1. Karakteristik teknis yang bersifat unik pada setiap kluster diidentifikasi secara sistematis. Hal ini dilakukan untuk mendeskripsikan peta kekuatan kompetisi secara objektif.

3.9. Algoritma K-Means Clustering

Tahap pengelompokan dilakukan menggunakan algoritma K-Means pasca penyelesaian fase eksplorasi. Penentuan jumlah kluster dilakukan melalui integrasi Elbow Method dan Silhouette Score. Pendekatan ini diterapkan guna mencapai keseimbangan antara kuantitas kelompok dengan kualitas pemisahan data. Jarak antara subjek pembalap dengan pusat kluster diukur menggunakan rumus *Euclidean Distance*. Perhitungan matematis tersebut menjamin akurasi posisi data terhadap sentroid yang ditetapkan.

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

Keterangan:

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

x_i = nilai data dari dimensi ke- i

y_i = nilai pusat cluster dari dimensi ke- i

n = banyaknya dimensi atau atribut data

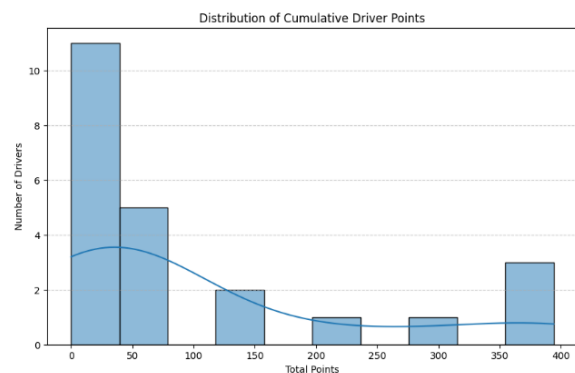
4. HASIL DAN PEMBAHASAN

4.1. Algoritma K-Means Clustering

Analisis data eksploratif (*Exploratory Data Analysis*) dilakukan sebagai tahap fundamental untuk mengidentifikasi struktur serta karakteristik intrinsik dataset. Prosedur ini diawali dengan inspeksi profil data yang komprehensif, mencakup evaluasi tipe data,

validasi integritas jumlah entri, serta identifikasi nilai yang hilang (*missing values*). Langkah ini krusial dalam menentukan protokol penanganan data yang tepat guna menjamin validitas analisis selanjutnya.

Selain itu, analisis distribusi pada variabel performa pengemudi dilakukan melalui integrasi teknik *Histogram* dan *Kernel Density Estimation* (KDE) pada parameter perolehan poin kumulatif. Penerapan histogram dengan penyesuaian interval (*bins*) bertujuan untuk memetakan frekuensi observasi secara diskrit, sementara kurva KDE digunakan untuk memodelkan kontur penyebaran data secara kontinu. Visualisasi ini berfungsi sebagai instrumen diagnostik dalam merumuskan strategi pra-pemrosesan data yang akurat, seperti normalisasi atau transformasi statistik, guna meningkatkan performa dan ketepatan algoritma pada tahap pemodelan atau pengelompokan (*clustering*).



Gambar 3. Distribusi Poin Pembalap

Hasil eksplorasi menunjukkan variabilitas performa teknis yang lebar antar pembalap pada musim 2025. Konsistensi waktu lap ditemukan tidak selalu berkorelasi linear dengan akumulasi poin kejuaraan. Perbedaan karakteristik data ini mendasari perlunya standarisasi variabel sebelum tahap pengelompokan.

4.2. Preprocessing Data

Konsistensi format data dijamin melalui tahap pra-pemrosesan sebelum penerapan algoritma pengelompokan. Prosedur ini dilakukan untuk mengeliminasi inkonsistensi pada dataset. Tiga langkah utama diterapkan untuk menyiapkan variabel penelitian.

