Analisis Pemain Sepak Bola Profesional dengan Kerja Sama Tim Terbaik di Kompetisi Resmi Indonesia Menggunakan Metode K-Means Klastering

Wayan Ardike
Program Studi Statistika, Fakultas Sains dan Teknologi,
Universitas Terbuka, Bandar Lampung, Indonesia
Email: ardiwayan470@gmail.com

Abstrak

Klastering dengan algoritma K-Means menjadi teknologi analisis data yang handal dalam mengelompokkan pemain sepak bola. Metode ini kebanyakan digunakan dalam pengelompokan pemain berdasarkan aspek individual. Sangat jarang ditemukan penggunaan K-Means dalam pengelompokan pemain berdasarkan kerja sama dalam tim. Padahal kerja sama adalah fundamental dasar dalam sepak bola, dengan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan pemain sepak bola berdasarkan variabel yang mencerminkan kerja sama tim, khususnya di kompetisi resmi Indonesia. Diharapkan penelitian ini bermanfaat dalam strategi rekrutmen pemain sepak bola yang lebih efektif dan efisien. Tahapan klastering dimulai dari pengumpulan data, pembersihan, pembangunan model, evaluasi, validasi hingga visualisasi dan analisis klaster. Hasil yang diperoleh menunjukkan model K-Means dengan evaluasi Elbow mampu mengelompokkan pemain dengan optimal ke dalam 2 klaster, serta validasi dengan Silhouette Score mendapatkan nilai 0,7739 menunjukkan bahwa 77,39% data sudah berada pada klaster yang tepat. Sehingga dapat diketahui perbedaan karakteristik masing-masing klaster dengan cara perbandingan, dari hasil tersebut menunjukkan bahwa klaster 1 merupakan kelompok pemain dengan kerja sama tim terbaik karena mayoritas variabelnya memiliki nilai rata-rata yang tinggi. Sehingga strategi rekrutmen yang efektif dan efisien dapat dilakukan dengan memprioritaskan daftar pemain yang berada di klaster 1.

Kata kunci: analisis, klastering, K-Means, kerja sama, sepak bola.

1 PENDAHULUAN

Teknologi analisis data berkembang sangat cepat dari waktu ke waktu, dengan penggunaannya meliputi berbagai bidang termasuk olahraga sepak bola. Data statistik pemain seperti jumlah gol, assist, penyelamatan, tekel sukses, hingga akurasi umpan menjadi pondasi penting bagi analisis data dalam mengevaluasi performa pemain secara objektif (Wardhana & Winarno, 2024). Teknologi analisis data memiliki peran penting dalam sepak bola, penerapannya dapat membantu strategi pertandingan hingga manajemen tim. Salah satu strategi yang bisa diterapkan adalah proses seleksi pemain sepak bola untuk menciptakan kerja sama tim yang efektif dan efisien. Seleksi pemain berbasis data sangat penting untuk dilakukan terutama berdasarkan kerja sama tim, dikarenakan proses pengambilan keputusan dalam strategi sepak bola masih sering bias dan bersifat subjektif atau individu karena referensi pelatih hingga keterbatasan dalam interpretasi data yang kompleks.

Salah satu teknologi analisis data yang populer digunakan dalam seleksi atau rekrutmen pemain yaitu klastering. K-Means menjadi metode klastering yang dapat digunakan dalam pengelompokan berdasarkan kesamaan yang dimiliki (Mulaab, 2021), seperti posisi pemain dan performa lari dalam pertandingan (de Haan et al., 2025). Berbagai penelitian sebelumnya

telah menggunakan K-Means untuk menemukan pola tersembunyi dalam data statistik pemain, seperti pengelompokan pemain berdasarkan performa dan atribut fisik (Yuris et al., 2025), seleksi pemain yang tepat berdasarkan harga dan riwayat performa (Prasetyo et al., 2023), hingga sistem pendukung penyusunan line up pemain yang sesuai (Nurzahputra et al., 2017). Berdasarkan hasil penelitian tersebut, terbukti bahwa K-Means dapat menjadi alat seleksi yang strategis, efektif dan efisien dalam menemukan pemain dengan kriteria yang sesuai.

