Vol. 7 No. 1 Juni 2026 Hal : 117 - 132 E-ISSN: 2723-1089
P-ISSN: 2776-7779

Penerapan Algoritma K-Nearest Neighbors untuk Prediksi Negara-Negara Asia Berpotensi Lolos Piala Dunia 2026

Andrian Sahputra [1)], Ilka Zufria [2)] & Mhd Ikhsan Rifki [3)]*

[1,2,3)] Prodi Ilmu Komputer, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

*Coresponding Email: andrianmunthe750@gmail.com, ilkazufria@uinsu.ac.id, rifki.mhdikhsan@uinsu.ac.id

Abstrak

Penelitian ini bertujuan menerapkan algoritma K-Nearest Neighbors dalam memprediksi potensi kelolosan negara-negara Asia dari babak ketiga kualifikasi menuju putaran final Piala Dunia 2026. Data yang digunakan terdiri atas 36 data historis kualifikasi Piala Dunia periode 2014–2022 sebagai data latih dan 18 data peserta kualifikasi zona Asia tahun 2026 sebagai data uji. Pendekatan yang digunakan adalah kuantitatif dengan teknik data mining berbasis klasifikasi. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan melalui pembersihan dan transformasi fitur menjadi rasio performa, normalisasi data, serta pemodelan menggunakan algoritma K-Nearest Neighbors. Hasil evaluasi menunjukkan bahwa model dengan parameter k = 4 menghasilkan akurasi sebesar 83,33%, dengan performa yang baik pada kategori lolos dan tidak lolos, meskipun kategori play-off menunjukkan hasil yang lebih rendah. Nilai rata-rata makro untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,79, 0,77, dan 0,77. Hasil prediksi menunjukkan enam negara berpotensi lolos langsung, empat negara berada pada posisi play-off, dan delapan negara lainnya tidak lolos. Temuan ini menunjukkan bahwa algoritma K-Nearest Neighbors memiliki kemampuan yang cukup baik dalam memprediksi kelolosan tim berdasarkan performa historis.

Kata Kunci: K-Nearest Neighbors, data mining, prediksi, kualifikasi Piala Dunia, AFC.

Abstract

This study aims to apply the K-Nearest Neighbors algorithm to predict the qualification potential of Asian countries from the third round of qualification to the 2026 FIFA World Cup finals. The dataset consists of 36 historical qualification records from 2014–2022 as training data and 18 Asian qualification participants in 2026 as testing data. A quantitative approach with a data mining classification technique is employed. The research process includes data collection, preprocessing through data cleaning and feature transformation into performance ratios, normalization, and model development using the K-Nearest Neighbors algorithm. The evaluation results show that the model with k = 4 achieves an accuracy of 83.33%, with strong performance in the qualified and non-qualified categories, while the play-off category remains less optimal. The macro average values for precision, recall, and f1-score are 0.79, 0.77, and 0.77, respectively. The prediction results indicate six countries are likely to qualify directly, four countries enter the play-off stage, and eight countries fail to qualify. These findings demonstrate that the K-Nearest Neighbors algorithm is effective in predicting team qualification based on historical performance.

Keywords: K-Nearest Neighbors, data mining, prediction, World Cup qualification, AFC.

PENDAHULUAN

Perkembangan teknologi informasi dalam beberapa dekade terakhir telah menghadirkan transformasi yang substansial pada beragam sektor kehidupan, termasuk bidang olahraga. Integrasi data serta teknik analisis berbasis komputasi kini menjadi elemen esensial dalam proses pengambilan keputusan, baik untuk meningkatkan kinerja tim, merumuskan strategi pertandingan, maupun melakukan evaluasi secara objektif (Zein & Gunawan, 2022). Dalam perkembangan sepak bola modern, pemanfaatan data statistik tidak lagi terbatas pada analisis pasca pertandingan, melainkan telah berkembang ke arah prediksi performa tim melalui pendekatan berbasis data (data-*driven*). Sepak bola sebagai olahraga dengan tingkat popularitas global yang tinggi memiliki intensitas kompetisi yang sangat ketat, khususnya pada ajang internasional seperti Piala Dunia, sehingga setiap negara berupaya meningkatkan kualitas tim nasional melalui pembinaan pemain, penguatan taktik, serta optimalisasi teknologi analisis data (Karim et al., 2023).