4.3. Konversi Waktu ke Detik

Data waktu balapan ditransformasikan dari format temporal menjadi satuan detik numerik. Pemrosesan dilakukan dengan memisahkan *string* berdasarkan indikator titik dua guna memfasilitasi perhitungan matematis.

```
import pandas as pd

# 1. Fungsi Konversi Waktu ke Detik
def time_to_seconds(time_str):
    if pd.isna(time_str) or time_str == '':
        return None
    try:
        parts = str(time_str).split(':')
        if len(parts) == 3: # hh:mm:ss.sss
            return int(parts[0]) * 3600 + int(parts[1]) * 60 + float(parts[2])
        elif len(parts) == 2: # mm:ss.sssfwcss
            return int(parts[0]) * 60 + float(parts[1])
        else:
            return float(parts[0])
    except:
        return None

# Terapkan konversi pada kolom waktu
df['Race Time Seconds'] = df['Time/Retired'].apply(time_to_seconds)
df['Fastest Lap Seconds'] = df['Fastest Lap Time'].apply(time_to_seconds)
```

Gambar 4. Konversi Waktu ke Detik

4.4. Penanganan Nilai Hilang

Nilai kosong pada variabel waktu diatasi melalui teknik imputasi median. Pendekatan ini dipilih untuk mereduksi dampak pencilan terhadap validitas hasil analisis.

```
# 2. Penanganan Nilai Hilang dengan Median
df['Race Time Seconds'] = df['Race Time Seconds'].fillna(df['Race Time Seconds'].median())
df['Fastest Lap Seconds'] = df['Fastest Lap Seconds'].fillna(df['Fastest Lap Seconds'].median())
```

Gambar 5. Penanganan Nilai Hilang

4.5. Normalisasi Data

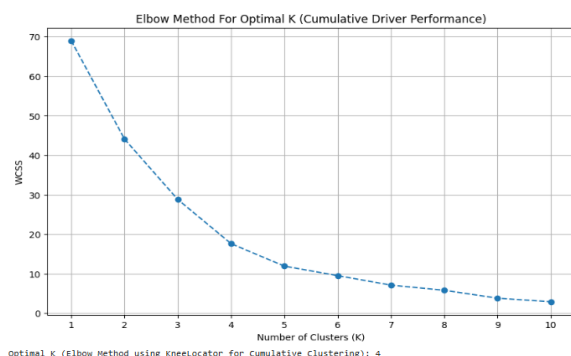
Variabel dengan skala berbeda diseragamkan melalui metode *StandardScaler*. Transformasi ini menghasilkan distribusi fitur dengan rerata nol dan standar deviasi satu. Penyetaraan bobot dilakukan agar setiap variabel memiliki kontribusi seimbang dalam klusterisasi.

```
# 4. Normalisasi Data (StandardScaler)
# Ambil fitur numerik utama: Points dan Position Gained/Lost
features = df[['Points', 'Position Gained/Lost']]
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)
```

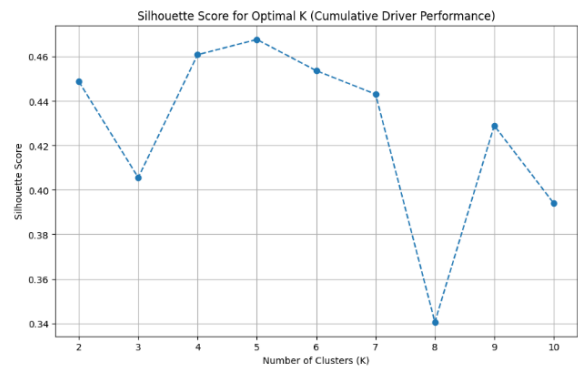
Gambar 6. Normalisasi Data

4.6. Menentukan Jumlah Kluster (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 $K = 5$, dan berdasarkan perhitungan *Silhouette Score*, nilai tertinggi diperoleh pada $K = 5$. Oleh karena itu, jumlah kluster yang digunakan dalam penelitian ini adalah 5 kluster.