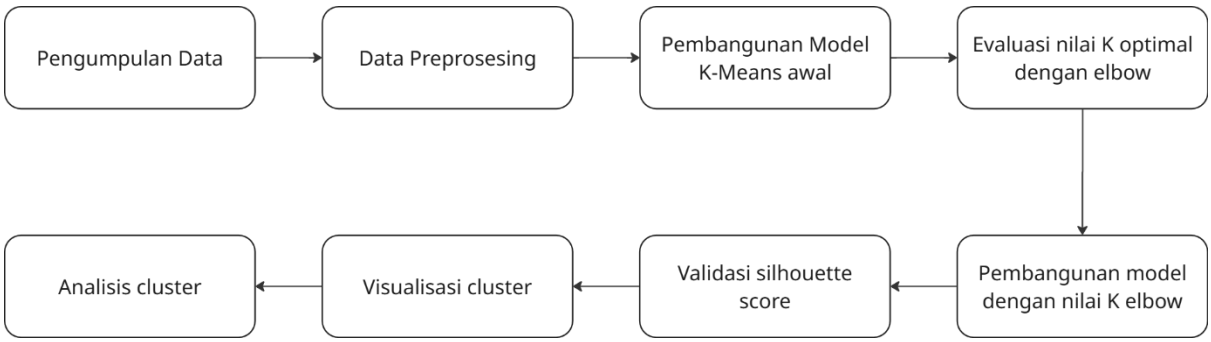
Walaupun sudah terbukti handal dalam pengelompokan pemain sepak bola, namun penggunaan K-Means pada penelitian terdahulu kebanyakan berfokus pada pemain dengan skill individu seperti performa, posisi hingga harga transfer yang dilakukan oleh oleh (de Haan et al., 2025; Nurzahputra et al., 2017; Yuris et al., 2025). Metode serupa masih sangat terbatas digunakan dalam pengelompokan pemain berdasarkan kerja sama tim, khususnya pemain profesional yang berkompetisi di liga resmi Indonesia. Faktanya dalam pertandingan sepak bola sebuah kemenangan tidak bergantung pada skill individu saja melainkan oleh kemampuan pemain dalam bekerja sama secara efektif dalam sebuah tim (Muslimin, 2020). Dengan begitu penerapan metode K-Means klastering sangat diperlukan dalam menyeleksi pemain dengan kerja sama tim yang bagus dikarenakan kerja sama merupakan fundamental dasar dalam pertandingan sepak bola (Putro et al., 2025).

Dengan kondisi tersebut penelitian ini dilakukan untuk mengelompokkan pemain profesional yang berkompetisi resmi di Indonesia dengan metode K-Means klastering. Tujuan utamanya untuk mengidentifikasi pola dan karakteristik dalam menentukan kelompok pemain terbaik berdasarkan indikator kerja sama tim. Penelitian ini juga diharapkan dapat menjadi acuan objektif dalam proses rekrutmen sekaligus penyusunan strategi permainan berbasis data yang dapat meningkatkan kualitas tim dan kompetisi itu sendiri kemudian pada bidang akademik diharapkan bisa menjadi referensi dalam penelitian lanjutan terkait penggunaan metode machine learning dalam analisis pemain sepak bola.

2 METODE

2.1 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah analisis klastering dengan menggunakan algoritma K-Means dalam pengelompokan pemain berdasarkan atribut yang dimiliki. Proses klastering ini dilakukan dalam beberapa tahapan utama yaitu mulai dari proses pengumpulan data dilanjutkan pembersihan dan pembangunan model kemudian dilakukan evaluasi dan validasi hingga pada tahapan visualisasi serta analisis klaster (Jollyta et al., 2021; Mulaab, 2021). Proses metodologi pada penelitian ini ditunjukkan pada Gambar 1 beserta penjelasannya.



Gambar 1. Tahapan utama dalam proses klastering meliputi pengumpulan, pembersihan, pembangunan model, evaluasi dan validasi hingga hasil akhir berupa klaster.

Berikut penjelasan langkah-langkah metodologi penelitian pada Gambar 1:

- Pengumpulan data menjadi tahapan awal dalam metodologi ini untuk mengumpulkan data yang dibutuhkan dalam proses klastering.
- Supaya data dapat dianalisis dan menghasilkan output yang optimal dapat menerapkan tahapan pra-prosesing lebih lanjut terhadap data dengan nilai kosong maupun outlier serta data dengan skala yang berbeda.
- Parameter K penting ditentukan saat pembangunan model serta menjadi langkah awal dalam menentukan jumlah klaster.
- Evaluasi parameter K dengan metode Elbow untuk mendapatkan jumlah klaster yang optimal.
- Silhouette Score digunakan pada saat memvalidasi model yang berguna untuk memastikan kejelasan dan konsistensi pemisah untuk setiap klasternya.
- Sebaran dan distribusi data harus dapat diamati dengan jelas sehingga diperlukan tahapan visualisasi klaster.
- Analisis klaster bertujuan membandingkan karakteristik setiap kelompok untuk menemukan perbedaan serta mencari yang lebih unggul.

2.2 Mengumpulkan Data

2.2.1 Populasi Dan Sampel

Populasi merupakan wilayah generalisasi yang terdiri atas obyek dengan karakteristik tertentu untuk dipelajari atau diteliti (Santosa Ir., 2021). Wilayah generalisasi yang digunakan dalam penelitian ini adalah seluruh pemain sepak bola yang terdaftar dalam kompetisi resmi Indonesia yaitu Indonesia Super League pada musim 2025/2026.

Kemudian sampel merupakan bagian dari populasi dengan karakteristik yang sama untuk diteliti atau dipelajari (Santosa Ir., 2021). Sampel pada penelitian ini adalah pemain sepak bola yang berkompetisi pada divisi Indonesia Super League, sampel diambil dengan metode purposive sampling yang merupakan teknik sampling berdasarkan kriteria dari jumlah populasi yang diketahui (Santosa Ir., 2021).