Prediksi performa tim menjadi aspek penting dalam mengukur peluang keberhasilan suatu negara dalam kompetisi berskala besar. Piala Dunia 2026 menandai fase baru dalam perkembangan sepak bola dunia dengan perubahan format turnamen dari 32 menjadi 48 peserta. Perubahan tersebut membuka peluang lebih luas bagi negara-negara dari berbagai kawasan, termasuk Asia, untuk tampil pada putaran final. Kendati demikian, pendekatan prediksi konvensional masih menghadapi keterbatasan dalam mengelola kompleksitas data performa tim, sementara tingkat persaingan pada babak kualifikasi semakin intens dan dinamis (Dip et al., 2024).

Performa tim-tim Asia dalam beberapa tahun terakhir menunjukkan tren peningkatan yang signifikan, baik dari sisi teknik permainan, strategi, maupun kualitas individu pemain. Pengalaman bertanding di liga internasional, sistem pembinaan pemain muda yang semakin terstruktur, serta penerapan taktik modern berkontribusi terhadap meningkatnya daya saing tim. Situasi tersebut menjadikan

proses prediksi semakin kompleks dan menuntut pendekatan analisis yang lebih objektif serta berbasis data empiris (Bhowmick & Sarvaiya, 2021).

Kemajuan teknologi turut mendorong pemanfaatan metode pembelajaran mesin sebagai instrumen analisis dalam olahraga. Data statistik kualifikasi, seperti jumlah pertandingan, kemenangan, hasil imbang, kekalahan, jumlah gol, dan perolehan poin, dapat diolah menjadi informasi prediktif yang bernilai (Evitania, 2023). Salah satu metode yang relevan adalah algoritma K-Nearest Neighbors (KNN), yaitu metode berbasis kedekatan antar data yang digunakan untuk memprediksi kategori suatu objek. Penentuan hasil dilakukan dengan mengacu pada mayoritas kelas dari sejumlah tetangga terdekat. Metode ini memiliki keunggulan dalam kesederhanaan implementasi serta kemampuannya mengolah data historis tanpa memerlukan asumsi distribusi tertentu (Munif et al., 2022).

Sejumlah penelitian telah mengkaji penerapan metode pembelajaran mesin dalam prediksi berbasis data, termasuk pada bidang olahraga. Zein dan Gunawan (2022) menunjukkan bahwa algoritma seperti Naïve Bayes, Neural Network, dan Support Vector Machine menghasilkan tingkat akurasi yang beragam dalam memprediksi hasil Piala Dunia. Penelitian oleh (Maskuri et al., 2022) juga memperlihatkan efektivitas algoritma K-Nearest Neighbor (KNN) dalam prediksi penyakit stroke dengan akurasi mencapai 95% melalui pemodelan data menggunakan metode split validation. Temuan tersebut menegaskan bahwa metode klasifikasi memiliki kapasitas prediktif yang kuat, meskipun kajian yang ada masih terbatas pada prediksi hasil atau kasus tertentu dan belum mengembangkan analisis yang menitikberatkan pada peluang kelolosan tim berdasarkan performa kualifikasi.

