

ANALISIS PERFORMA FC BARCELONA BERDASARKAN STATISTIK PERTANDINGAN MENGGUNAKAN K-MEANS CLUSTERING

Muhammad Habib Rifai, Ridwan Dwi Irawan, Tri Djoko Santosa

Teknik Informatika, Universitas Duta Bangsa

Jl. Bhayangkara No.55, Tipes, Serengan, Kota Surakarta, Jawa Tengah 57154

habibrifai13@gmail.com

ABSTRAK

Penggunaan data statistik dalam sepak bola terus mengalami perkembangan dan menjadi salah satu pendekatan penting untuk menilai kinerja tim secara lebih objektif. Meskipun demikian, sebagian besar analisis yang dilakukan masih menitikberatkan pada hasil akhir pertandingan sehingga belum memberikan gambaran yang utuh mengenai pola permainan tim. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan pola performa pertandingan FC Barcelona dalam kompetisi La Liga berdasarkan data statistik pertandingan. Metode yang diterapkan adalah K-Means Clustering dengan memanfaatkan variabel penguasaan bola, jumlah tembakan, tembakan tepat sasaran, jumlah gol, jumlah kebobolan, serta akurasi umpan. Data yang digunakan berjumlah 114 pertandingan yang diperoleh dari FBref dan telah melalui proses preprocessing berupa pembersihan data, pemilihan fitur, serta normalisasi. Penentuan jumlah cluster yang optimal dilakukan menggunakan Metode Elbow, sedangkan evaluasi kualitas cluster dilakukan dengan Silhouette Score. Hasil analisis menunjukkan bahwa jumlah cluster terbaik adalah tiga kelompok performa, yaitu agresif, stabil, dan dominan. Cluster stabil mendominasi dengan persentase sebesar 50%, disusul cluster dominan sebesar 27,27% dan cluster agresif sebesar 22,73%. Selain itu, nilai Silhouette Score sebesar 0,209 mengindikasikan bahwa kualitas pengelompokan yang dihasilkan berada pada kategori cukup baik. Oleh karena itu, K-Means Clustering dapat digunakan secara efektif untuk mengidentifikasi dan mengelompokkan pola performa tim berdasarkan statistik pertandingan.

Kata kunci : K-Means Clustering, data mining, performa tim, analisis statistik, sepak bola

1. PENDAHULUAN

Sepak bola adalah salah satu olahraga yang paling terkenal di seluruh dunia dan memiliki dampak yang signifikan dalam bidang ekonomi, sosial, serta teknologi. Perkembangan industri sepak bola saat ini ditandai dengan semakin banyaknya penggunaan teknologi dan data statistik untuk menilai kinerja tim dan pemain[1]. Analisis yang didasarkan pada data kini sangat penting dalam proses pengambilan keputusan strategis karena dapat memberikan gambaran yang objektif tentang seberapa baik permainan sebuah tim[2].

Dalam dunia data mining dan machine learning, berbagai teknik telah diterapkan untuk menganalisis data sepak bola. Penelitian yang telah dilakukan sebelumnya menggunakan metode klasifikasi seperti Naïve Bayes dan Decision Tree untuk meramalkan hasil dari pertandingan sepak bola[3]. Selain itu, penelitian yang dilakukan oleh Ari Yanto dan Testiana (2024) menggunakan metode Naïve Bayes untuk memprediksi pemenang La Liga. Namun, metode klasifikasi ini adalah bagian dari pendekatan pembelajaran terawasi yang memerlukan data yang sudah diberi label, sehingga kurang efektif dalam mengidentifikasi pola yang tersembunyi dalam data hasil pertandingan[4].

Kompetisi sepak bola profesional seperti La Liga menawarkan berbagai data statistik pertandingan yang rinci dan terorganisir, seperti total gol, penguasaan bola, jumlah tembakan, akurasi umpan, hingga expected goals[5]. Salah satu klub yang menunjukkan konsistensi dalam data performa yang baik adalah FC

Barcelona. Data statistik pertandingan FC Barcelona dalam tiga musim bisa digunakan untuk menjelaskan pola kinerja tim dengan lebih mendetail dan objektif. Namun, analisis kinerja tim sepak bola biasanya masih terbatas pada hasil akhir pertandingan atau peringkat di klasemen, sehingga belum dapat menggambarkan pola konsistensi dan karakteristik kinerja pertandingan secara menyeluruh[6].

Pendekatan lain yang dapat digunakan adalah clustering, yaitu teknik pengelompokan data berdasarkan kemiripan karakteristik tanpa memerlukan label pada data terlebih dahulu. Salah satu algoritma clustering yang paling banyak diterapkan adalah K-Means Clustering. Metode ini terkenal karena prosesnya yang sederhana, mudah untuk diterapkan, dan sangat efektif dalam mengelompokkan data yang berbentuk angka. Metode ini dapat mengelompokkan data pertandingan berdasarkan kesamaan statistik, sehingga dapat menghasilkan segmentasi performa, seperti performa yang tinggi, sedang, dan rendah[7].