Sampel untuk penelitian ini berasal dari populasi yang berjumlah 576 pemain yang terdaftar dalam kompetisi Indonesia Super League pada musim 2025/2026 dengan kriteria sebagai berikut, yaitu:

- Terdaftar secara resmi sebagai pemain di salah satu klub kompetisi
- Memiliki jumlah penampilan setidaknya 1 kali dalam kompetisi
- Memiliki catatan statistik yang diperlukan sebagai variabel penelitian

2.2.2 Teknik Pengumpulan Data

Pengumpulan data pada platform Sofascore dilakukan dengan pemilahan manual sesuai kriteria dan diorganisir satu persatu. Data yang dikumpulkan dari profil pemain disimpan dalam satu dataset, yang berisikan variabel dasar kerja sama tim dari setiap jenis posisi pemain, mulai dari posisi kiper, bek, gelandang, sayap hingga striker (Avianto, 2020). Variabel yang dikumpulkan meliputi jumlah pertandingan, assist, key passes, passing accuracy, Successsful dribble, tackles, interceptions, aerial duels won hingga error leading to goal.

2.2.3 Variabel

Total terdapat 2 variabel utama yaitu nama pemain digunakan untuk mengidentifikasi pemain dan variabel kerja sama tim yang terdiri dari 9 variabel dengan masing-masing mempunyai fungsi sendiri sebagai penentu kualitas kerja sama pemain dalam tim, variabel kerja sama tim dapat dilihat pada Tabel 1 beserta deskripsi dan fungsinya.

Tabel 1. Deskripsi dan fungsi variabel kerja sama tim

Variabel	Deskripsi	Fungsi
MP	Match Played	Jumlah penampilan dalam kompetisi
AST	Assists	Jumlah assist yang diciptakan
KEYP	Key Passes	Jumlah peluang yang diciptakan
APS	Accuracy Passing	Akurasi distribusi bola
SDR	Success Dribble	Kemampuan membuka ruang menggiring bola
TACK	Tackles	Jumlah kontribusi bertahan
INT	Interceptions	Jumlah pemotongan aliran lawan
ADW	Aerial duels won	Jumlah kemenangan duel diudara
ELTG	Error leading to goal	Jumlah penalti yang dari kesalahan pemain

2.3 Persiapan Data

2.3.1 Data Hilang Atau Kosong

Data yang hilang atau kosong dapat sangat mempengaruhi performa model dalam melakukan klastering sehingga sangat diperlukan data dengan keadaan lengkap atau terisi sepenuhnya kemudian data dengan nilai kosong bisa diatasi dengan dihapus (Mulaab, 2021).

2.3.2 Data Duplikat

Kondisi data terduplikasi yaitu saat beberapa data mempunyai nilai sama persis yang bisa berdampak cukup serius terhadap akurasi model kemudian untuk mengatasi permasalahan ini bisa dilakukan dengan penghapusan data yang sama tersebut.

2.3.3 Data Outlier

Selain itu data dengan sejumlah nilai yang menyimpang jauh atau outlier juga menjadi permasalahan serius yang dapat terlihat dengan visualisasi boxplot atau histogram namun dapat diperbaiki dengan transformasi seperti metode Z-score normalization.

$$Z = \frac{x - \mu}{\sigma}$$

dengan :
$Z =$ banyaknya sampel nilai z
$X =$ nilai data individu
$\mu =$ nilai rata-rata
$\sigma =$ nilai standar deviasi

Penyelesaian data outlier dapat dilakukan dengan melakukan transformasi data menggunakan beberapa metode seperti box-cox hingga yeo-johnson untuk membuat distribusi data menjadi normal.

2.3.4 Menemukan variabel yang redundan

Variabel yang redundan menandakan adanya pengaruh kuat antara variabel satu dengan variabel lainnya. Variabel yang redundan harus dihindari dengan mendeteksi lebih dini, untuk variabel dengan nilai numerik dapat menggunakan koefisien korelasi dan kovarian (Mulaab, 2021). Karena semua variabel yang digunakan untuk klastering penelitian ini bertipe data numerik kontinu maka dapat menggunakan korelasi Pearson.

$$r_{xy} = \frac{n\sum x_i\ y_i - (\sum x_i\ )(\sum y_i\ )}{\sqrt{(n\sum x_i^2 - (\sum x_i\ )^2)(n\sum y_i^2 - (\sum y_i\ )^2)}}$$

dengan:
r_{xy} = korelasi antara x dan y
n = banyaknya sampel
x_i = nilai x ke-i
y_i = nilai y ke-i

2.3.5 Normalisasi Data

Normalisasi data merupakan proses untuk mengubah satuan ukur dari variabel yang berbeda menjadi ukuran yang sama dalam bentuk skala atau range (Mulaab, 2021). Dengan data yang sudah dinormalisasi maka bobotnya akan sama sehingga tidak akan terlalu signifikan dalam mempengaruhi analisa model klastering.

$$f(x) = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Dengan:
x_i = data ke-i
x_{min} = data dengan nilai minimum
x_{max} = data dengan nilai maksimum

