

Aplikasi *K-Medoid* dalam Regenerasi Pemain Sepak Bola

WAHYU SA'DUN AKBAR¹, RAHMADI YOTENKA², ROHMATUL FAJRIYAH³

^{1, 2, 3}) Program Studi Statistika Fakultas MIPA Universitas Islam Indonesia, Indonesia

e-mail: 966110101@uii.ac.id

ABSTRAK

Sepakbola merupakan olahraga paling populer. Dalam permainan sepakbola, peran gelandang sangat penting sehingga setiap klub sepakbola terutama di Liga-Liga top Eropa perlu meregenerasi gelandang untuk kompetisi Liga di musim-musim berikutnya. Penelitian ini bertujuan untuk melakukan pencarian pemain tengah pengganti yang memiliki kesamaan karakteristik bermain menggunakan metode *K-Medoid Clustering*. Metode *Principal Component Analysis* diimplementasikan untuk mengatasi multikolinieritas dan untuk mereduksi variabel yang digunakan, sehingga didapatkan 2 komponen utama. Kedua komponen utama ini menjelaskan 72.9% variansi populasi yang ada. *K-Medoid Clustering* menghasilkan 2 kelompok dan berdasarkan pengukuran jarak dengan metode *Euclidean* didapatkan bahwa Tchouameni merupakan gelandang yang paling mirip dengan Casemiro. Hasil ini mendukung keputusan Real Madrid untuk membeli Tchouameni sebagai keputusan yang tepat.

Kata Kunci: Sepakbola, Gelandang, *K-Medoids Clustering*, *Principal Component Analysis*.

ABSTRACT

Football is the most popular sport. In the game of football, the role of midfielders is very important so that every football club, especially in the top European leagues, needs to regenerate midfielders for league competitions in the following seasons. This research aims to search for substitute midfielders who have similar playing characteristics using the K-Medoid Clustering method. The Principal Component Analysis method was implemented to overcome multicollinearity and to reduce the variables used, so that 2 main components were obtained. These two main components explain 72.9% of the population variance. K-Medoid Clustering produced 2 groups and based on distance measurements using the Euclidean method it was found that Tchouameni was the midfielder most similar to Casemiro. These results support Real Madrid's decision to buy Tchouameni as the right decision.

Keywords: Football, Midfielders, *K-Medoids Clustering*, *Principal Component Analysis*.

1. PENDAHULUAN

Sepakbola adalah olahraga paling diminati dan populer di dunia. Posisi gelandang sangat krusial, pemain dalam posisi ini terlibat dalam semua fase bermain dari bertahan, membangun permainan, hingga menyerang. Di lingkup benua Eropa sebagai kiblat sepakbola dunia, klub Real Madrid memiliki sejarah gemilang menghasilkan gelandang gelandang kelas dunia seperti Casemiro.

Casemiro di lini tengah real madrid sangat penting untuk menutup ruang yang ditinggalkan pemain tengah lain. Peranya sebagai gelandang bertahan dilakukan Casemiro dengan sangat baik, di Laliga pada musim 2017/2018, Casemiro mampu memenangkan 78 Tekel dari 108 percobaan, serta melakukan 56 kali *blocks* bola baik itu dari *shooting* maupun *passing*. Selain itu dalam distribusi bola Casemiro berperan sebagai distributor bola dari area bertahan ke area tengah maupun area menyerang. Peran tersebut dilakukan Casemiro dengan baik seperti contoh data statistik La Liga pada musim 2017/2018 Casemiro mampu melakukan umpan dengan akurasi sebesar 87.7%. Dengan performa dan kemampuan Casemiro tersebut membuat peranya di lini tengah sangat penting dalam Real Madrid menjuarai liga domestik maupun luar domestik.

Pada musim 2022/2023 Casemiro memutuskan untuk pergi dari Real Madrid dan bermain untuk klub asal Inggris yaitu Manchester United. Sehingga sebelum Casemiro pergi, Real Madrid membeli Aurélien Tchouaméni dari AS Monaco. Aurélien Tchouaméni merupakan gelandang muda berusia 22 tahun. Dengan usianya yang masih muda tentunya konsistensi dan kualitas Aurélien Tchouaméni masih diragukan untuk tim sekelas Real Madrid.

Lantas muncul pertanyaan gelandang seperti apa yang memenuhi kriteria Casemiro sebagai gelandang, serta apa yang mendasari Real Madrid melakukan pembelian Aurélien Tchouaméni untuk menggantikan Casemiro sebagai gelandang bertahan di Real Madrid dan apakah ada alternatif pemain lain yang cocok menggantikan Casemiro.

Dalam penelitian ini *K-Medoid* digunakan untuk mengelompokkan jenis-jenis gelandang yang ada di liga top Eropa berdasarkan kesamaan atau kedekatan antara titik-titik data di dalamnya. Setelah mengelompokkan pemain gelandang dan mengetahui kelompok kategori dari Casemiro maka dilakukan perhitungan kembali Jarak *Euclidean* semua pemain dengan Casemiro untuk melihat pemain dengan jarak paling dekat. Jarak *Euclidean* adalah ukuran jarak yang paling umum digunakan dalam analisis data dan pengukuran geometri.

2. METODE PENELITIAN

2.1 Sepak Bola

Sepakbola adalah sebuah olahraga dimana dua tim bertanding untuk memasukkan bola ke gawang lawan. Setiap tim terdiri dari sebelas pemain, sehingga disebut kesebelasan, di mana dalam permainan Sepak bola, diperlukan kemampuan dasar seperti menendang bola, mengontrol bola, menggiring bola, menyundul bola, gerakan tipu, merebut bola, melempar bola ke dalam lapangan, dan teknik penjaga gawang.

Dalam upaya mencapai tujuan mencetak gol, permainan sepak bola terdiri dari dua bagian, yaitu menyerang dan bertahan. Menyerang adalah situasi di mana tim berupaya untuk mencetak gol ke gawang lawan, sementara bertahan adalah situasi di mana tim berusaha untuk mencegah tim lawan mencetak gol. Pada saat menyerang, tim berusaha untuk menciptakan peluang melalui pemberian umpan dan tembakan, sedangkan pada saat bertahan, tim berusaha untuk mencegah peluang dengan cara memotong umpan dan menghalangi tembakan.

2.2 Gelandang

Dalam tim sepak bola, posisi pemain tengah atau gelandang kerap dijuluki sebagai "otak" atau "pengatur siasat". Tugas mereka adalah untuk melakukan *passing* bola secara cepat ke wilayah lawan dengan cara menendang bola ke tengah lapangan. Posisi gelandang salah satu posisi yang memegang peranan penting dalam mengatur strategi permainan. Mereka harus memiliki kemampuan berpikir cepat dan mengambil keputusan yang tepat. Selain itu, gelandang juga bertanggung jawab dalam mengatur serangan tim. Salah satu tugas penting gelandang adalah melakukan operan bola secara cepat dan tepat ke daerah pertahanan lawan, serta memberikan umpan kepada rekan satu tim.