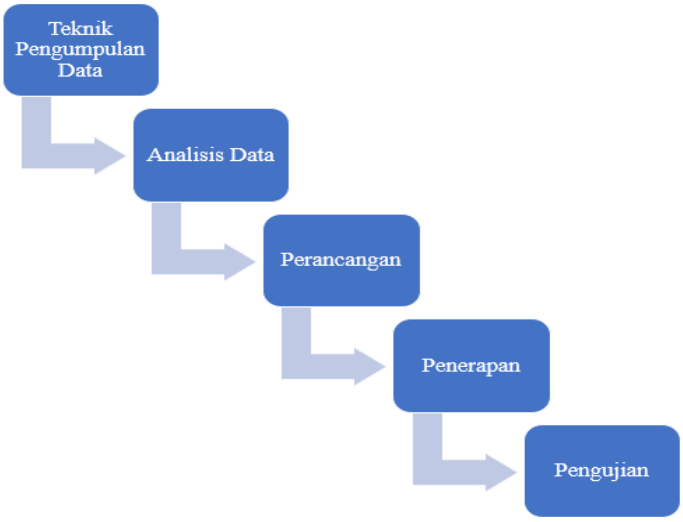
Penelitian ini bertujuan untuk membangun model prediksi berbasis KNN serta menganalisis hasil klasifikasi guna memberikan gambaran potensi kelolosan tim secara lebih objektif. Penelitian ini juga mencakup evaluasi terhadap kinerja model yang dihasilkan, sehingga dapat diketahui tingkat akurasi serta keandalan metode dalam proses prediksi. Luaran penelitian diharapkan memberikan kontribusi substansial bagi pengembangan analisis olahraga berbasis data, sekaligus

memperkaya khazanah kajian ilmiah yang dapat dijadikan rujukan dalam penelitian lanjutan pada bidang yang sejenis.

METODE PENELITIAN

Dalam penelitian ini digunakan metode kuantitatif dengan pendekatan data mining klasifikasi berbasis algoritma K-Nearest Neighbors (K-NN) untuk memprediksi potensi kelolosan negara-negara Asia ke Piala Dunia 2026 (Mulyadi et al., 2025). Data yang dianalisis berupa informasi numerik, seperti jumlah pertandingan, kemenangan, hasil imbang, kekalahan, gol tercetak, gol kebobolan, selisih gol, dan total poin, yang kemudian ditransformasikan menjadi parameter turunan berupa rasio kemenangan, rasio imbang, rasio kekalahan, rata-rata gol cetak, rata-rata kebobolan, serta rata-rata poin per pertandingan. Proses penelitian diawali dengan pengumpulan data performa tim nasional Asia dari babak ketiga kualifikasi zona AFC dan data edisi sebelumnya sebagai data latih, dilanjutkan dengan tahap pra-pemrosesan, normalisasi, pemisahan data latih dan data uji, serta pemodelan menggunakan algoritma K-NN.

Guna mencapai hasil penelitian yang selaras dengan tujuan yang telah ditetapkan, diperlukan suatu rancangan kerja yang tersusun secara sistematis. Rancangan tersebut kemudian disajikan dalam bentuk kerangka kerja penelitian yang divisualisasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

1. Teknik Pengumpulan Data

   Tahap ini bertujuan memperoleh data yang relevan dan akurat melalui studi pustaka dan web scraping. Studi pustaka dilakukan dengan menelaah berbagai sumber ilmiah untuk memperkuat landasan teori. Web scraping dimanfaatkan untuk mengekstraksi data statistik dari situs resmi dan platform olahraga, yang kemudian disimpan dalam format terstruktur untuk keperluan analisis (Dewi & Rahayu, 2022).

2. Analisis Data

   Tahapan ini difokuskan pada pengolahan data untuk menemukan pola dan informasi yang bermakna. Proses dimulai dari pengumpulan data, pembersihan, serta seleksi fitur yang relevan, kemudian dilanjutkan dengan penentuan model KNN, pelatihan data, dan evaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* guna menilai kinerja model (Urva et al., 2023).

3. Perancangan Sistem

   Tahap ini bertujuan menyusun alur kerja sistem secara sistematis melalui representasi flowchart algoritma KNN. Proses meliputi penentuan nilai k, perhitungan jarak antar data, pengurutan jarak, pemilihan tetangga terdekat, hingga penentuan kelas berdasarkan mayoritas (Amalia et al., 2022).