2.4 Pembangunan Model K-Means

K-Means merupakan salah satu algoritma klastering dengan metode partisi (partitioning method) yang berbasis titik pusat (centroid) selain algoritma k-Medoids yang berbasis obyek (Irwansyah & Faisal, 2015). K-Means sendiri salah satu metode yang sering digunakan karena mudah untuk diimplementasikan (Mulaab, 2021). Cara kerja dari K-Means sendiri adalah dengan membentuk klaster yang semirip mungkin di dalam satu kelompok, sambil memastikan bahwa perbedaan antar klaster sebesar mungkin atau kesamaan antar klaster serendah mungkin (Yuris et al., 2025). Apabila ditulis dalam bentuk rumus matematika, bentuk algoritma K-Means adalah sebagai berikut:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{n} \left(x_{ij} - c_{kj}\right)^2}$$

dengan :
$d(x_i, c_k)$ = jarak antara titik data ke-i dan pusat klaster (centroid)ke-k
x_i = vektor fitur dari data ke-i
c_k = vektor centroid dari klaster ke-k
x_{ij} = nilai fitur ke-j dari data ke-i
c_{kj} = nilai fitur ke-j dari centroid klaster ke-k
n = jumlah fitur (variabel) yang digunakan dalam analisis

Berikut beberapa tahapan dalam menggunakan algoritma K-Means dalam klastering data dijelaskan sebagai berikut (Mulaab, 2021):

a. Menentukan jumlah klaster awal yang ingin dihasilkan, biasanya jumlahnya lebih kecil dari jumlah keseluruhan baris data.
b. Menentukan centroid atau pusat awal klaster, langkah ini dapat dilakukan secara acak atau secara manual dengan mengambil beberapa titik data.
c. Menghitung jarak dari setiap data yang dihitung dari jarak pusat klaster dengan menggunakan rumus Euclidean distance, sebagai berikut:

$$d_{p,q} = \sqrt{\sum_{k=1}^{n}(p_k - q_k)}$$

Dengan :
$d =$ jarak objek antara x dan y
$n =$ dimensi data
$p_k =$ objek p ke k
$q_k =$ objek q ke k

d. Membandingkan hasil jarak setiap klaster, dan memilih jarak yang terdekat
e. Menghitung kembali untuk menentukan centroid baru untuk setiap klaster dengan menghitung rata-rata nilai vektor, dengan rumus sebagai berikut:

$$C_k = \frac{1}{k}\sum d_i$$

Dengan:
$C_k =$ centroid pada klaster ke k
$n_k =$ jumlah dokumen dalam satu klaster
$d_i =$ jumlah dari nilai jarak yang masuk dalam masing-masing klaster

f. Lakukan langkah b dan c sampai tidak ada perubahan yang terjadi pada centroid

2.5 Evaluasi Dengan Metode Elbow

Model K-Means yang dibangun perlu dilakukan evaluasi supaya menghasilkan klaster yang sesuai dengan data. Salah satunya adalah mengevaluasi nilai k awal, dapat dilakukan denga menggunakan metode Elbow. Metode ini merupakan optimasi klaster yang dihasilkan melalui perbandingan persentase antar klaster yang membentuk siku pada suatu titik (Jollyta et al., 2021). Perbandingan dilakukan dengan mencari nilai Sum of Square Error (SSE) dengan rumus sebagai berikut:

$$SSE = \sum_{k=1}^{k}\sum_{x_i es_k} \|x_i - c_k\|_2^2$$

Dengan :
$k =$ jumlah klaster
$s_k =$ himpunan data klaster ke-i
$x_i =$ titik data ke i yang berada di klaster ke-k
$c_k =$ centroid pusat dari klaster ke-k
$\|x_i - c_k\|_2^2 =$ jarak euclidean kuadrat

2.6 Validasi Dengan Silhouette Score (skor siluet)

Hasil klaster yang didapat perlu dilakukan evaluasi secara mendalam, sebab algoritma seperti K-Means bekerja secara supervisi dalam mengelompokkan data (Hasan, 2024). Maka diperlukan serangkaian validasi klaster yang dapat memberikan pemahaman akurat dari hasil klastering. Salah satu metode yang bisa digunakan adalah Silhouette Score (skor siluet) yang memberikan ukuran seberapa baik setiap titik data ditempatkan dalam klaster mereka sendiri dibandingkan dengan klaster lain (Hasan, 2024). Silhouette Score (skor siluet) secara rumus matematis ditulis sebagai berikut (Mulaab, 2021):

$$Si = \frac{(bi - ai)}{max(ai, bi)}$$

dengan:
$Si =$ nilai silhouette coefficient untuk data ke-i.
$bi =$ rata-rata jarak terkecil antara data ke-i dan seluruh data di klaster lain
$ai =$ rata-rata jarak antara data ke-i dan semua data lain dalam klaster yang sama
$max(ai, bi) =$ digunakan sebagai pembagi agar nilai Si berada pada rentang standar

2.7 Visualisasi Klaster

Setelah terbentuknya klaster yang optimal melalui proses evaluasi dan validasi model K-Means maka selanjutnya adalah mempresentasikan klaster. Sejumlah klaster yang didapatkan divisualisasi menggunakan scatterplot untuk melihat sebaran atau distribusi dari masing-masing klaster.