2.3 Statistika Deskriptif

Statistika deskriptif adalah metode yang digunakan untuk mengumpulkan, merangkum, dan mempresentasikan data dalam bentuk yang mudah dipahami dan berguna untuk analisis selanjutnya. Dalam statistika deskriptif, data dijabarkan dan diilustrasikan dengan berbagai cara agar dapat memberikan informasi yang bermakna. Salah satu tujuan utama dari statistika deskriptif adalah mengorganisir data secara sistematis sehingga mudah dianalisis dan dipahami. Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu data sehingga memberikan informasi yang berguna. Salah satu ukuran pemusatan data yang biasa digunakan adalah *mean*.

Grafik merupakan alat yang kuat dalam statistika deskriptif karena dapat menyajikan informasi secara visual, sehingga memudahkan pemahaman dan interpretasi data. Dengan menggunakan grafik, kita dapat melihat pola, tren, perbandingan, dan karakteristik data dengan lebih jelas.

2.4 Multikolinieritas

Multikolinieritas terjadi ketika dua atau lebih variabel independen dalam suatu model regresi memiliki korelasi linear tinggi antara satu sama lain. Selain itu, multikolinieritas juga dapat mengurangi efisiensi dan keandalan perkiraan koefisien regresi. Standar *error* menjadi lebih besar, yang dapat menyebabkan interval kepercayaan yang lebih lebar dan mengurangi kepastian dalam kesimpulan statistik.

Salah satu metode untuk mengidentifikasi adanya multikolinieritas adalah dengan menggunakan koefisien korelasi antara variabel-variabel independen. Koefisien korelasi merupakan sebuah indikator yang digunakan untuk mengukur tingkat hubungan antara variabel-variabel tersebut. Jika nilai koefisien korelasi antara variabel melebihi 0,8, dapat disimpulkan bahwa terdapat multikolinieritas antara variabel-variabel tersebut

$$r = \frac{n \sum_{i=1}^n (x_i y_i) - (\sum_{i=1}^n x_i \sum_{i=1}^n y_i)}{((n \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i)^2 ((n \sum_{i=1}^n y_i^2 - \sum_{i=1}^n y_i)^2)^{1/2}}$$

- r = nilai koefisien korelasi
- x_i = nilai x ke- i
- y_i = nilai y ke- i
- n = banyaknya data

Selain menggunakan koefisien korelasi, deteksi multikolinieritas juga dapat dilakukan dengan menggunakan nilai VIF (*Variance Inflation Factor*).

$$VIF = \frac{1}{(1 - R_j^2)}$$

- VIF = nilai VIF (*Variance Inflation Factor*)
- R_j^2 = Koefisien determinasi variabel bebas ke- j dengan variabel lain
- j = 1,2, ... k

2.5 Principal Component Analysis

Principal Component Analysis (PCA) merupakan metode analisis yang berguna untuk mengurangi dimensi data pada dataset yang besar dan kompleks. Dalam analisis PCA, arah-arah utama atau komponen-komponen utama dari data diidentifikasi berdasarkan variabilitas dalam dataset. Komponen-komponen utama ini kemudian dapat dimanfaatkan untuk mentransformasi data ke dalam ruang baru dengan dimensi yang lebih rendah (Suhail, Rajpoot, & Markham, 2011).

PCA memiliki 3 tahapan utama yaitu standarisasi data, menghitung matriks kovariansi, dan menghitung nilai dan *vector eigen*.

1. Standarisasi data. Standarisasi data dilakukan untuk memastikan bahwa setiap variabel memiliki skala yang sama dan memiliki pengaruh yang seimbang pada hasil analisis PCA berikut rumus standarisasi data menggunakan *Z-Score*.

$$x_{baru} = \frac{x_i - \bar{x}}{\sigma}$$

- x_i = nilai x ke i
- \bar{x} = rata rata x
- σ = standar deviasi

2. Menghitung matriks kovariansi, Berikut ini rumus untuk menghitung matriks kovarian.

$$Cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

- x_i = nilai x ke i
- \bar{x} = rata-rata x
- \bar{y} = nilai rata-rata y
- y_i = nilai y ke i
- n = banyak data

3. Menghitung nilai eigen. Nilai eigen merepresentasikan variansi yang dapat dijelaskan oleh komponen utama yang terbentuk. Nilai eigen ini nantinya juga akan berperan dalam pemilihan banyaknya dimensi (K). Rumus menghitung nilai eigen adalah sebagai berikut ini $(A - \lambda I)x = 0$
4. Setelah nilai eigen terbentuk langkah selanjutnya adalah menghitung *vector* eigen dan memilih nilai K.
5. Menghitung nilai *score* K komponen utama yang akan digunakan pada analisis *K-Medoid Clustering*.

2.6 Jarak Euclidean

Euclidean distance adalah perhitungan untuk mengukur jarak dua titik dalam *Euclidean space* yang mempelajari hubungan antara sudut dan jarak. Dalam matematika *Euclidean distance* digunakan untuk mengukur dua titik dalam satu dimensi yang memberikan hasil seperti perhitungan pythagoras (Miftahuddin, Umaroh, & Karim, 2020).

Jarak *Euclidean*, juga dikenal sebagai jarak *Euclidean* atau jarak, adalah metrik yang digunakan untuk mengukur jarak antara dua titik dalam ruang *Euclidean*. Konsep ini didasarkan pada geometri *Euclidean*, yang merupakan sistem geometri yang umum digunakan dalam matematika.

$$d_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

- x_i = nilai x ke i
- x_j = nilai x ke j
- y_i = nilai y ke i
- y_j = nilai y ke j

2.7 K-Medoid Clustering

K-Medoid merupakan suatu algoritma dalam pengelompokan data yang bertujuan untuk memisahkan data berdasarkan kesamaan di antara mereka. Metode *K-Medoid* merupakan metode yang efisien untuk mengelompokkan objek berdasarkan kesamaan multivariat, terutama ketika data memiliki beberapa variabel atau atribut yang saling berhubungan (Mardia, et al., 1979). Tujuan dari algoritma ini adalah untuk mengurangi suatu fungsi objektif yang menghitung jarak antara data dan pusat kluster yang telah ditetapkan (Kaufman & Rousseeuw, 1990).

Algoritma *K-Medoid* adalah bentuk modifikasi dari algoritma *K-Means* yang menggunakan *medoid* sebagai pusat kluster, bukan rata-rata seperti pada algoritma *K-Means*. *Medoid* adalah titik data di dalam kluster yang memiliki jarak rata-rata terdekat dengan semua titik data lain di dalam kluster. Dengan menggunakan *medoid* sebagai pusat kluster, algoritma *K-Medoid* lebih tahan terhadap data-data yang berbeda jauh atau *outlier* dibandingkan dengan algoritma *K-Means* (Park & Jun, 2009).