Beberapa penelitian terdahulu telah menerapkan metode K-Means Clustering pada bidang olahraga untuk melakukan pengelompokan dan analisis data berdasarkan karakteristik tertentu. Penelitian tentang segmentasi kinerja pembalap Formula 1 menunjukkan bahwa metode K-Means dapat memberikan pengelompokan kinerja yang cukup baik. Namun, penelitian itu belum menggunakan evaluasi cluster dengan cara yang paling efektif untuk menentukan jumlah cluster[8].

Oleh sebab itu, penelitian ini menerapkan evaluasi menggunakan Metode Elbow dan Silhouette Score guna memperoleh hasil klaster yang lebih optimal serta mampu merepresentasikan data dengan lebih baik. Berdasarkan permasalahan tersebut, penelitian ini bertujuan menganalisis performa tim FC Barcelona pada kompetisi La Liga menggunakan metode K-Means Clustering berdasarkan data statistik pertandingan. Variabel yang digunakan dalam analisis meliputi penguasaan bola, jumlah tembakan, tembakan tepat sasaran, jumlah gol, jumlah kebobolan, dan akurasi umpan. Melalui penelitian ini, diharapkan dapat diperoleh gambaran yang lebih objektif mengenai pola konsistensi serta kecenderungan performa tim sepanjang musim kompetisi.

2. TINJAUAN PUSTAKA

Beberapa penelitian sebelumnya telah membahas penerapan data mining dan machine learning dalam analisis performa olahraga, khususnya sepak bola. Analisis data statistik pertandingan digunakan untuk membantu proses evaluasi performa tim maupun pemain secara lebih objektif berdasarkan data numerik pertandingan. Berbagai metode seperti klasifikasi dan clustering telah diterapkan untuk mengubah data statistik olahraga menjadi informasi yang mendukung proses pengambilan keputusan.

Metode klasifikasi sering dipakai dalam penelitian sepak bola untuk meramalkan hasil pertandingan serta mengelompokkan pemain berdasarkan data statistik. Namun, metode klasifikasi yang termasuk dalam pendekatan pembelajaran terawasi memerlukan data yang sudah diberi label, sehingga kurang efektif untuk mengidentifikasi pola tersembunyi dalam data pertandingan[9]. Dengan demikian, metode clustering dapat dijadikan alternatif karena mampu mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label data terlebih dahulu.

Metode K-Means Clustering adalah salah satu algoritma pengelompokan yang sering digunakan dalam pengolahan data angka karena prosesnya yang mudah dan efisien. Dalam analisis sepak bola, metode ini bisa digunakan untuk mengelompokkan kinerja pertandingan berdasarkan statistik seperti penguasaan bola, jumlah tembakan, jumlah gol, dan akurasi umpan. Hasil dari pengelompokan ini dapat dimanfaatkan untuk memahami pola konsistensi kinerja tim sepanjang satu musim kompetisi.

2.1. Penelitian Terdahulu

Banyak penelitian telah dilakukan mengenai penggunaan machine learning dalam sepak bola dengan menggunakan metode klasifikasi. Studi yang menggunakan algoritma Naïve Bayes dan Decision Tree menunjukkan bahwa metode klasifikasi bisa dipakai untuk meramalkan hasil pertandingan sepak bola berdasarkan data statistik pertandingan[3]. Selain itu, metode Naïve Bayes juga diterapkan untuk memprediksi siapa yang akan menjadi juara La Liga

dengan berdasarkan pada kinerja klub selama kompetisi berlangsung[4].

Penelitian lain menerapkan metode klasifikasi untuk mengidentifikasi jenis pemain sepak bola berdasarkan data statistik kinerja mereka. Metode Naïve Bayes dapat digunakan untuk mengklasifikasikan jenis penyerang berdasarkan statistik permainan dengan tingkat akurasi yang tertentu[4]. Selain itu, algoritma C4.5 juga diterapkan dalam pemilihan pemain tengah untuk tim nasional sepak bola, berdasarkan pada atribut statistik dari permainan mereka[10].

Selain metode klasifikasi, penelitian yang menggunakan clustering juga mulai banyak digunakan dalam dunia olahraga. Penelitian tentang segmentasi kinerja pembalap Formula 1 dengan menggunakan metode K-Means Clustering berhasil menciptakan beberapa kelompok kinerja berdasarkan karakteristik statistik dari para pembalap[8]. Penelitian itu menunjukkan bahwa metode pengelompokan dapat digunakan untuk mengidentifikasi pola kinerja tanpa perlu adanya label data sebelumnya.

Metode K-Means Clustering dalam data mining memiliki sejumlah kelebihan, seperti kemudahan dalam penerapan dan efektivitasnya dalam mengolah data numerik. Namun, cara ini memiliki kekurangan dalam menentukan jumlah cluster yang paling tepat dan sangat dipengaruhi oleh posisi centroid awal[11]. Karena itu, metode evaluasi klaster seperti Metode Elbow dan Skor Silhouette sering digunakan untuk meningkatkan kualitas hasil klastering.

Dari beberapa penelitian yang ada, penggunaan machine learning dalam analisis sepak bola masih banyak menekankan pada metode klasifikasi untuk meramalkan hasil pertandingan dan mengklasifikasikan pemain. Penelitian ini berbeda karena menggunakan metode K-Means Clustering untuk mengelompokkan kinerja pertandingan FC Barcelona berdasarkan statistik pertandingan dalam satu musim kompetisi La Liga. Selain itu, penelitian ini juga menggunakan evaluasi kluster dengan metode Elbow dan Skor Silhouette untuk mendapatkan jumlah kluster yang optimal dan dapat mewakili.