2.8 Analisis Klaster

Klaster yang berhasil terbentuk dianalisis lebih lanjut untuk melihat ciri-ciri atau karakteristik. Hal ini dilakukan untuk menemukan pembeda serta pembanding untuk masing-masing klaster, dengan begitu dapat melihat dan menyeleksi kelompok pemain dengan karakteristik kerja sama tim yang bagus.

3 HASIL DAN PEMBAHASAN

3.1 Persiapan Data

Data yang digunakan untuk penelitian telah melewati berbagai pemrosesan data, dari jumlah awal 576 pemain profesional yang terdaftar dalam kompetisi Indonesia Super League, diperoleh 425 pemain yang memenuhi kriteria penelitian, berikut hasil analisis statistik deskriptif variabel kerja sama tim ditunjukkan pada Tabel 2.

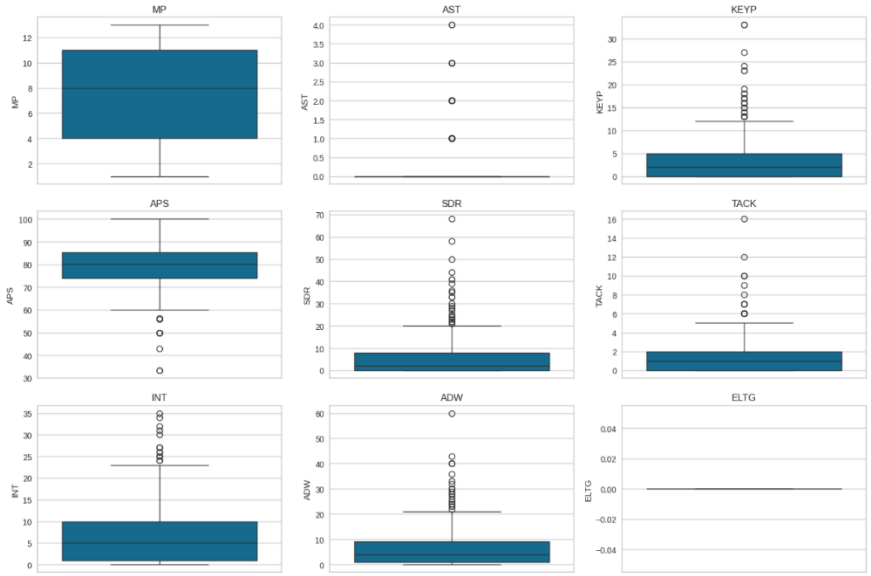
Tabel 2. Statistik deskriptif variabel kerja sama tim (contoh)

Variabel	Mean	Std.Dev	Min	Max
MP	7,4	3,6	1	13
AST	0,3	0,6	0	4
KEYP	3,5	5	0	33
APS	79,2	9,6	33	100
SDR	6	8,9	0	68
TACK	1,5	2,1	0	16
INT	6,8	7,2	0	35
ADW	6,5	8,1	0	60
ELTG	0	0	0	0

Berdasarkan hasil analisis statistik deskriptif tersebut diketahui bahwa mayoritas pemain memiliki kemampuan kerja sama yang cukup bervariasi, dengan rata-rata tiap variabel lebih besar dari nilai minimum. Sebagian besar pemain cukup aktif dalam mengikuti pertandingan ditunjukkan nilai rata-rata variabel MP sebesar 7.4 dan maksimumnya 13, mayoritas pemain memiliki kemampuan menciptakan peluang (KEYP) dan tackel (TACK) yang cukup baik, namun beberapa pemain lebih menonjol karena nilai maksimum yang jauh lebih besar. Akurasi passing (APS) memiliki rata-rata 79.2 dengan standar deviasi 9.6 menunjukkan bahwa terdapat variasi yang cukup besar kemampuan passing antar pemain, kemudian untuk variabel ELTG memiliki nilai 0 pada semua pemain, menunjukkan bahwa tidak terdapat pemain yang melakukan kesalahan dalam menciptakan gol.

3.2 Transformasi Data

Pada tahapan pemeriksaan kualitas data menggunakan visualisasi box plot ditemukan bahwa sejumlah variabel memiliki nilai outlier yang cukup ekstrim, seperti pada variabel Tackles, Accuracy Passing, Interceptions dan beberapa variabel lain yang dapat dilihat pada Gambar 2 berikut.