Dalam setiap iterasinya, algoritma *K-Medoid* akan memilih *k-medoid* awal secara acak dari dataset yang akan dikelompokkan. Selanjutnya, setiap data akan ditempatkan ke dalam kluster yang memiliki *medoid* terdekat dengan data tersebut. Untuk setiap kluster, *medoid* baru akan dihitung dengan menggunakan teknik *exhaustive search*. Iterasi tersebut akan diulang terus-menerus sampai tidak terjadi lagi perubahan dalam penempatan data ke dalam kluster. Dalam akhir proses, algoritma *K-Medoid* akan menghasilkan k kluster dengan *medoid* masing-masing kluster (Park & Jun, 2009).

Menurut (Kaufman & Rousseeuw, 1990) dan (Park & Jun, 2009) untuk menerapkan algoritma *K-Medoid*, langkah-langkah yang harus dilakukan adalah sebagai berikut:

1. Tentukan jumlah k (jumlah kluster yang diinginkan) dan pilih k data acak sebagai *medoid* awal.
2. Hitung jarak total antara setiap data dengan *medoid* dalam kluster tersebut.

$$\sum_{i=1}^n d_{im}$$

d_{im} = jarak *Euclidean* antara data ke- i dan *medoid* nya.

3. Pilih data yang memiliki jarak total terkecil sebagai *medoid* baru dalam setiap kluster.

4. Ulangi langkah 2-3 hingga tidak ada perubahan dalam penempatan data ke dalam kluster atau iterasi telah mencapai batas maksimum yang ditentukan sebelumnya.

Algoritma *K-Medoid* menghasilkan k kluster dengan masing-masing kluster memiliki satu *medoid*.

2.8 Populasi Penelitian

Populasi penelitian ini menggunakan data pemain gelandang dari liga sepakbola di Eropa pada musim 2021/2022. Terdapat 5 liga di Eropa yang digunakan sebagai sampel pada penelitian ini, diantaranya *La Liga* (Liga Spanyol), *Premiere League* (Liga Inggris), *Serie A* (Liga Italia), *Bundesliga* (Liga Jerman), dan *Ligue 1* (Liga Prancis). Sampel data yang digunakan dalam penelitian ini berjumlah 1216.

2.9 Data dan Sumber Data

Data pada penelitian ini adalah data sekunder, dimana data pada penelitian ini diambil dari website <https://fbref.com/>. Website tersebut menyediakan data beraneka macam yang berkaitan dengan olahraga salah satunya sepak bola. Data yang diunduh berupa 1216 gelandang sepakbola yang bermain di *La Liga* (Liga Spanyol), *Premiere League* (Liga Inggris), *Serie A* (Liga Italia), *Bundesliga* (Liga Jerman), dan *Ligue 1* (Liga Prancis).

2.10 Variabel Penelitian

Pada Penelitian ini, terdapat beberapa variabel yang digunakan, berikut ini adalah daftar variabel beserta definisi variabel yang digunakan dalam penelitian ini.

Tabel 1. Keterangan Variabel

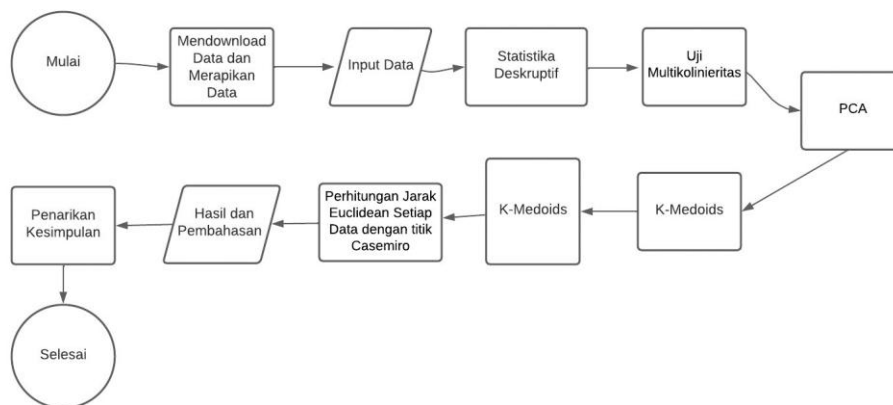
Variabel	Definisi
<i>Player</i>	Nama pemain
<i>Nation</i>	Kebangsaan Pemain
<i>Pos (Position)</i>	Posisi
<i>Squad</i>	Klub bermain pemain
<i>Age</i>	Umur Pemain pada musim 2021/2022
<i>Cmp(Pass Completion)</i>	Presentase keakuratan umpan dalam satu musim
<i>PrgDist(Progressive Distance)</i>	Jumlah umpan progresif dalam satu musim
<i>Ast(Asist)</i>	Jumlah asist pemain dalam satu musim
<i>Tkl.Int (Tekel & Interceptions)</i>	Jumlah tekel dan intersep yang dilakukan pemain dalam satu musim
<i>Sh.90(Shoot/90minutes)</i>	Rata rata jumlah <i>shoot</i> yang dilakukan pemain dalam 90 menit
<i>Def.3rd(Defense 1/3)</i>	Jumlah sentuhan di <i>area</i> bertahan selama satu musim
<i>Mid.3rd(Middle 1/3)</i>	Jumlah sentuhan di <i>area</i> tengah lapangan selama satu musim
<i>Att.3rd(Attack 1/3)</i>	jumlah sentuhan di <i>area</i> menyerang selama satu musim

2.11 Tahapan Penelitian

Pada penelitian ini metode analisis yang digunakan adalah metode deskriptif dan metode *non hierarki* yaitu *K-Medoid*. Rincian tahapan penelitian yang dapat dilihat sebagai berikut.

1. Mendownload data dari website <https://fbref.com/>

2. Merapikan data.
3. *Input* data.
4. Melakukan analisis deskriptif.
5. Melakukan *Bartlett Test* untuk mengetahui gejala multikolinieritas.
6. Karena data mengalami masalah multikolinieritas maka dilakukan analisis *Principal Component Analysis (PCA)*.
7. Melakukan pengelompokan data menggunakan *K-Medoid Clustering*.
8. Melakukan penghitungan Jarak *Euclidean* seluruh data dengan titik Casemiro.
9. Pilih jarak yang paling dekat dengan Casemiro dan seleksi dengan kriteria yang ada.
10. Membandingkan performa Casemiro dengan pemain pengganti yang dipilih.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Statistika Deskriptif

Sebelum menganalisis data diharuskan untuk meneliti gambaran umum data yang akan dianalisis. Data yang dianalisis adalah data karakteristik pemain gelandang 5 liga Eropa. Berikut ini adalah data rata-rata performa tiap karakteristik pemain yang dianalisis dan performa Casemiro.