2.2. Data Mining

Data mining merupakan proses pengolahan data yang bertujuan untuk menemukan pola, keterkaitan, maupun informasi tertentu dari sejumlah data berukuran besar[12]. Teknik data mining sering dipakai di berbagai bidang, termasuk olahraga, karena dapat mendukung analisis data dengan cara yang lebih teratur dan objektif. Dalam studi ini, teknik data mining diterapkan untuk menganalisis statistik pertandingan sepak bola agar dapat mengidentifikasi pola kinerja tim berdasarkan karakteristik data yang ada dalam pertandingan.

2.3. Clustering

Clustering merupakan teknik pengelompokan data yang dilakukan berdasarkan tingkat kemiripan

karakteristik tertentu. Melalui metode ini, data dalam kelompok yang sama cenderung memiliki kesamaan yang lebih tinggi dibandingkan dengan data pada kelompok lainnya. Clustering adalah metode pembelajaran tanpa pengawasan karena pengelompokan dilakukan tanpa memanfaatkan label kelas yang sudah ada sebelumnya[13]. Metode ini sering dipakai untuk mengidentifikasi pola yang tidak terlihat dalam data dan mendukung proses pemisahan berdasarkan ciri-ciri tertentu.

Pada penelitian ini, teknik clustering digunakan untuk mengelompokkan kinerja pertandingan FC Barcelona ke dalam beberapa kategori berdasarkan statistik pertandingan seperti penguasaan bola, jumlah tembakan, jumlah gol, dan akurasi umpan.

2.4. K-means Clustering

Algoritma K-Means merupakan salah satu metode clustering dalam data mining yang melakukan pengelompokan data berdasarkan perhitungan jarak antar data. Metode ini memiliki kelebihan karena relatif mudah dipahami, sederhana dalam penerapan, serta mampu menghasilkan pengelompokan yang cukup baik sehingga banyak digunakan dalam berbagai penelitian. Meskipun demikian, K-Means juga memiliki kelemahan, yaitu hasil pengelompokan sangat dipengaruhi oleh penentuan centroid awal[7]. Langkah-langkah algoritma K-Means adalah sebagai berikut:

- Menentukan jumlah cluster (K).
- Menetapkan centroid awal secara acak.
- Menghitung jarak antara setiap data dengan centroid menggunakan metode seperti Euclidean Distance.
- Menempatkan data ke dalam cluster dengan centroid terdekat.
- Memperbarui nilai centroid berdasarkan rata-rata anggota pada masing-masing cluster..
- Mengulangi langkah sebelumnya hingga tidak ada perubahan anggota cluster.

Metode ini dipilih karena mudah diterapkan, efisien dalam mengolah data angka, dan cocok untuk analisis statistik pertandingan yang berbentuk data kuantitatif.

2.5. Data Statistik Pertandingan

Statistik dalam pertandingan sepak bola adalah informasi angka yang menunjukkan bagaimana kinerja tim selama pertandingan. Berbagai variabel statistik yang sering digunakan antara lain:

- Penguasaan bola
- Jumlah tembakan
- Tembakan tepat sasaran
- Jumlah gol
- Jumlah kebobolan
- Akurasi umpan

Dalam penelitian ini, statistik pertandingan yang dianalisis diambil dari pertandingan FC Barcelona selama tiga musim tertentu di kompetisi La Liga.

2.6. Evaluasi Cluster

Evaluasi cluster dilakukan untuk mengetahui tingkat kualitas dari hasil pengelompokan data yang telah diperoleh. Beberapa metode yang umum digunakan dalam proses evaluasi antara lain Metode Elbow dan Silhouette Score, yang digunakan untuk menentukan jumlah cluster yang optimal sekaligus mengukur kualitas hasil pengelompokan[14]. berikut penjelasan dari berbagai evaluasi cluster:

a. Elbow method

Metode Elbow merupakan teknik yang digunakan untuk menentukan jumlah cluster (K) yang optimal pada algoritma K-Means dengan melihat grafik hubungan antara jumlah cluster dan nilai Within Cluster Sum of Squares.

b. Silhouette Score

Evaluasi ini penting agar hasil pengelompokan tidak hanya valid secara matematis, tetapi juga dapat memberikan interpretasi yang jelas dan bermakna.

2.7. Machine Learning

Machine Learning merupakan bidang penelitian yang mempelajari algoritma serta model statistik yang memungkinkan komputer menjalankan tugas tertentu tanpa instruksi eksplisit. Sistem bekerja dengan mengenali pola dan membuat inferensi dari data yang tersedia[15]. Selain itu, machine learning termasuk cabang dari kecerdasan buatan yang memungkinkan sebuah sistem mempelajari data secara otomatis tanpa harus diprogram secara khusus.

3. METODE PENELITIAN

3.1. Jenis dan Sumber data

Penelitian ini menggunakan data kuantitatif yang berupa statistik pertandingan sepak bola FC Barcelona dalam kompetisi La Liga musim 2023/2024 sampai 2025/2026. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs penyedia statistik sepak bola, yaitu FBref. Dataset dibuat dalam format CSV atau Excel agar proses pengolahan data dengan menggunakan bahasa pemrograman Python menjadi lebih mudah.