4. Implementasi Sistem

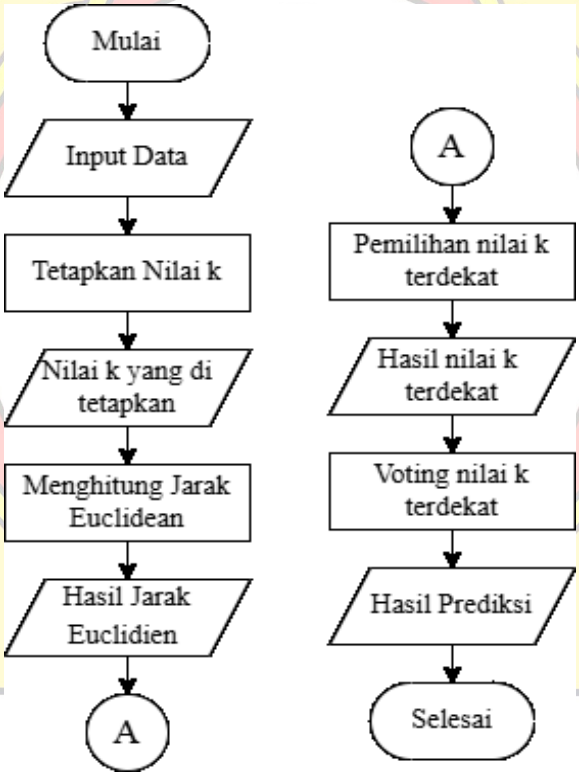
   Tahapan ini mencakup penerapan algoritma KNN menggunakan data yang telah diproses, termasuk normalisasi dan pembagian data menjadi data latih dan data uji. Model kemudian digunakan untuk menghitung jarak dan mengklasifikasikan potensi kelolosan tim berdasarkan kedekatan karakteristik data (Rismala et al., 2023).

5. Pengujian Sistem

   Tahap ini dilakukan untuk mengevaluasi performa model menggunakan confusion matrix. Hasil prediksi dibandingkan dengan data aktual untuk mengidentifikasi tingkat ketepatan dan kesalahan klasifikasi, yang

selanjutnya digunakan dalam perhitungan metrik evaluasi guna menilai efektivitas model (Rahmadinia et al., 2023).

Selain penyusunan kerangka kerja penelitian, pemilihan metode yang tepat menjadi komponen esensial dalam pengolahan data. Penelitian ini menggunakan algoritma K-Nearest Neighbors (K-NN) sebagai metode klasifikasi untuk memprediksi potensi kelolosan negara-negara Asia ke Piala Dunia 2026 (Damanik & Gunawan, 2024). Algoritma ini mengelompokkan data berdasarkan kedekatan karakteristik antarobjek, sehingga mampu menentukan kategori tim ke dalam kelas tertentu seperti lolos langsung, *play-off*, atau tidak lolos. Alur penerapan metode tersebut disajikan secara ringkas dalam bentuk flowchart pada Gambar 2.



Gambar 2. Flowchart algoritma K-Nearest Neighbor

HASIL DAN PEMBAHASAN

Analisis Data

Analisis data dimulai dari pengumpulan data historis performa negara peserta Piala Dunia 2014–2022 serta data kualifikasi zona AFC 2026 yang mencakup 18

negara Asia, dengan variabel statistik pertandingan dan status kelolosan sebagai acuan. Data kemudian dibersihkan, disusun ulang, dan ditransformasikan ke dalam bentuk rasio serta rata-rata per pertandingan agar lebih representatif. Dataset yang telah diproses digunakan dalam pemodelan K-Nearest Neighbors dengan perhitungan jarak menggunakan *Euclidean Distance* untuk menentukan prediksi berdasarkan kedekatan karakteristik. Evaluasi dilakukan pada data uji untuk menilai ketepatan model dalam mengklasifikasikan status kelolosan.

Representasi Data

Data penelitian bersumber dari arsip resmi FIFA dan AFC, yang terdiri atas data historis babak ketiga kualifikasi Piala Dunia 2014–2022 sebanyak 36 entri sebagai data latih serta data 18 negara peserta kualifikasi 2026 sebagai data uji. Setiap entri memuat variabel statistik pertandingan seperti jumlah laga, kemenangan, imbang, kekalahan, gol, kebobolan, selisih gol, dan poin. Data kemudian diproses melalui normalisasi dan transformasi fitur menjadi rasio serta rata-rata per pertandingan, lalu direpresentasikan dalam bentuk vektor numerik multidimensi untuk memungkinkan perhitungan kedekatan menggunakan *Euclidean Distance* dalam algoritma KNN guna menentukan klasifikasi kelolosan. Data latih yang digunakan dalam penelitian ini disajikan pada Tabel 1 berikut.