Tabel 2. Perbandingan Performa Rata-Rata dan Casemiro

Variabel	Rata Rata Gelandang Eropa	Casemiro
<i>Cmp</i>	75.79% per season	85.2% per season
<i>PrgDist</i>	2035.81 umpan per season	10039 umpan per season
<i>Ast</i>	1.424 asist per season	3 asist per season
<i>Tkl.Int</i>	31.48 tekel per season	117 tekel per season
<i>Sh.90</i>	1.491 shoot per match	1.44 shoot per match
<i>Def.3rd</i>	131.16 sentuhan per season	707 sentuhan per season
<i>Mid.3rd</i>	355.82 sentuhan per season	1300 sentuhan per season
<i>Att.3rd</i>	206.05 sentuhan per season	366 sentuhan per season

Tabel 2. diatas merupakan tabel perbandingan statistik Casemiro dengan rata-rata gelandang eropa pada musim 2021/2022. Dapat dilihat pada Tabel 2, Casemiro memiliki statistik

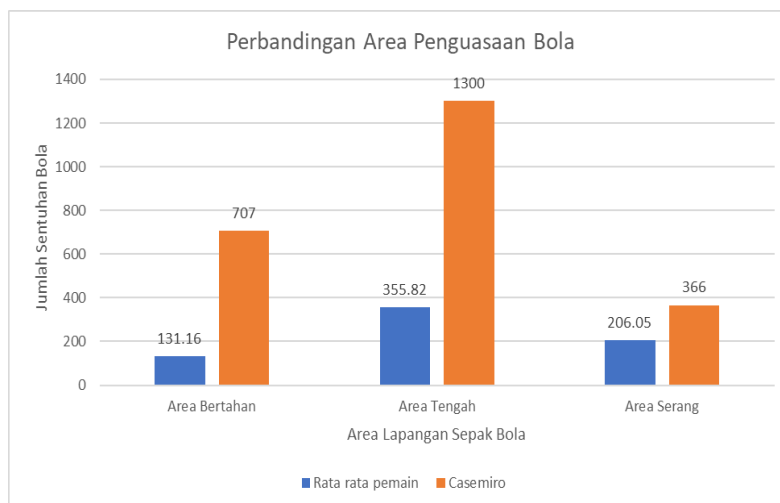
yang lebih unggul dibanding rata rata gelandang Eropa pada umumnya. Untuk lebih jelasnya berikut ini adalah visualisasi performa Casemiro pada musim 2021/2022.



Gambar 2. *Bar Chart* Perbandingan Performa Casemiro dan Rata-Rata Gelandang Eropa Musim 2021/2022

Pada Gambar 2. terlihat bahwa pada musim 2021/2022, Casemiro memiliki akurasi umpan(*Cmp*), umpan progresif(*PrgDist*), tekel dan intersep(*Tkl.Int*), Asist(*Ast*) diatas rata rata gelandang Eropa. Hanya saja *Shooting* per 90 menit Casemiro lebih rendah dibanding rata rata gelandang Eropa. Hal tersebut dikarenakan posisi Casemiro sebagai gelandang bertahan sehingga jarang melakukan percobaan *shooting*. Dapat dilihat pula visualisasi dibawah yang menggambarkan sentuhan Casemiro di *area* bertahan melebihi rata rata pemain eropa lainya.

Gambar 3. memvisualisasi perbandingan penguasaan bola Casemiro dengan rata rata gelandang eropa pada musim 2021/2022. Terlihat bahwa Casemiro cenderung lebih banyak melakukan penguasaan bola di *area* bertahan dibanding *area* menyerang. Itu menandakan posisi Casemiro sebagai gelandang bertahan yang cenderung melakukan built up serangan dan pergerakann di area pertahanan.



Gambar 3. *Bar Chart* Perbandingan Penguasaan Bola Casemiro dengan Rata Rata Gelandang Eropa Musim 2021/2022

3.2 Uji Multikolinearitas

Untuk mengidentifikasi adanya multikolinearitas, dapat dilakukan dengan menghitung nilai korelasi antar variabel atau menghitung nilai *Variance Inflation Factor* (VIF).

Tabel 3. Nilai VIF Variabel

Variabel	Nilai VIF
<i>PrgDist</i>	16.905
<i>Ast</i>	2.427
<i>Tkl.Int</i>	4.753
<i>Sh.90</i>	1.190
<i>Def.3rd</i>	9.370
<i>Mid.3rd</i>	14.744
<i>Att.3rd</i>	3.501

Berdasarkan Tabel 3. sebagai hasil pengujian didapatkan nilai VIF setiap variabel < 10 kecuali variabel *PrgDist* dan *Mid.3rd*, artinya terjadi multikolinearitas atau variabel *PrgDist* dan *Mid.3rd* terdapat korelasi dengan variabel lain. Salah satu pendekatan untuk mengatasi masalah ini adalah dengan mereduksi variabel menggunakan teknik seperti *Principal Component Analysis* (PCA).

Tabel 4. Tabel Standarisasi Data

No	Pemain	<i>Cmp</i>	<i>PrgDist</i>	<i>Ast</i>	<i>Tkl.Int</i>	<i>Mid.3rd</i>	<i>Att.3rd</i>
1	Paul Akouokou	1.183	-0.320	-0.681	-0.199	-0.294	-0.832
2	Tomas Alarcon	-0.632	0.166	-0.203	0.907	0.043	-0.220
...
1216	Edon Zhegrova	3.4E-05	-0.674	-0.203	-0.476	-0.805	0.026

Tabel 5. Matriks Varian-Kovarian

	<i>Cmp</i>	<i>PrgDist</i>	<i>Ast</i>	<i>Tkl.Int</i>	<i>Sh.90</i>	<i>Def.3rd</i>	<i>Mid.3rd</i>	<i>Att.3rd</i>
<i>Cmp</i>	1	0.310258	0.000728	0.194163	-0.14508	0.296964	0.343129	0.015871
<i>PrgDist</i>	0.310258	1	0.426894	0.770047	-0.16913	0.898177	0.951487	0.600019
<i>Ast</i>	0.000728	0.426894	1	0.296076	0.094828	0.226393	0.430973	0.747912
<i>Tkl.Int</i>	0.194163	0.770047	0.296076	1	-0.17157	0.841799	0.826278	0.517257
<i>Sh.90</i>	-0.14508	-0.16913	0.094828	-0.17157	1	-0.24313	-0.16333	0.091465
<i>Def.3rd</i>	0.296963	0.898177	0.226393	0.841799	-0.24313	1	0.871879	0.420179
<i>Mid.3rd</i>	0.343129	0.951487	0.430973	0.826278	-0.16333	0.871879	1	0.636031
<i>Att.3rd</i>	0.015871	0.600019	0.747912	0.517257	0.091465	0.420179	0.636031	1