Variabel yang digunakan dalam penelitian meliputi

- Penguasaan bola
- Jumlah tembakan
- Tembakan tepat sasaran
- Jumlah gol
- Jumlah kebobolan
- Akurasi umpan

3.2. Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini dilakukan melalui beberapa tahapan yang dijelaskan sebagai berikut.

a. Studi Literatur

Mengumpulkan referensi dari berbagai jurnal, buku, serta penelitian terdahulu yang berkaitan dengan data mining, clustering, machine learning, serta analisis performa dalam sepak bola.

b. Dokumentasi

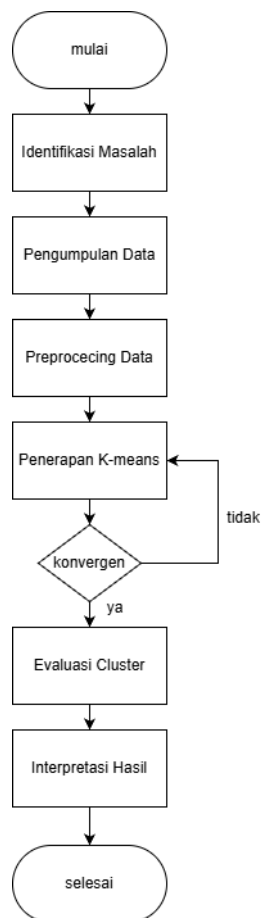
Mengumpulkan data statistik pertandingan FC Barcelona dari situs FBref dan kemudian menyusunnya dalam bentuk dataset.

c. Pengolahan Dataset

Dataset yang telah dikumpulkan kemudian melalui tahap praproses data, seperti pembersihan data, seleksi fitur, dan normalisasi sebelum dilakukan proses clustering.

3.3. Tahapan Penelitian

Proses penelitian dalam studi ini dilakukan secara teratur untuk mendapatkan hasil pengelompokan performa pertandingan FC Barcelona berdasarkan data statistik pertandingan. Tahapan penelitian yang dilakukan meliputi sebagai berikut



Gambar 1. Flochart alur penelitian

a. Identifikasi Masalah

Proses identifikasi masalah dilakukan untuk menetapkan fokus penelitian yang berkaitan dengan analisis kinerja pertandingan sepak bola. Analisis kinerja yang hanya menilai hasil pertandingan dianggap belum cukup untuk menggambarkan pola performa tim secara keseluruhan berdasarkan statistik pertandingan. Karena itu, penelitian ini memanfaatkan metode pengelompokan untuk mengidentifikasi pola kinerja berdasarkan data statistik pertandingan.

b. Pengumpulan Dataset

Proses pengumpulan dataset dilakukan dengan mengumpulkan data statistik pertandingan FC Barcelona di kompetisi La Liga untuk musim 2023/2024 sampai dengan musim 2025/2026 dari situs FBref. Dataset dibuat dalam format CSV atau Excel dan mencakup variabel seperti penguasaan bola, jumlah tembakan, tembakan yang tepat sasaran, jumlah gol, jumlah kebobolan, akurasi umpan, jumlah tendangan sudut, serta jumlah pelanggaran.

c. Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan dataset sebelum proses pengelompokan dilakukan. Pada tahap ini dilakukan pembersihan data guna menghilangkan data kosong maupun duplikat, seleksi fitur untuk menentukan atribut yang relevan, serta normalisasi data agar skala antarvariabel menjadi seragam sehingga proses clustering dapat berjalan lebih optimal.

d. Penerapan K-means Clustering

Metode K-Means Clustering diterapkan untuk mengelompokkan data berdasarkan kemiripan statistik pertandingan. Proses pengelompokan dilakukan dengan cara menetapkan jumlah kelompok, memilih titik pusat awal, menghitung jarak antara data dan titik pusat, lalu melakukan iterasi sampai kelompok-kelompok tersebut stabil. Perhitungan jarak menggunakan Euclidean Distance sebagai berikut:

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

Keterangan

$d(x, y)$ = jarak antara data dan centroid,

x_i = nilai data ke- i ,

y_i = nilai centroid ke- i ,

n = jumlah atribut.

Perhitungan centroid baru dilakukan menggunakan rata-rata data pada setiap cluster dengan persamaan berikut:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (2)$$

Keterangan:

C_k = centroid cluster ke- k ,

n_k = jumlah data pada cluster ke- k ,

x_i = data anggota cluster.

e. Evaluasi Cluster

Evaluasi cluster dilakukan untuk menentukan jumlah cluster optimal menggunakan Elbow Method berdasarkan nilai Within Cluster Sum of Squares (WCSS).

$$WCSS = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - c_i)^2 \quad (3)$$

Keterangan:

$WCSS$ = total variasi dalam cluster,

k = jumlah cluster,

x_j = data ke- j ,

c_i = centroid cluster,

S_i = himpunan data pada cluster ke- i .