Gambar 7. Elbow Method untuk Menentukan K Optimal



Gambar 8. Optimal K berdasarkan Silhouette Score

4.7. Hasil Clustering

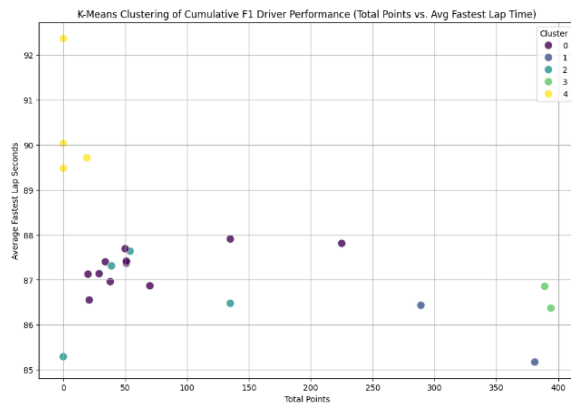
Parameter jumlah kelompok dalam algoritma K-Means ditetapkan melalui integrasi metode Elbow serta Silhouette Score. Analisis grafis terhadap *Within-Cluster Sum of Squares* (WCSS) mengidentifikasi kemunculan titik siku pada nilai $K=5$. Validasi tambahan melalui perhitungan Silhouette Score menunjukkan perolehan koefisien tertinggi pada parameter kluster yang sama. Konsistensi hasil dari kedua instrumen evaluasi tersebut mendasari penetapan struktur lima kluster dalam penelitian ini. Pemilihan nilai K ini ditujukan untuk menjamin derajat kohesi serta separasi objek data yang paling representatif. Jarak antar objek data dalam proses partisi kemudian diukur menggunakan pendekatan matematis Euclidean Distance. Perhitungan tersebut memastikan setiap anggota kelompok memiliki kemiripan fitur yang kuat dengan pusat klusternya.

Tabel 1. Data Hasil Clustering

No	Driver	Cluster
1	Alexander Albon	0
2	Charles Leclerc	0
3	Esteban Ocon	0
4	Fernando Alonso	0
5	Isack Hadjar	0
6	Kimi Antonelli	0
7	Lance Stroll	0
8	Liam Lawson	0
9	Nico Hulkenberg	0
10	Pierre Gasly	0
11	Yuki Tsunoda	0
12	George Russell	1
13	Oscar Piastri	1
14	Carlos Sainz	2
15	Franco Colapinto	2
16	Lewis Hamilton	2
17	Oliver Bearman	2
18	Lando Norris	3
19	Max Verstappen	3
20	Gabriel Bortoleto	4
21	Jack Doohan	4
22	Liam Lawson	4
23	Yuki Tsunoda	4

4.8. Interpretasi Hasil Clustering

Hasil analisis pengelompokan menunjukkan pembagian populasi pembalap ke dalam lima kategori performa utama. Kategorisasi didasarkan pada variabel akumulasi poin kejuaraan serta catatan waktu lap tercepat rata-rata. Visualisasi distribusi spasial antar-kelaster disajikan melalui sebaran data pada Gambar 9.



Gambar 9. Distribusi Driver Berdasarkan Cluster

4.9. Cluster 0

Cluster 0 diidentifikasi sebagai kelompok terbesar yang merepresentasikan persaingan pada zona papan tengah. Karakteristik utama kelompok ini meliputi perolehan poin moderat dengan rata-rata kecepatan lap 87,29 detik. Stabilitas performa terlihat dari perubahan posisi yang relatif rendah dengan nilai sekitar -6,89. Anggota kelompok ini terdaftar secara sistematis dalam Tabel 2.

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

4.10. Cluster 1

Cluster 1 mengelompokkan pembalap dengan profil teknis kecepatan tinggi dan efisiensi manuver menyalip. Rerata waktu lap tercepat pada kelompok ini mencapai 85,80 detik. Peningkatan posisi rata-rata sebesar -68,31 menunjukkan kemampuan pemulihan posisi yang tinggi saat balapan berlangsung. Daftar pembalap pada kategori ini disajikan dalam Tabel 3.

Tabel 3. Cluster 1

No	Driver	Cluster
1	George Russell	1
2	Oscar Piastri	1

4.11. Cluster 2

Klaster 2 mencakup pembalap dengan orientasi taktik bertarung atau *racecraft* yang dominan di lintasan. Meskipun akumulasi poin tidak setinggi klaster elit, kemampuan kenaikan posisi tercatat rata-rata -44,45. Kelompok ini menunjukkan efektivitas dalam mengonversi situasi sulit menjadi raihan poin teknis. Detail anggota kelompok terlampir pada Tabel 4.

Tabel 4. Cluster 2

No	Driver	Cluster
1	Franco Colapinto	2
2	Franco Colapinto	2
3	Lewis Hamilton	2
4	Oliver Bearman	2

4.12. Cluster 3

Cluster 3 diklasifikasikan sebagai kelompok elit yang mendominasi persaingan perebutan gelar juara dunia. Kelompok ini mencatatkan rerata perolehan poin tertinggi sebesar 391,50 poin. Nilai pergerakan posisi sebesar 0,77 mengonfirmasi dominasi kualifikasi pada baris terdepan lintasan. Identifikasi pembalap kelompok ini dipaparkan dalam Tabel 5.