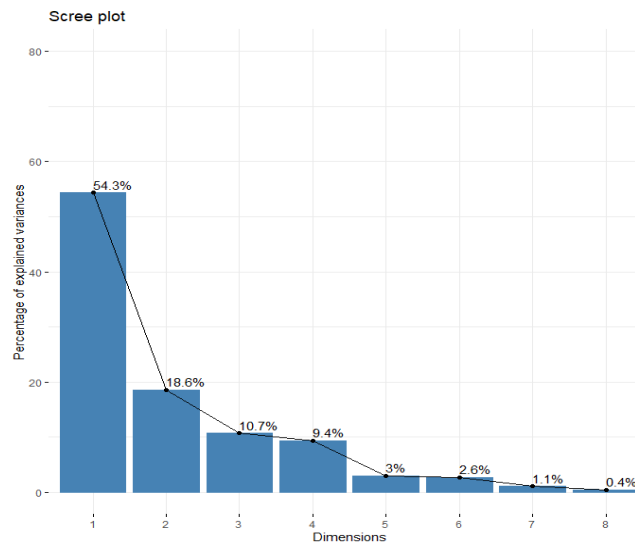
Tabel 6. Nilai Eigen

No	Nilai Eigen
1	4.3
2	1.5
3	0.9
4	0.8
5	0.2

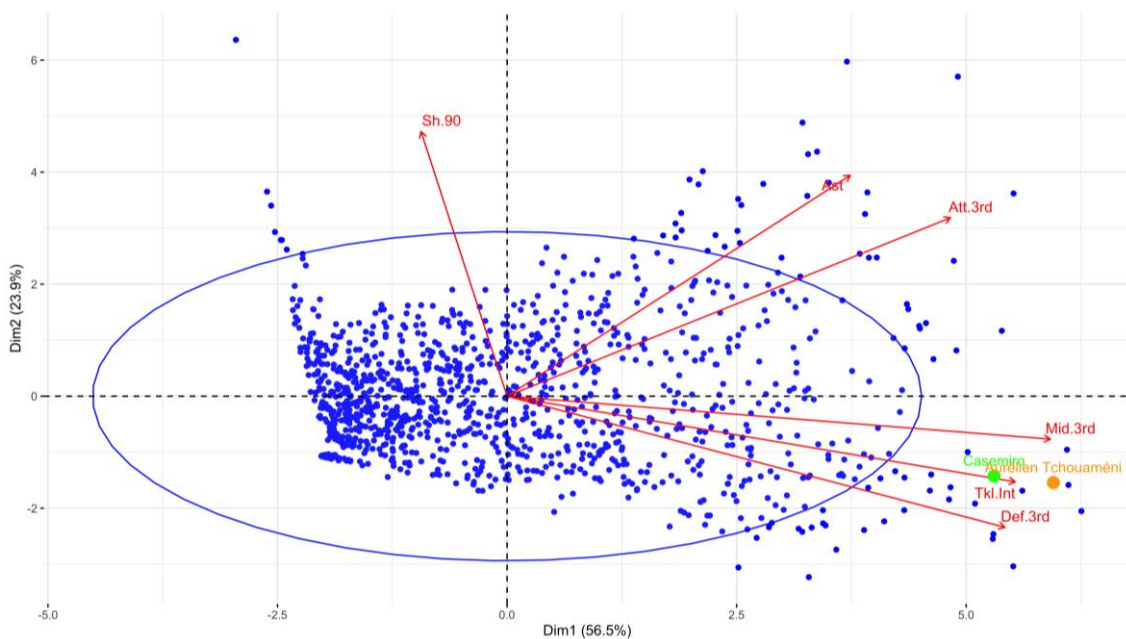
Tabel 5 merupakan tabel hasil perhitungan nilai varian-kovarian. Pada tabel tersebut, setiap sel menunjukkan kovarian antara pasangan variabel yang sesuai. Matriks kovarian simetris, dengan

diagonal utama yang terdiri dari varians masing-masing variabel, karena kovarian antara variabel dengan dirinya sendiri bernilai 1.

Langkah selanjutnya dalam penelitian ini adalah menghitung nilai *eigen* dari data untuk menentukan jumlah komponen utamanya. Proses perhitungan nilai *eigen* dilakukan menggunakan RStudio. Hasilnya dapat dilihat pada Tabel 6. Berdasarkan hasil hitung nilai *eigen* terdapat 2 nilai *eigen* yang melebihi 1, yang dapat disimpulkan bahwa jumlah komponen utama yang terbentuk sebanyak 2 faktor. Hal ini juga diperkuat dengan visualiasi *scree plot* pada Gambar 4. Pada Gambar 4, dengan menggunakan 2 faktor, variabilitas data nya 72.9% populasi.



Gambar 4. *Scree Plot* Nilai *Eigen*



Gambar 5. *Biplot* Variabel PCA

Berdasarkan Gambar 5, jelas bahwa PC1 (Dim1) menjelaskan 56,5 persen dari variasi keseluruhan data. Komponen PC1 memiliki dampak yang signifikan dari variabel seperti Def.3rd, Mid.3rd, dan Tkl.Int. Ini menunjukkan bahwa PC1 terutama mencerminkan kemampuan

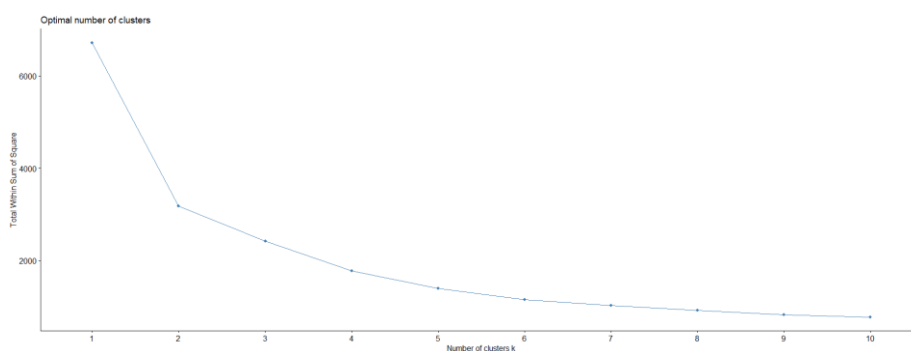
bertahan dan penguasaan lapangan tengah. Namun, PC2 (Dim2) bertanggung jawab atas 23,9 persen dari variasi keseluruhan data. Komponen PC2 dipengaruhi oleh faktor-faktor seperti Sh.90 dan Ast. Ini berarti bahwa PC2 mencerminkan kontribusi dalam penyerangan, terutama terkait dengan tembakan ke gawang dan assist.

Sesuai dengan Biplot ini, kriteria dalam menentukan pemain tipe Gelandang Bertahan harus didasarkan pada pemain yang memiliki nilai PC1 yang tinggi dan dekat dengan variabel seperti Def.3rd, Mid.3rd, dan Tkl.Int. Dalam kasus Gelandang Serang, pemain dengan nilai PC2 yang tinggi dan dekat dengan variabel seperti Sh.90, Ast, dan Att.3rd adalah yang harus dipilih. Untuk Gelandang Tengah Seimbang, pemain yang berada di tengah-tengah antara PC1 dan PC2, yang menunjukkan kemampuan bertahan dan menyerang yang seimbang, harus dipilih. Pada Gambar 5., jelas terlihat bahwa Aurelien Tchouameni dan Casemiro berada di kuadran yang sama, menunjukkan bahwa mereka berdua termasuk dalam kategori gelandang bertahan.