Selain Elbow Method, evaluasi kualitas cluster juga dilakukan menggunakan Silhouette Score dengan persamaan berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

Keterangan:

$a(i)$ = rata-rata jarak data dalam cluster yang sama,

$b(i)$ = rata-rata jarak data ke cluster terdekat,

$s(i)$ = nilai silhouette

f. Interpretasi Hasil

Tahap terakhir adalah menganalisis hasil pengelompokan untuk memahami karakteristik setiap cluster berdasarkan statistik pertandingan. Hasil dari pengelompokan kemudian dianalisis

untuk menunjukkan pola performa pertandingan FC Barcelona selama satu musim di kompetisi La Liga, yang dibagi menjadi kategori performa stabil, agresif, dan dominan.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang dipakai dalam penelitian ini adalah data statistik pertandingan FC Barcelona di kompetisi La Liga untuk musim 2023/2024 hingga 2025/2026 yang diambil dari situs FBref. Dataset ini disusun dalam format CSV dan mencakup 110 data pertandingan yang meliputi berbagai variabel statistik, seperti penguasaan bola, jumlah tembakan, tembakan yang tepat sasaran, total umpan, akurasi umpan, jumlah gol, jumlah kebobolan, serta hasil dari setiap pertandingan.

Data statistik yang telah dikumpulkan selanjutnya digunakan sebagai landasan untuk menganalisis performa pertandingan FC Barcelona dengan menggunakan metode K-Means Clustering.

Tabel 1. Data match Barcelona

Matc hday	Lawan	Penguasaan Bola(%)	Temba kan	Tembakan Tepat Sasaran	Jumlah Umpan	Akurasi Umpan(%)	Jumlah Gol	Jumlah Kebobolan	Hasil
1	getafe	75	14	4	729	90	0	0	seri
2	cadiz	75	23	10	613	91	2	0	menang
3	villareal	61	21	11	592	91	4	3	menang
4	osasuna	70	9	4	635	87	2	1	menang

Berdasarkan Tabel 1, terlihat bahwa dataset yang digunakan dalam penelitian ini mencakup statistik pertandingan FC Barcelona di kompetisi La Liga untuk musim 2024/2025. Dataset yang digunakan berisi total 114 data pertandingan yang diambil dari situs FBref dan disusun dalam format CSV. Setiap data pertandingan memiliki beberapa atribut statistik yang mencakup penguasaan bola, jumlah tembakan, tembakan tepat sasaran, jumlah umpan, akurasi umpan, jumlah gol, jumlah kebobolan, serta hasil pertandingan.

Data dalam tabel menunjukkan variasi kinerja pertandingan yang didasarkan pada nilai statistik yang diperoleh di setiap matchday. Sebagai contoh, pertandingan melawan Real Betis mencatatkan jumlah gol terbanyak, yaitu 5 gol, dengan penguasaan bola mencapai 65% dan akurasi umpan sebesar 91%. Sementara itu, pertandingan melawan Getafe berakhir imbang dengan skor 0-0, di mana kedua tim tidak mencetak gol maupun kebobolan. Perbedaan nilai statistik di setiap pertandingan menunjukkan bahwa penampilan tim memiliki ciri-ciri yang berbeda-beda sepanjang kompetisi.

Dataset yang terdiri dari 114 pertandingan tersebut selanjutnya digunakan sebagai input dalam proses preprocessing dan pengelompokan menggunakan metode K-Means Clustering. Metode ini digunakan untuk mengelompokkan performa pertandingan berdasarkan tingkat kemiripan data statistik permainan.

4.2. Hasil Preprocessing Data

Tahap preprocessing dilakukan sebelum proses clustering guna meningkatkan kualitas data yang akan dianalisis. Pada tahap ini dilakukan pembersihan data untuk memastikan tidak terdapat data kosong maupun data yang tidak sesuai. Selain itu, dilakukan juga seleksi fitur untuk menentukan variabel yang dianggap penting dalam analisis performa pertandingan.

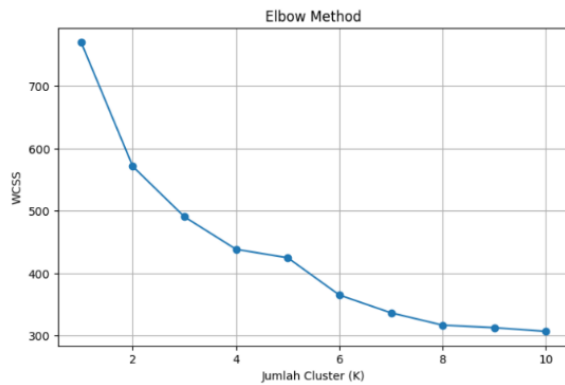
4.3. Hasil Penentuan Jumlah Cluster

Penentuan jumlah cluster optimal dilakukan menggunakan Metode Elbow berdasarkan nilai Within Cluster Sum of Squares (WCSS). Hasil pengujian menunjukkan bahwa titik elbow berada pada nilai $k=3$, sehingga jumlah cluster yang paling sesuai untuk digunakan dalam penelitian ini adalah tiga cluster.

Tabel 2. Data jumlah cluster

Cluster	Karakteristik
Cluster 0	Agresif
Cluster 1	Stabil
Cluster 2	Dominan

Menurut nilai centroid, setiap cluster menunjukkan ciri-ciri performa yang berbeda. Kluster dominan memiliki penguasaan bola dan jumlah tembakan yang lebih tinggi dibandingkan dengan kluster lainnya.