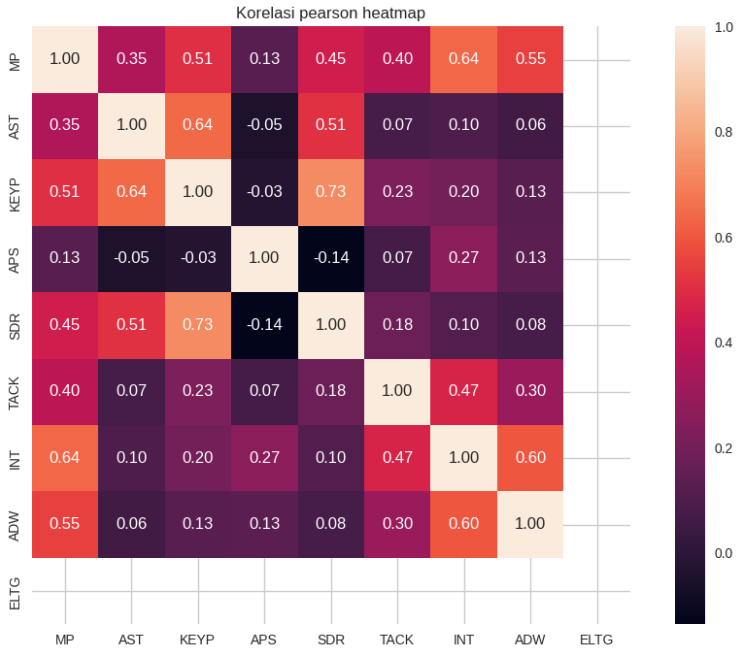
Gambar 2. Box plot data outlier

Data dengan nilai outlier kemudian dilakukan transformasi menggunakan metode yeo-johnson, metode ini dipilih karena rentang data dimulai dari 0, transformasi bertujuan untuk membuat distribusi data menjadi normal, serta melakukan standarisasi supaya semua variabel memiliki skala nilai yang sama.

3.3 Penggabungan Variabel

Beberapa variabel dalam datasets ternyata mengalami redundan atau mempunyai korelasi yang kuat dengan variabel lainnya, seperti variabel Match Played dengan Aerial Duels Won, Assits dengan Key Passes, hingga Interceptions dengan Aerial Duels Won. Selain beberapa variabel tersebut masih ada variabel lain yang mengalami korelasi dapat dilihat pada visualisasi korelasi pearson dengan tabel heatmap pada Gambar 3.

Solusi untuk mengatasi masalah korelasi tersebut dapat diselesaikan dengan menggabungkan beberapa variabel. Penggabungan variabel menggunakan metode UMAP (Uniform Manifold Approximation and Projection), variabel awal yang berjumlah 9 kemudian digabungkan menjadi 3 variabel baru.



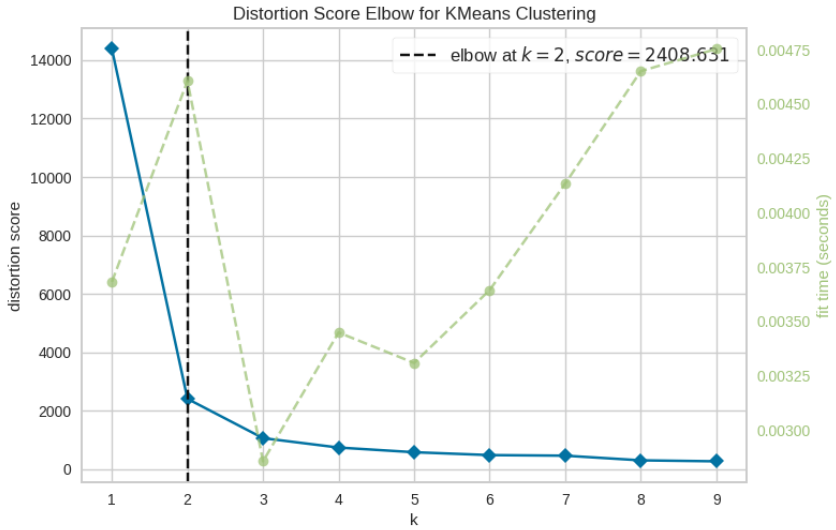
Gambar 3. Heatmap korelasi pearson

3.4 Pembangunan Dan Validasi Model K-Means Awal

Setelah data sudah dipastikan layak maka dapat dilakukan pembangunan model K-Means awal, pada model ini nilai K atau jumlah klaster ditentukan secara manual. Nilai K awal yang ditetapkan yaitu 3 artinya jumlah klaster yang akan terbentuk pada model awal ini adalah 3 klaster. Hasil validasi model K-Means dengan 3 klaster menghasilkan Silhouette Score sebesar 0.58, hasil tersebut menunjukkan bahwa baru 58% data berada pada klaster yang tepat, hasil tersebut masih belum menunjukkan kualitas klaster yang bagus.

3.5 Evaluasi Dengan Metode Elbow

Pada model awal ditetapkan 3 klaster atau kelompok pemain berdasarkan variabel kerja sama, model kemudian dilakukan evaluasi untuk menemukan nilai K yang optimal dengan metode Elbow, berikut grafik evaluasi metode Elbow pada Gambar 4.



Gambar 4. Grafik evaluasi nilai K dengan metode Elbow

Hasil evaluasi menunjukkan nilai K yang optimal adalah 2, yang lebih kecil dari nilai sebelumnya. Ini menunjukkan bahwa jumlah klaster yang lebih kecil memberikan pemisahan serta pembeda yang lebih jelas dan representatif antar klaster. Maka dilakukan pembangunan ulang model K-Means dengan nilai K = 2.