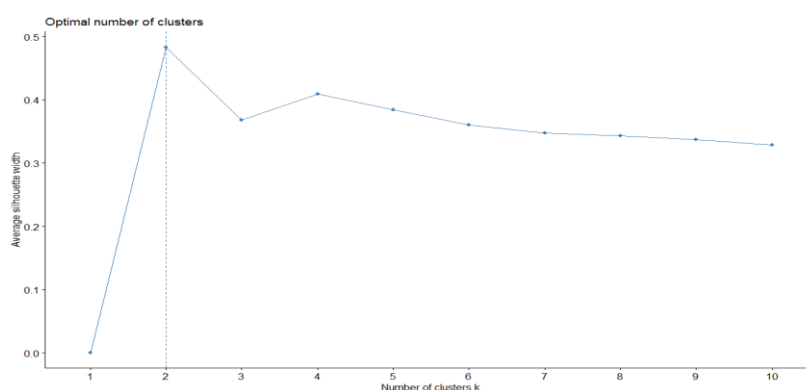
3.3 K-Medoid Clustering

K-medoid mengelompokkan data ke dalam beberapa kelompok yang serupa berdasarkan karakteristik tertentu. Sebelum melaksanakan pengelompokan data, peneliti akan menentukan jumlah kelompok yang akan dibentuk berdasarkan data komponen utama 1 dan 2 menggunakan menghitung nilai *within sum of square* (meminimalkan jumlah varians dalam kelompok) dan metode *silhouette*.

Visualisasi dari nilai *within sum of square* dan metode *silhouette* dapat dilihat pada Gambar 6 dan 7.



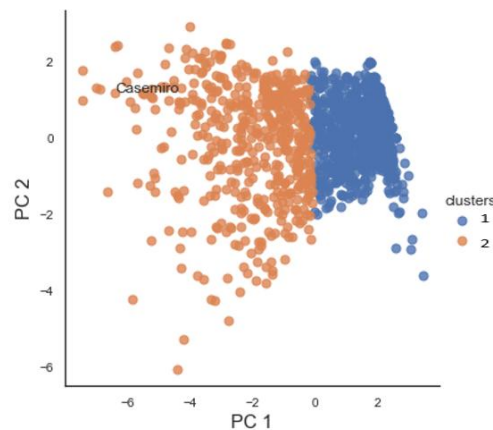
Gambar 6. Visualisasi *Within Sum Of Square*



Gambar 7. Visualisasi Metode *Silhouette*.

Pada Gambar 6. dapat dilihat bahwa nilai sum of square mulai melandai pada kelompok 1 ke 2 sehingga dapat disimpulkan bahwa kelompok maksimal yang terbentuk dari *K-Medoid Clustering* sebanyak 2. Gambar 7. memberikan nilai *Optimal Number of Cluster* pada angka 2 sehingga dapat disimpulkan bahwa kelompok maksimal yang terbentuk dari analisis *K-Medoid Clustering* adalah 2.

Setelah menentukan jumlah kelompok, dilanjutkan dengan melakukan analisis *K-Medoid Clustering*. Dalam analisis ini, pengelompokan dilakukan berdasarkan jarak antar titik data dengan pusat kelompok terdekat. Dengan demikian, pengelompokan data yang dilakukan dapat memberikan informasi yang lebih detail dan mudah dipahami bagi peneliti dalam mengambil keputusan atau menyusun strategi. Berikut ini adalah visualisasi *cluster* yang terbentuk dari analisis analisis *K-Medoid Clustering*.



Gambar 8. Visualisasi analisis *K-Medoid Clustering*

Gambar 8. memvisualisasikan 2 kelompok yang terbentuk dengan menggunakan warna yang berbeda, yaitu: kelompok 1 berwarna biru, kelompok 2 berwarna oranye. Banyaknya anggota untuk setiap kelompok dapat dilihat pada Tabel 8.

Tabel 7. Banyak nya Anggota Kelompok

Kelompok	Banyaknya anggota
1	737
2	479

Selanjutnya akan ditampilkan profiling kelompok berdasarkan nilai rata-rata untuk setiap variabelnya. Untuk kelompok klaster 1 memiliki rata rata umur sebesar 24.11. Untuk keakuratan umpan kelompok klaster1 memiliki akurasi sebesar 73.98% dengan rata jumlah umpan progresif sebesar 729.60. Jumlah Tekel dan *intercept* pemain di kelompok ini memiliki rata rata 12.41 kali . Kelompok ini memiliki rata rata *assist* sebanyak 0.55 dengan rata rata percobaan *shooting* sebanyak 1.59 qkali setiap 90 menit. Untuk *area* penguasaan bola pada kelompok ini dominan pada *area* tengah dengan rata rata sentuhan sebanyak 144.03 kali, sedangkan untuk *area* menyerang dan bertahan masing masing memiliki rata rata sentuhan 105.22 dan 51.25.

Untuk Kelompok klaster 2 memiliki rata rata umur sebesar 25.97. Untuk keakuratan umpan kelompok klaster1 memiliki akurasi sebesar 78.60% dengan rata-rata jumlah umpan progresif sebesar 4045.55. Jumlah Tekel dan *intercept* pemain di kelompok ini memiliki rata rata 60.82 kali. Kelompok ini memiliki rata rata *assist* sebanyak 2.77 dengan rata rata percobaan *shooting* sebanyak 1.34 kali setiap 90 menit. Untuk *area* penguasaan bola pada kelompok ini dominan pada *area* tengah dengan rata rata sentuhan sebanyak 681.69 kali, sedangkan untuk area menyerang dan bertahan masing masing memiliki rata rata sentuhan 361.19 dan 254.12.

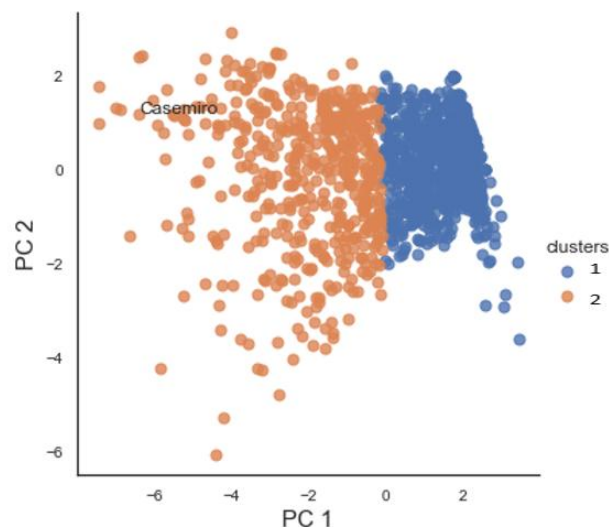
Berdasarkan perbandingan rata rata kedua kelompok, dapat diambil kesimpulan bahwa kelompok 2 memiliki keunggulan di semua variabel kecuali variabel *shoot* per 90 menit jika dibandingkan dengan kelompok 1.