Gambar 2. Hasil elbow method

Berdasarkan Gambar 2, terlihat adanya penurunan nilai Within Cluster Sum of Squares (WCSS) yang cukup signifikan dari cluster pertama hingga cluster ketiga. Setelah mencapai nilai $k=3$, penurunan WCSS cenderung melambat dan membentuk titik siku (elbow). Kondisi tersebut menunjukkan bahwa jumlah cluster yang paling optimal dalam penelitian ini adalah tiga cluster.

Pemakaian tiga cluster dianggap mampu memberikan pemisahan data yang cukup baik

berdasarkan kesamaan statistik pertandingan FC Barcelona. Karena itu, dalam penelitian ini, proses pengelompokan menggunakan tiga kategori kinerja, yaitu agresif, stabil, dan dominan.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

Hasil pengujian memperoleh nilai Silhouette Score sebesar 0,209. Nilai tersebut menunjukkan bahwa hasil pengelompokan termasuk dalam kategori cukup baik karena masing-masing cluster masih memiliki karakteristik yang dapat dibedakan satu sama lain, meskipun ada beberapa data yang mirip di antara kelompok-kelompok tersebut.

4.4. Hasil Clustering

Hasil dari pengelompokan dengan metode K-Means Clustering menunjukkan tiga kategori dalam performa pertandingan, yaitu agresif, stabil, dan dominan. Pengelompokan dilakukan berdasarkan kemiripan statistik pertandingan, seperti penguasaan bola, jumlah tembakan, tembakan tepat sasaran, jumlah gol, dan jumlah kebobolan.

Tabel 3. Hasil Clustering

Matc hday	Lawan	Penguas aan Bola(%)	Temb akan	Tembakan Tepat Sasaran	Jumlah Umpan	Akurasi Umpan (%)	Jumlah Gol	Jumlah Kebobo lan	Hasil	Clust er	Kategori_ Performa
1	getafe	75	14	4	729	90	0	0	seri	01.00	stabil
2	cadiz	75	23	10	613	91	2	0	menang	02.00	dominan
3	villareal	61	21	11	592	91	4	3	menang	00.00	agresif
4	osasuna	70	9	4	635	87	2	1	menang	01.00	stabil

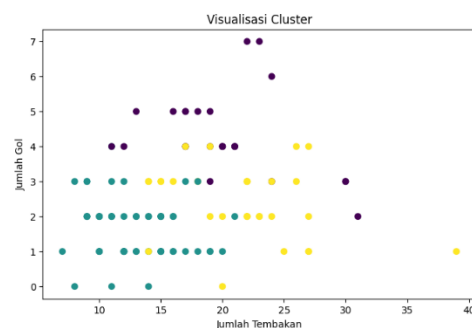
Hasil dari pengelompokan dengan metode K-Means Clustering menunjukkan tiga kategori dalam performa pertandingan, yaitu agresif, stabil, dan dominan. Pengelompokan dilakukan berdasarkan kesamaan statistik pertandingan seperti penguasaan bola, jumlah tembakan, tembakan yang tepat sasaran, jumlah gol, dan jumlah gol yang kebobolan.

Pertandingan melawan Getafe di matchday pertama bisa dianggap stabil karena statistik permainan menunjukkan keseimbangan yang baik dengan penguasaan bola yang tinggi, meskipun tidak ada gol yang tercipta. Pertandingan melawan Cadiz dapat dianggap sebagai penampilan yang dominan karena tim memiliki banyak tembakan dan penguasaan bola yang besar, ditambah dengan hasil kemenangan. Sementara itu, pertandingan melawan Villareal tergolong agresif karena terdapat banyak tembakan yang mengarah ke gawang dan jumlah gol yang tinggi, meskipun tim juga kebobolan cukup banyak.

Hasil dari pengelompokan menunjukkan bahwa setiap pertandingan memiliki ciri khas performa yang bervariasi, berdasarkan statistik permainan yang dikumpulkan selama pertandingan berlangsung.

4.5. Visualisasi Hasil cluster

Visualisasi hasil pengelompokan dilakukan untuk memahami sebaran data pertandingan sesuai dengan kelompok yang sudah terbentuk. Dalam penelitian ini, kami menggunakan visualisasi dalam bentuk grafik scatter plot dengan dua variabel utama, yaitu jumlah tembakan dan jumlah gol. Kedua variabel ini cukup menggambarkan kinerja ofensif tim.



Gambar 3. Visualisasi cluster

Visualisasi pada gambar tersebut menunjukkan hasil pengelompokan data pertandingan menggunakan metode K-Means Clustering berdasarkan dua variabel, yaitu jumlah tembakan pada sumbu X dan jumlah gol

pada sumbu Y. Setiap titik pada grafik merepresentasikan satu pertandingan, sedangkan perbedaan warna menunjukkan kategori cluster yang terbentuk.