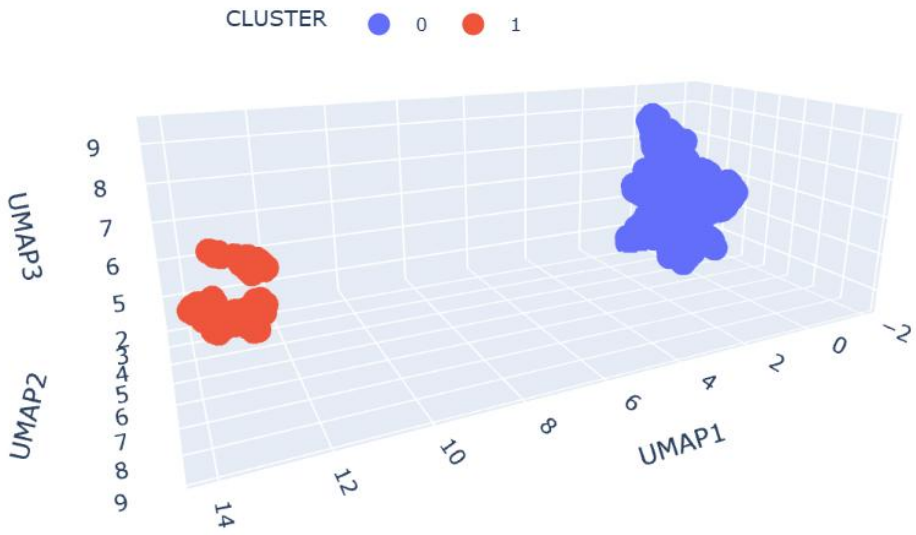
3.6 Validasi Akhir Model K-Means

Pembangunan model hasil evaluasi dengan metode Elbow kemudian dilakukan validasi menggunakan Silhouette Score. Tujuan dari validasi ini adalah untuk melihat tingkat kesesuaian data antar klaster yang terbentuk. Hasil validasi Silhouette Score yang didapatkan pada model ini sebesar 0.7739.

Nilai tersebut menunjukkan bahwa tingkat kesesuaian data antar klaster adalah 77.39%. Artinya bahwa 77.39% data sangat cocok dengan klasternya serta sudah cukup baik dalam menginterpretasikan karakteristik pemain dengan baik berdasarkan kemampuan kerja sama.

3.7 Visualisasi Klaster

Klaster yang dihasilkan kemudian di visualisasikan dengan scatterplot, karena jumlah variabel UMAP atau gabungan yang terbentuk adalah 3 maka bentuk visualisasi yang digunakan adalah 3 dimensi, distribusi atau sebaran dari masing-masing klaster dapat dilihat pada Gambar 5.



Gambar 5: Distribusi klaster dengan 3d scatterplot

Hasil visualisasi tersebut menunjukkan grafik scatterplot dengan 2 warna yang berada pada ruang 3 dimensi, yaitu warna biru muda (klaster 0) dan warna merah tomat (klaster 1). Berdasarkan hasil tersebut dapat diketahui bahwa klaster 0 memiliki distribusi lebih besar dibandingkan klaster 1 yang menunjukkan bahwa sebagian besar pemain sepak bola berada di klaster 0.

3.8 Analisis Klaster

Masing-masing klaster memiliki ciri-ciri atau karakteristiknya sendiri, untuk melihat pembeda tersebut digunakan analisis deskriptif dalam menampilkan rata-rata nilai variabel pada tiap klasternya. Hasil analisis dapat dilihat pada Tabel 3 berikut.

Tabel 3. Tabel rata-rata kemampuan kerja sama

Variabel	Klaster 0	Klaster 1
MP	6,6	10
AST	0,0	1,4
KEYP	1,9	8,9
APS	79,5	78,2
SDR	3,6	14,1
TACK	1,4	1,7
INT	6,3	8,4
ADW	6,0	8,0
ELTG	0	0

Melalui Tabel 3 dapat diketahui bahwa klaster 0 menunjukkan sebagian besar pemain memiliki tingkat kemampuan kerja sama yang kurang bagus jika dibandingkan dengan klaster 1. Nilai rata-rata untuk variabel MP, KEYP, dan SDR pada kalster 0 lebih rendah yang menunjukkan bahwa pemain belum memiliki kemampuan yang optimal dalam kontribusi dengan anggota team, keterlibatan dalam kerja sama serta kemampuan dalam menyelesaikan masalah. Kemampuan interaksi yang rendah, seperti inisiatif dalam membantu tim, terlihat pada kecilnya nilai rata-rata variabel AST, TACK, INT, dan ADW.

Sebaliknya, pemain pada klaster 1 memiliki nilai rata-rata lebih tinggi hampir di semua variabel kerja sama tim. Karakteristik yang solid serta efektif pada klaster 1 tercermin dari tingginya nilai variabel MP, KEYP, dan SDR. Hal tersebut menandakan bahwa pemain lebih aktif dalam berkolaborasi, serta inisiatif dalam membantu. Selain itu, nilai rata-rata variabel INT dan ADW yang lebih tinggi mengindikasikan bahwa pemain lebih adaptif dan disiplin.