Casemiro, merupakan pemain yang masuk ke kelompok 2. Hal ini sangat sesuai, karena Casemiro posisinya adalah sebagai pemain gelandang bertahan yang jarang melakukan serangan sehingga *shoot* yang dilakukan lebih rendah dibanding rata-rata pemain lain.

Tabel 8. Perbandingan Rata-rata Setiap Variabel

Variabel	Kelompok 1	Kelompok 2
<i>Age</i>	24.11	25.97
<i>Cmp</i>	73.98	78.60
<i>PrqDist</i>	729.60	4045.55
<i>Ast</i>	0.55	2.77
<i>Tkl.Int</i>	12.41	60.82
<i>Sh.90</i>	1.59	1.34
<i>Def.3rd</i>	51.25	254.12
<i>Mid.3rd</i>	144.03	681.69
<i>Att.3rd</i>	105.22	361.19

Langkah selanjutnya adalah melakukan regenerasi pemain dengan memilih pemain dengan jarak *Euclidean* paling dekat dengan Casemiro.



Gambar 9. Casemiro Sebagai Anggota Kelompok 2

3.4 Pengukuran Jarak *Euclidean*

Pengukuran jarak *Euclidean* membantu mengetahui seberapa dekat setiap pemain dengan Casemiro. Tabel 9 adalah 5 daftar pemain dengan jarak *Euclidean* paling dekat dengan titik Casemiro.

Tabel 9. Daftar Pemain Paling Dekat dengan Casemiro

No	Player	Jarak <i>Euclidean</i>
1	Florian Tardieu	0.443625
2	Aurélien Tchouaméni	0.506547

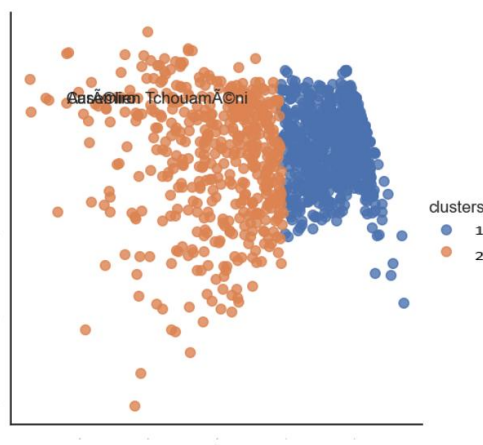
No	Player	Jarak <i>Euclidean</i>
3	Christian Nørgaard	0.551718
4	Declan Rice	0.618342
5	Cheick Doucouré	0.768339

Regenerasi dilakukan untuk mengganti pemain yang keluar dengan pemain muda, sehingga dilakukan seleksi berdasarkan kriteria: usia pemain dibawah 23 tahun dan jaraknya dekat dengan Casemiro. Hasil pemilihan dapat dilihat pada Tabel 10.

Berdasarkan Tabel 10 kita lihat bahwa pemain muda yang paling dekat dengan Casemiro adalah Aurélien Tchouaméni (0.517). Oleh karena itu kita simpulkan bahwa pemain yang paling tepat untuk menggantikan Casemiro berdasarkan jarak *Euclidean* adalah Aurélien Tchouaméni. Hal ini dapat kita lihat secara visual, pada Gambar 10., bahwa Casemiro dan Aurélien Tchouaméni berada pada klaster yang sama, yaitu klaster 2. Dengan demikian Tchouaméni merupakan pemain yang paling tepat untuk menggantikan Casemiro di Real Madrid.

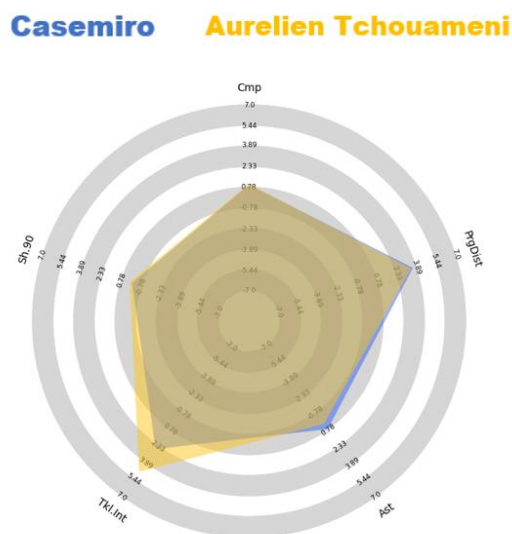
Tabel 10. Daftar Pemain Muda Paling Dekat dengan Casemiro berdasarkan jarak *Euclidean*

No	Player	Jarak <i>Euclidean</i>	Usia
1	Aurélien Tchouaméni	0.51	21
2	Declan Rice	0.62	22
3	Cheick Doucouré	0.77	21
4	Boubacar Kamara	1.19	21
5	Mahdi Camara	1.89	23



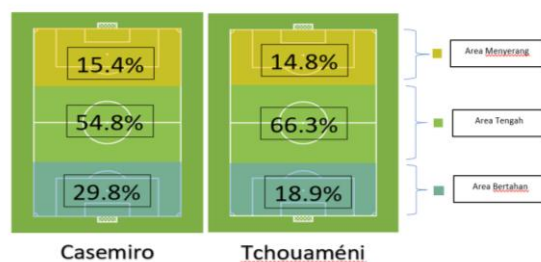
Gambar 10. Letak titik Casemiro, Tchouaméni

Selanjutnya diperlihatkan perbandingan profil Casemiro dan Tchouaméni dalam statistik bermain nya, lihat Gambar 11.



Gambar 11. Visualisasi Perbandingan Casemiro dengan Tchouameni

Berdasarkan visualisasi pada Gambar 11. terlihat bahwa Casemiro dan Tchouameni memiliki banyak kemiripan dalam hal statistik bermain. Perbedaan mereka adalah Tchouameni memiliki tingkat tekel dan *intercept* yang lebih tinggi dibanding Casemiro.



Gambar 12. Visualisasi Perbandingan Penguasaan Bola

Dalam hal area penguasaan bola, Gambar 12., kedua pemain memiliki karakteristik penguasaan bola yang cenderung lebih banyak dilakukan di *area* bertahan daripada *area* menyerang. Hal ini menunjukkan bahwa kedua pemain memiliki peranan yang sama untuk menjadi gelandang yang lebih banyak bertahan dan menyalurkan bola dari *area* bertahan ke *area* tengah. Namun Tchouameni lebih bermain ke *area* lapangan tengah dibanding Casemiro.

4. SIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, diketahui bahwa pada musim 2021/2022 Casemiro memiliki akurasi umpan (*CMP*), umpan progresif (*PrgDist*), tekel dan intersep (*Tkl.Int*), *Asist* (*Ast*) di atas rata-rata gelandang Eropa. Hanya saja *shooting* per 90 menit Casemiro lebih rendah dibanding rata-rata gelandang Eropa. Hal tersebut dikarenakan posisi Casemiro sebagai gelandang bertahan sehingga jarang melakukan percobaan *shooting*. Pada musim 2021/2022 Casemiro cenderung lebih banyak melakukan penguasaan bola di *area* bertahan dibanding *area* menyerang. Hal ini menandakan posisi Casemiro sebagai gelandang bertahan yang cenderung melakukan *built up* serangan dan pergerakan di *area* pertahanan.