Tabel 5. Cluster 3

No	Driver	Cluster
1	Lando Norris	3
2	Max Verstappen	3

4.13. Cluster 4

Cluster 4 merepresentasikan kelompok dengan tingkat performa teknis paling rendah di dalam kompetisi. Rerata waktu lap tercatat paling lambat sebesar 90,39 detik dengan perolehan poin minimal. Mobilitas posisi pada kelompok ini sangat terbatas akibat kendala kompetitivitas kendaraan atau pengalaman teknis. Rincian data klaster ini disajikan dalam Tabel 6.

Tabel 6. Cluster 4

No	Driver	Cluster
1	Gabriel Bortoleto	4
2	Jack Doohan	4
3	Liam Lawson	4
4	Yuki Tsunoda	4

4.14. Evaluasi

Kualitas pengelompokan diukur menggunakan indikator Silhouette Score sebagai parameter evaluasi utama. Instrumen ini menghitung tingkat pemisahan setiap objek data terhadap kelompok klaster lainnya. Validasi dilakukan untuk memastikan derajat kohesi dan separasi hasil partisi data.

```
print("Silhouette score for cumulative clustering with k=5: ", silhouette_scores_cumulative[silhouette_scores_cumulative.index.max()])
```

```
Silhouette score for cumulative clustering with k=5: 0.467808628628
```

Gambar 10. Silhouette Score

Hasil perhitungan menunjukkan perolehan skor sebesar 0,4678 bagi model pengelompokan tersebut. Capaian nilai ini merepresentasikan bahwa

proses segmentasi yang diterapkan memiliki akurasi yang memadai. Sebagian besar subjek penelitian teridentifikasi berada dalam kelompok yang sesuai dengan karakteristik performanya. Namun, terdapat sejumlah kecil objek data yang menempati posisi pada area perbatasan antar klaster. Validasi statistik ini mengonfirmasi objektivitas hasil partisi data performa pembalap musim 2025.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil melakukan segmentasi performa pembalap Formula 1 musim 2025 secara objektif menggunakan algoritma K-Means Clustering dengan menetapkan lima klaster (K=5) sebagai jumlah kelompok paling representatif berdasarkan integrasi Elbow Method dan Silhouette Score. Hasil analisis menunjukkan pembagian yang jelas mulai dari kelompok elit pada Klaster 3 dengan dominasi poin tertinggi (rata-rata 391,50) hingga Klaster 4 yang menunjukkan performa teknis paling rendah dengan waktu lap rata-rata 90,39 detik. Validasi menggunakan Silhouette Score sebesar 0,4678 mengonfirmasi bahwa model memiliki akurasi pemisahan yang memadai dalam mengidentifikasi pola performa antar-klaster. Sebagai saran untuk pengembangan selanjutnya, penelitian berikutnya dapat mengintegrasikan variabel telemetri yang lebih dinamis seperti degradasi ban, pengaruh cuaca, serta penggunaan model machine learning yang bersifat prediktif guna memberikan gambaran strategi balap yang lebih komprehensif dalam menghadapi transisi regulasi teknis tahun.