4 KESIMPULAN

Berdasarkan hasil dan pembahasan terhadap penggunaan K-Means dalam pengelompokan pemain berdasarkan variabel kerja sama tim diperoleh kesimpulan sebagai berikut:

- Klastering menggunakan algoritma K-Means mampu mengelompokan pemain sepak bola berdasarkan indikator kerja sama tim secara komprehensif. Evaluasi dengan metode Elbow membuat model secara optimal mengelompokan pemain kedalam 2 klaster, serta validasi dengan Silhouette Score menunjukan model K-Means dapat mengelompokkan pemain pada klaster yang sesuai dengan akurasi 77,39%.
- Dengan analisis nilai rata-rata yang berbeda sudah menunjukan bahwa masing-masing klaster memiliki pemain dengan karakteristik tersendiri. Klaster 1 memiliki nilai rata-rata yang cenderung lebih tinggi artinya pemain memiliki karakteristik kerja sama tim yang lebih baik dibandingkan pemain pada klaster 0.
- Dapat disimpulkan bahwa pemain yang berada pada klaster 1 bisa dijadikan acuan dalam pembentukan strategi, seperti penyusunan posisi pemain yang optimal, hingga rekrutment efektif dan efisien dengan memprioritaskan daftar pemain yang berada pada klaster 1.

DAFTAR PUSTAKA

Avianto, L. (2020). *Mengenal sepak bola*. PT Balai Pustaka (Persero). https://books.google.co.id/books?id=1Qh9DQAAQBAJ

de Haan, M., van der Zwaard, S., Sanders, J., Beek, P. J., & Jaspers, R. T. (2025). Beyond Playing Positions: Categorizing Soccer Players Based on Match-Specific Running Performance Using Machine Learning. *Journal of Sports Science and Medicine*, *24*(3), 565–577. https://doi.org/10.52082/jssm.2025.565

Hasan, Y. (2024). Pengukuran Silhouette Score Dan Davies-Bouldin Index Pada Hasil Klaster K-Means Dan Dbscan. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1), 60–74. https://doi.org/10.23960/jitet.v12i3s1.5001

Irwansyah, E., & Faisal, M. (2015). *Advanced Klastering: Teori dan Aplikasi*. DeePublish. https://books.google.co.id/books?id=8y80BgAAQBAJ

Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (2021). *Teknik Evaluasi Klaster Solusi Menggunakan Python Dan Rapidminer - Deny Jollyta , Muhammad Siddik , Herman Mawengkang , Syahril Efendi - Google Books*. Deepublish. https://books.google.co.id/books?hl=en&lr=&id=3rcgEAAAQBAJ&oi=fnd&pg=PP1&dq=D.+Jollyta,+M.+Siddik,+H.+Mawengkang,+dan+S.+Efendi,+Teknik+Evaluasi+Klaster+Solusi+Menggunakan+Python+Dan+Rapidminer.+Deepublish,+2021.+&ots=likSFeqa9F&sig=ivOBJJbQq3fkkDWcqSRvnp

Mulaab. (2021). *Data Mining: Konsep dan Aplikasi*. Media Nusa Creative (MNC Publishing). https://books.google.co.id/books?id=X1FKEAAAQBAJ

Muslimin, M. (2020). Meningkatkan Hasil Belajar PJOK Materi Operan dalam Permainan Sebak Bola dengan Tekhnik Give and Go Pada Siswa Kelas VI SDN Esot Kecamatan Pringgarata Tahun Pelajaran 2019 /2020. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, *4*(1). https://doi.org/10.36312/jisip.v4i1.1015

Nurzahputra, A., Pranata, A. R., & Puwinarko, A. (2017). Decision Support System for Football Players Lineup Selection using Fuzzy Multiple Attribute Decision Making and K-Means Klastering Methods. *Jurnal Teknologi Dan Sistem Komputer*, *5*(3), 106–109. https://doi.org/10.14710/jtsiskom.5.3.2017.106-109

Prasetyo, E., Priyatama, A. S., & Setyatama, F. (2023). Constellation of Football Players Determination Based on Cost and Performance History Using the K-Means Klastering. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, *14*(2), 194–205. https://doi.org/10.31849/digitalzone.v14i2.17106

Putro, G. C., Yudistrio, I. T., Bahtiar, M. A., & Windiarti, S. R. (2025). *Sepak bola Bukan Hanya Sekedar Tendang Bola*. Kramantara JS. https://books.google.co.id/books?id=JOhlEQAAQBAJ

Santosa Ir., P. I. (2021). *Metodologi Penelitian*. Universitas Terbuka. https://pustaka.ut.ac.id/lib/msim4312-metodologi-penelitian/

Wardhana, F. P., & Winarno, S. (2024). Analisis Pemain Terbaik Sepak Bola dengan menggunakan Algoritma K-Means. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 409–417. https://doi.org/10.29408/edumatic.v8i2.27105

Yuris, S., Azzami, A., Al Zami, F., Hadi, H. P., Irawan, C., Nurhindarto, A., & Sulistyono, M. T. (2025). Klastering and Profiling Analysis for FIFA Football Player using K-Means. *Jurnal Informatika: Jurnal Pengembangan IT*, *10*(1), 178–189. https://doi.org/10.30591/jpit.v9ix.xxx