Tabel 1. Dataset Latih

No	Tahun	Negara	Main	Menang	Seri	Kalah	GF	GA	GD	Poin	Status
1	2022	Iran	10	8	1	1	15	4	11	25	Lolos
2	2022	Korea Selatan	10	7	1	2	13	3	10	22	Lolos
3	2018	Uzbekistan	10	4	1	5	6	7	-1	13	*Play-off*
4	2018	China	10	3	3	4	8	10	-2	12	Tidak Lolos
5	2014	Iraq	6	3	1	2	8	6	2	10	Lolos
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
36	2014	Korea Utara	6	0	2	4	2	10	-8	2	Tidak Lolos

Data kualifikasi Piala Dunia zona Asia pada babak ketiga dimanfaatkan sebagai data uji untuk memprediksi status kelolosan setiap negara, baik lolos langsung, melaju ke babak *play-off*, maupun tidak lolos. Data tersebut disajikan pada Tabel 2 berikut.

Tabel 2. Data Uji

No	Tahun	Negara	Main	Menang	Seri	Kalah	GF	GA	GD	Poin	Status
1	2026	Iran	10	7	2	1	19	8	11	23	?
2	2026	Korea Selatan	10	6	4	0	20	7	13	22	?
3	2026	Uzbekistan	10	6	3	1	14	7	7	21	?
4	2026	China	10	3	0	7	7	20	−13	9	?
5	2026	Iraq	10	4	3	3	9	9	0	15	?
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
6	2026	Korea Utara	10	0	3	7	9	21	−12	3	?

Data yang diperoleh akan melalui tahap pre-processing sebelum diimplementasikan dalam model. Tahapan ini merupakan bagian esensial untuk memastikan data berada dalam kondisi bersih, terstruktur, dan konsisten. Proses yang dilakukan mencakup penyelarasan skala, pengurangan noise, serta pembentukan fitur yang relevan bagi algoritma K-Nearest Neighbor (KNN), sehingga kualitas data meningkat dan mendukung ketepatan hasil prediksi.

1. Pembersihan Data (*Cleaning*)

   Tahap ini berfokus pada penghapusan duplikasi, nilai kosong, serta data yang tidak lengkap pada data latih dan data uji. Proses tersebut menjaga konsistensi dan kualitas data agar layak digunakan pada tahap pengolahan berikutnya.

2. Transformasi Fitur

   Atribut numerik diolah menjadi metrik yang lebih representatif, seperti rasio kemenangan, imbang, kekalahan, serta rata-rata gol dan poin per pertandingan. Transformasi ini menghasilkan ukuran performa yang proporsional sehingga memudahkan perbandingan antarnegara dalam proses klasifikasi.

3. Seleksi Fitur (*Feature Selection*)

   Tahap ini memilih fitur yang paling relevan, yaitu rasio performa tim seperti *WinRatio, DrawRatio, LossRatio, GoalAvg, ConcedeAvg, DiffAvg*, dan *PointAvg*. Kolom status digunakan sebagai label pada data latih, sedangkan data uji hanya memanfaatkan fitur tanpa label untuk keperluan prediksi.

Setelah melalui tahapan pre-processing, diperoleh hasil data yang telah terstruktur dan siap digunakan, sebagaimana disajikan pada tabel data latih dan data uji berikut.