Berdasarkan grafik tersebut, dapat dijelaskan sebagai berikut

a. Cluster dominan (warna kuning)

Terlihat berada di lokasi dengan jumlah tembakan yang cukup tinggi, yaitu sekitar 20 hingga 40, sementara jumlah gol yang dicetak bervariasi. Hal ini menunjukkan bahwa tim memiliki tingkat intensitas serangan yang tinggi dan dapat mengendalikan permainan, meskipun tidak selalu menghasilkan banyak gol.

b. Cluster agresif (warna ungu)

Cenderung mencetak banyak gol, bisa mencapai 7 gol, dengan jumlah tembakan yang juga cukup banyak. Cluster ini menunjukkan pertandingan dengan gaya permainan menyerang yang sangat efisien, meskipun di beberapa situasi juga terdapat risiko kebobolan.

c. Cluster stabil (warna hijau kebiruan)

Terlihat berkumpul di area dengan jumlah tembakan dan gol yang lebih rendah hingga sedang. Kluster ini menunjukkan kinerja yang lebih stabil dan seimbang, tetapi tidak terlalu agresif dalam mencetak gol.

Secara keseluruhan, visualisasi tersebut menunjukkan bahwa metode K-Means mampu mengelompokkan data pertandingan berdasarkan pola performa yang berbeda. Meskipun masih terdapat beberapa titik yang posisinya berdekatan antarcluster, perbedaan pola antar kelompok tetap dapat terlihat dengan cukup jelas. Hal ini sejalan dengan nilai Silhouette Score sebesar 0,209 yang menunjukkan bahwa kualitas hasil clustering berada pada kategori cukup baik.

4.6. Pembahasan Hasil Penelitian

Hasil penelitian menunjukkan bahwa metode K-Means Clustering mampu mengelompokkan data pertandingan FC Barcelona ke dalam tiga kategori performa, yaitu agresif, stabil, dan dominan, berdasarkan kemiripan statistik pertandingan. Proses pengelompokan tersebut menggunakan beberapa variabel, seperti penguasaan bola, jumlah tembakan, tembakan tepat sasaran, total gol yang dicetak, serta jumlah kebobolan.

Berdasarkan rata-rata setiap kelompok, terlihat perbedaan karakteristik yang cukup jelas antara kelompok-kelompok tersebut. Kluster agresif memiliki rata-rata jumlah tembakan (20,08) dan gol (4,32) yang cukup tinggi dibandingkan dengan kluster lainnya, tetapi juga disertai dengan jumlah kebobolan yang lebih banyak, yaitu (1,24). Hal ini menunjukkan bahwa cara bermain dalam kelompok ini cenderung menyerang dengan intensitas tinggi, tetapi belum sepenuhnya diimbangi dengan pertahanan yang kuat.

Kluster dominan menunjukkan persentase penguasaan bola yang tertinggi, yaitu 74,93%, serta jumlah tembakan yang juga tinggi, yaitu 21,70. Namun, rata-rata jumlah gol (2,40) tidak setinggi cluster yang agresif. Ini menunjukkan bahwa dominasi dalam permainan lebih terlihat dari penguasaan bola dan kontrol atas pertandingan, bukan hanya dari seberapa efektif penyelesaian akhir yang dilakukan.

Sementara itu, cluster yang stabil menunjukkan nilai rata-rata yang cenderung lebih rendah pada banyak variabel, seperti jumlah tembakan yang mencapai 13,07 dan jumlah gol yang hanya 1,63, serta jumlah kebobolan yang tergolong kecil, yaitu 0,89. Ini menunjukkan bahwa pertandingan di cluster ini memiliki pola permainan yang lebih seimbang, dengan intensitas serangan yang tidak terlalu tinggi.

Distribusi jumlah data menunjukkan bahwa kategori cluster stabil adalah yang paling mendominasi dengan persentase mencapai 50%, diikuti oleh cluster dominan yang memiliki persentase 27,27%, dan cluster agresif dengan persentase 22,73%. Ini menunjukkan bahwa mayoritas pertandingan FC Barcelona selama periode penelitian termasuk dalam kategori performa yang cukup stabil.

Visualisasi hasil clustering juga menunjukkan adanya perbedaan pola antarcluster, meskipun masih terdapat beberapa titik data yang saling berdekatan. Nilai Silhouette Score sebesar 0,209 menunjukkan bahwa kualitas cluster berada pada kategori cukup baik. Hal tersebut menandakan bahwa struktur pengelompokan telah terbentuk dengan cukup jelas, walaupun masih terdapat beberapa data yang memiliki karakteristik serupa.

Secara umum, hasil penelitian ini menunjukkan bahwa metode K-Means Clustering dapat digunakan untuk mengidentifikasi pola performa pertandingan berdasarkan data statistik yang tersedia. Pengelompokan yang dibuat memberikan pemahaman tentang perbedaan karakteristik permainan tim, sehingga bisa dijadikan landasan untuk menganalisis kinerja dengan cara yang lebih teratur.

5. KESIMPULAN DAN SARAN

Hasil penelitian memperlihatkan bahwa metode K-Means Clustering berhasil digunakan untuk mengelompokkan pola performa pertandingan FC Barcelona ke dalam tiga kelompok, yaitu agresif, stabil, dan dominan, berdasarkan kemiripan statistik pertandingan. Dari hasil tersebut, sebagian besar pertandingan berada pada kategori stabil yang menunjukkan keseimbangan dalam pola permainan. Di sisi lain, kategori agresif dan dominan memperlihatkan perbedaan pada aspek intensitas serangan serta penguasaan bola. Nilai Silhouette Score sebesar 0,209 menunjukkan bahwa kualitas hasil clustering berada pada tingkat yang cukup baik. Hal ini menandakan bahwa metode K-Means mampu memberikan gambaran awal mengenai pola performa tim, walaupun masih terdapat beberapa data yang memiliki kedekatan antarcluster. Untuk

pengembangan penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar serta menambahkan variabel statistik lain, seperti expected goals dan jumlah peluang tercipta. Selain itu, kombinasi metode clustering dengan teknik analisis lainnya juga berpotensi meningkatkan kualitas dan ketepatan hasil analisis.