Analisis *K-Medoid Clustering* menempatkan Casemiro pada kelompok 2 gelandang top dunia. Kelompok 2 ini memiliki karakteristik berisikan pemain yang unggul pada setiap variabel yang diteliti kecuali pada variabel *shoot* per 90 menit. Untuk variabel *shoot* per 90 menit ini, kelompok 1 lebih unggul. Selanjutnya, kedekatan karakteristik pemain berdasarkan pengukuran jarak

Euclidean memberikan masukan bahwa pemain yang paling tepat menggantikan Casemiro di Real Madrid adalah Tchouameni.

Walaupun Tchouameni adalah pemain muda yang paling tepat menggantikan Casemiro di Real Madrid, namun terdapat perbedaan statistik Tchouameni dengan Casemiro. Tchouameni memiliki tingkat tekel dan *intercept* yang lebih tinggi dibanding Casemiro. Juga Tchouameni sedikit lebih unggul dalam hal umpan progresif. Casemiro dan Tchouameni, memiliki karakteristik penguasaan bola yang cenderung lebih banyak dilakukan di *area* bertahan. Walaupun memiliki peranan yang sama untuk menjadi gelandang bertahan, Tchouameni lebih bermain ke *area* lapangan tengah dibanding Casemiro.

Untuk mendapatkan hasil yang lebih *comprehensive*, variabel lain semisal harga pemain dalam bursa *transfer* setiap akhir musim, dapat dilakukan untuk dimasukkan ke dalam variabel penelitian.

DAFTAR PUSTAKA

- Amisatun, S., Wakhidah, N., & Putri, A. N. (2020). Penerapan Metode *K-Medoid* Untuk Pengelompokan Kondisi Jalan di Kota Semarang, *JATISI*, Vol. 6, No. 2..
- Atmaja, E. H. S. (2019). Implementation of K-Medoid Clustering Algorithm to Cluster Crime Patterns in Yogyakarta, *International Journal of Applied Sciences and Smart Technologies*, Vol 01, Issue 01.
- BBC. (2014, December 30). *AirAsia QZ8501: Does bad weather cause plane crashes?* Retrieved from BBC: <http://www.bbc.com/news/world-30631968>
- Brilliantro, B. A. (2021, Maret 2). *Muda dan Konsisten*. Retrieved from <https://theflanker.id/>: <https://theflanker.id/muda-dan-konsisten-b3fe1f>
- Dihni, V. A. (2021, Oktober 5). *Sepak Bola Jadi Olahraga Paling Populer di Dunia*. Retrieved from databoks.katadata.co.id: <https://databoks.katadata.co.id/datapublish/2021/10/05/sepak-bola-jadi-olahraga-paling-populer-di-dunia>
- Ebejer. (n.d.). *Formation In Football Deffensive and Attacking*. Retrieved from fieldinsider.com: <https://fieldinsider.com/defensive-attacking-formations/>
- Fazrul, I. (2023, Februari 15). *11 Posisi Pemain Sepak Bola Dan Tugasnya Di Lapangan. Ulasan Paling Lengkap*. Retrieved from berita.99.co: <https://berita.99.co/posisi-pemain-sepak-bola/>
- Febriyanti, S., & Nugraha, J. (2022). Application of K-Medoid Clustering to Increase the 2020 Family Planning Program in Sleman Regency. *Enthusiastic : International Journal of Applied Statistics and Data Science*, 2(1), 10–18.
- Habib, D. (2022). *Top 50 Best Football Leagues in the World Ranking 2022*. Retrieved from thefootballlovers.com..
- Johnson, R. A., & Bhattacharyya, G. K. (2010). *Statistics Principles & Methods*. USA: John Wiley & Sons.
- Johnson, R., & Wichern, D. (2007). Applied Multivariate Statistical Analysis. In R. Johnson, & D. Wichern, *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Education, Inc.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Leidiana, H. (2013). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, 1(1), 65–76.
- Li, X. (2013). Comparison and Analysis between Holt Exponential Smoothing and Brown Exponential Smoothing Used for Freight Turnover Forecast. *Third International Conference on Intelligent System Design and Engineering Applications* (pp. 453-456). IEEE.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
- Miftahuddin, Y., Umaroh, S., & Karim, F. (2020). PERBANDINGAN METODE PERHITUNGAN JARAK EUCLIDEAN, HAVERSINE, DAN MANHATTAN DALAM PENENTUAN POSISI KARYAWAN. *Jurnal Tekno Insentif*, 14(2), 69-77.

- Nusufi, M. (2016). Hubungan Kemampuan Motor Ability dengan Keterampilan Bermain, *Jurnal Ilmiah Mahasiswa Pendidikan Jasmani, Kesehatan dan Rekreasi*, Vol. 1, No. 1, Pages 1-10.
- Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoid clustering. *Expert Systems with Applications*, Volume 36, Issue 2, Part 2, Pages 3336-3341.
- Pratama, A. (2017). PENILAIAN KETERAMPILAN DASAR SEPAK BOLA PASSING DAN . 6, 41.
- Riyanto, B. (2019). Penerapan Algoritma K-medoid Clustering Untuk Pengelompokan Penyebaran Diare di Kora Medan.
- Saputra, Z. (2011). Analisis Perancangan Program Aplikasi Untuk Klasifikasi Menggunakan KNN. Skripsi, Teknik Informatika FTI UII, Yogyakarta.
- Sousa. (1984). The role of disturbance in natural communities. *Annual Review of Ecology and Systematics*, 353-391.
- Sudrajat, I. (2021, Agustus 6). *Mengenal Posisi Pemain Lini Tengah dalam Sepak Bola*. Retrieved from skor.id: <https://skor.id/post/skorpedia-mengenal-posisi-pemain-lini-tengah-dalam-sepak-bola-01389067>
- Suhail, Y., Rajpoot, N., & Markham, M. (2011). Principal Component Analysis for Dimension Reduction in Massive Data Sets with Applications to Gene Expression Analysis.
- Tukey. (1977). *Exploratory data analysis*. Addison-Wesley.
- Utari, D. (2019). *Modul Praktikum Analisis Regresi Terapan dengan R*. Yogyakarta: Universitas Islam Indonesia.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). *Probability & Statistics for Engineers & Scientists 9th Ed*. USA: Pearson.
- Wicaksono, R. H. (2022). Penerapan Principal Component Analysis, K-Means Clustering dan K-Nearest Neighbors Dalam Mencari Pemain Belakang.
- Yadav, R., & Sharma, A. (2012). Advanced Methods to Improve Performance of K-Means Global Journal of Computer Science and Technology Volume XII Issue IX Version I, Global Journals Inc. (US).