DAFTAR PUSTAKA

- [1] J. Xu, "Thermodynamic evaluation of 2026 Power Unit technical regulation changes in Formula 1," *Appl. Comput. Eng.*, vol. 26, pp. 145–150, Nov. 2023, doi: 10.54254/2755-2721/26/20230818.
- [2] L. Monsees and H. E. Næss, "Deprive to survive? – investigating Formula One's environmental policy using a WPR approach," *Int. J. Sport Policy Polit.*, vol. 0, no. 0, pp. 1–23, Jan. 2026, doi: 10.1080/19406940.2026.2618784.
- [3] A. Sasikumar, A. A. Leema, and P. Balakrishnan, "Data-driven pit stop decision support for Formula 1 using deep learning models," *Front. Artif. Intell.*, vol. 8, Nov. 2025, doi: 10.3389/frai.2025.1673148.
- [4] J. De Smedt, A. Yeshchenko, A. Polyvyanyy, J. De Weerd, and J. Mendling, "Process model forecasting and change exploration using time series analysis of event sequence data," *Data Knowl. Eng.*, vol. 145, p. 102145, May 2023, doi: 10.1016/j.datak.2023.102145.
- [5] W. Junthopas and C. Wongoutong, "Pre-Determining the Optimal Number of Clusters for k-Means Clustering Using the Parameters Package in R and Distance Metrics," *Appl. Sci.*, vol. 15, no. 21, Oct. 2025, doi: 10.3390/app152111372.
- [6] F. A. Sulaeman and B. Nurhakim, "Penerapan Algoritma K-Means Dalam Mengelompokkan Atlet PPLP Keolahragaan Di Jawa Barat," no. 2, 2023.
- [7] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci.*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
- [8] M. Afrizal, I. Saputra, R. Satria, and R. Rahmaddeni, "Analisis Performa Algoritma K-Means Clustering untuk Segmentasi Pasar di UMKM," *VISA J. Vis. Ideas*, vol. 5, no. 2, pp. 647–657, May 2025, doi: 10.47467/visa.v5i2.6942.
- [9] A. T. Alhasani, "Unsupervised Clustering of Multivariate Sports Activity Data Using K-Means: A Study on the Sport Data Multivariate Time Series Dataset," *J. Trans. Syst. Eng.*, vol. 3, no. 2, pp. 367–381, Jul. 2025, doi: 10.15157/JTSE.2025.3.2.367-381.
- [10] M. Sahira, A. Salsabila, S. Salsabila, A. N. Putri, K. D. Tania, and W. K. Sari, "Penerapan Metode K-Means Clustering untuk Segmentasi Performa Pembalap F1 Season 2024," *Infomatek*, vol. 27, no. 1, pp. 113–122, Jun. 2025, doi: 10.23969/infomatek.v27i1.24297.
- [11] Y. Guan and W. Xue, "Research on the optimization system of athlete selection and training effect based on big data," *Appl. Math. Nonlinear Sci.*, vol. 10, no. 1, p. 20250563, Jan. 2025, doi: 10.2478/amns-2025-0563.
- [12] A. R. Rivadulla, X. Chen, D. Cazzola, G. Trewartha, and E. Preatoni, "Clustering analysis across different speeds reveals two distinct running techniques with no differences in running economy," *Sports Biomech.*, vol. 25, no. 1, pp. 22–45, Jan. 2026, doi: 10.1080/14763141.2024.2372608.
- [13] B. Hartono and V. Lusiana, "Analisis Metode Elbow SSE, Silhouette Score, dan Jaccard Stability dalam Pemilihan Jumlah Klaster Data yang Optimal," *TIN Terap. Inform. Nusantara*, vol. 6, no. 8, pp. 1521–1532, Jan. 2026, doi: 10.47065/tin.v6i8.9271.
- [14] J. J. A. Limbong and F. M. Likumahwa, "Implementation of the K-Means Algorithm for Clustering Students' Web Programming Course Grades Using Silhouette Score," *J. Syst. Comput. Eng.*, vol. 6, no. 3, pp. 271–280, Aug. 2025, doi: 10.61628/jsce.v6i3.2100.
- [15] D. Hartama and S. Oktaviani, "OPTIMIZATION OF K-MEANS AND K-MEDOIDS CLUSTERING USING DBI SILHOUETTE ELBOW ON STUDENT DATA," *JURTEKSI J. Teknol. Dan Sist. Inf.*, vol. 11, no. 2, pp. 289–296, Mar. 2025, doi: 10.33330/jurteks.v11i2.3531.
- [16] A. Urdhwarsh, "The Use of Machine Learning in Predicting Formula 1 Race Outcomes," Apr.

- 18, 2025, *Preprints*: 2025041471. doi: 10.20944/preprints202504.1471.v1.
- [17] G. Megala and R. Venkatesan, "Driving Sequence Analysis for Intelligent Driver Assistance: Insights on Clustering Techniques," vol. 52, no. 10, 2025.
- [18] A. Sasikumar, A. A. Leema, and P. Balakrishnan, "Data-driven pit stop decision support for Formula 1 using deep learning models," *Front. Artif. Intell.*, vol. 8, Nov. 2025, doi: 10.3389/frai.2025.1673148.
- [19] A. Urdhwareshe, "The Use of Machine Learning in Predicting Formula 1 Race Outcomes," Apr. 18, 2025, *Preprints*: 2025041471. doi: 10.20944/preprints202504.1471.v1.
- [20] S. Krzysztoń and J. Smółka, "Application of machine learning for predicting Formula 1 race results," *J. Comput. Sci. Inst.*, vol. 38, pp. 81–86, Mar. 2026, doi: 10.35784/jcsi.8462.