DAFTAR PUSTAKA

- [1] A. A. Bausad, A. Pujiati, And D. Lukitasari, "Analisis Manajemen Sekolah Sepak Bola: Perspektif Sepak Bola Sebagai Industri Olahraga Football School Management Analysis: Perspective On Football As A Sports Industry," *Jambura Journal Of Sports Coaching*, Vol. 6, No. 2, 2024, Doi: 10.37311/Jjsc.V6i2.25014.
- [2] A. Ma'ruf, "Peran Teknologi Dalam Olah Raga Sepakbola," *Healing: Health Education, Advocacy, And Learning*, 2024, Doi: 10.12345/Novara.V1i2.48.
- [3] B. I. N. K. M. L. K. M. R. A. F. A. A. K. D. T. W. K. S. M Fathan A Syahputra, "Perbandingan Akurasi Algoritma Naïve Bayes Dan Decision Tree Dalam Prediksi Pemenang Pertandingan Pes 2021," *Jati(Jurnal Mahasiswa Teknik Informatika)*, 2025, Doi: 10.36040/Jati.V9i4.13941.
- [4] A. F. Ari Yanto And G. Testiana, "Implementasi Metode Klasifikasi Naïve Bayes Untuk Memprediksi Juara La Liga," *Jurnal Teknologi Sistem Informasi*, Vol. 5, No. 2, Pp. 128–139, Sep. 2024, Doi: 10.35957/Jtsi.V5i2.8028.
- [5] E. Triawan, N. Suarna, And A. Rinaldi Dikananda, "Klasifikasi Tipe Penyerang Sepak Bola Liga Inggris Berdasarkan Data Statistik Pemain Menggunakan Metode Naive Bayes," 2024.
- [6] S. Fauzi Prameswari And N. Latifah Dwi Mutiara Sari, "Literatur Review: Implementasi Metode Dan Dataset Pada Data Mining," *Seminar Nasional Informatika-Fti Upgris*, Vol. 2, 2024, Accessed: Apr. 14, 2026. [Online]. Available: <https://conference.upgris.ac.id/index.php/infe-st/article/view/6480>
- [7] Sekar Setyaningtyas, B. Indarmawan Nugroho, And Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, Vol. 10, No. 2, Pp. 52–61, Oct. 2022, Doi: 10.21063/Jtif.2022.V10.2.52-61.
- [8] M. Sahira, A. Salsabila, S. Salsabila, A. Najibah Putri, K. Ditha Tania, And W. Kurnia Sari, "Penerapan Metode K-Means Clustering Untuk Segmentasi Performa Pembalap F1 Season 2024," *Jurnal Buffer Informatika*, Vol. 11, No. 2, 2025, [Online]. Available: <https://journal.fkom.uniku.ac.id/index.php/buffer>
- [9] F. Dwi Agustina, M. Arif, S. Ahmad, P. Studi Sistem Dan Teknologi Informasi, And F. Sains Dan Teknologi, "Systematic Literature Review Atas Kinerja Algoritma Knn, Naïve Bayes, Dan Decision Tree Pada Berbagai Studi Prediksi Dan Klasifikasi," 2025.
- [10] M. Eric Adyatma Et Al., "Analisis Kontribusi Pemain Sepak Bola Eropa Berbasis C4.5," *Jurnal Ilmu Komputer Dan Informatika*, Vol. 2, No. 3, 2026, [Online]. Available: <https://jurnal.globalscients.com/index.php/jiki.998>
- [11] A. Putra Pratama Et Al., "Systematic Literature Review: Penerapan K-Means Clustering," 2025. Doi: 10.23969/Jp.V10i03.29785.
- [12] F. Khalish, N. M. Piranti, And O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering Dengan Metode K-Means," 2025. Doi: 10.54371/Jiip.V8i5.7874.
- [13] D. Restu Adji, E. Lutfina, B. Eka Ferdianto, E. Prashanti, And K. Amelia Putri Anwarri, "Metode Dan Algoritma Dalam Data Clustering: Systematic Literature Review," *Science, Technology And Management Journal*, 2025, Doi: 10.53416/Stmj.V5i1.326.
- [14] S. Ganiyah Ramadhani, D. Yuliana, N. Safrudin, And A. Budiman, "Tinjauan Sistematis Penerapan Klaster K-Means Dan Metode Elbow Untuk Analisis Data Stunting," 2026.
- [15] R. S. Nurhalizah, R. Ardianto, And P. Purwono, "Analisis Supervised Dan Unsupervised Learning Pada Machine Learning: Systematic Literature Review," *Jurnal Ilmu Komputer Dan Informatika*, Vol. 4, No. 1, Pp. 61–72, Aug. 2024, Doi: 10.54082/Jiki.168