Tabel 3. Hasil Pra-pemrosesan Dataset Latih

No	*Win Ratio*	*Draw Ratio*	*Loss Ratio*	*GoalAvg*	*Concede Avg*	*Diff Avg*	*Point Avg*	Status
1	0,80	0,10	0,10	1,50	0,40	1,10	2,50	Lolos
2	0,70	0,20	0,10	1,30	0,30	1,00	2,30	Lolos
3	0,40	0,10	0,50	0,60	0,70	-0,10	1,30	*Play-off*
4	0,30	0,30	0,40	0,80	1,00	-0,20	1,20	Tidak Lolos
5	0,50	0,17	0,33	1,33	1,00	0,33	1,67	Lolos
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
36	0,00	0,33	0,67	0,33	1,67	-1,33	0,33	Tidak Lolos

Tabel 4. Hasil Pra-pemrosesan Dataset Uji

No	*Win Ratio*	*Draw Ratio*	*Loss Ratio*	*Goal Avg*	*Concede Avg*	*Diff Avg*	*Point Avg*	Status
1	0,70	0,20	0,10	1,90	0,80	1,10	2,30	?
2	0,60	0,40	0,00	2,00	0,70	1,30	2,20	?
3	0,60	0,30	0,10	1,40	0,70	0,70	2,10	?
4	0,30	0,00	0,70	0,70	2,00	-1,30	0,90	?
5	0,40	0,30	0,30	0,90	0,90	0,00	1,50	?
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
6	0,00	0,30	0,70	0,90	2,10	-1,20	0,30	?

Algoritma K-Nearest Neighbor

Setelah tahap pra-pemrosesan data, dilakukan penerapan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan status kelolosan negara berdasarkan performa pertandingan. Data historis Piala Dunia zona Asia tahun 2014, 2018, dan 2022 digunakan sebagai data latih, dengan penentuan nilai k sebagai jumlah tetangga terdekat yang menjadi acuan klasifikasi. Model menghitung jarak antar data pada ruang fitur yang telah ditransformasi, kemudian menetapkan kelas berdasarkan mayoritas tetangga terdekat. Pendekatan ini memungkinkan penentuan status kelolosan melalui kesamaan pola performa dengan data historis.

Pada tahap implementasi, seluruh data historis sebanyak 36 entri dimanfaatkan sebagai data latih terlihat pada Tabel 3. Fitur yang digunakan meliputi WinRatio, DrawRatio, LossRatio, GoalAvg, ConcedeAvg, *DiffAvg*, dan *PointAvg* sebagai representasi performa tim. Kolom Status berfungsi sebagai label kelas yang terdiri dari kategori Lolos, *Play-off*, dan Tidak Lolos. Struktur ini memungkinkan model melakukan proses klasifikasi secara sistematis dan terarah, serta dilanjutkan dengan pengujian terhadap data ke-1 pada Tabel 4 untuk menilai hasil prediksi yang dihasilkan.

Langkah selanjutnya adalah menghitung jarak *Euclidean* antara data latih dan data uji dengan membandingkan setiap atribut numerik untuk memperoleh tingkat kedekatan. Pada implementasi ini digunakan nilai k = 5, sehingga lima tetangga terdekat dijadikan acuan dalam penentuan klasifikasi. Pemilihan nilai ganjil membantu menghindari hasil yang ambigu serta menjaga kestabilan prediksi. Secara matematis, jarak *Euclidean* dihitung dengan persamaan 1 sebagai berikut:

$$d_{x,y} = \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2} \quad \text{.....(1)}$$

Rumus jarak *Euclidean* pada algoritma KNN digunakan untuk mengukur tingkat kedekatan antara data latih dan data uji berdasarkan seluruh atribut numerik yang dimiliki. Notasi $d(x, y)$ menyatakan jarak antara dua data, di mana x merepresentasikan data latih dan y menunjukkan data uji..

$$d(1,1) = \sqrt{(0{,}80 - 0{,}70)^2 + (0{,}10 - 0{,}20)^2 + (0{,}10 - 0{,}10)^2 + (1{,}50 - 1{,}90)^2}$$
$$\sqrt{+(0{,}40 - 0{,}80)^2 + (1{,}10 - 1{,}10)^2 + (2{,}50 - 2{,}30)^2} = 0{,}6164$$

$$d(2,1) = \sqrt{(0{,}70 - 0{,}70)^2 + (0{,}20 - 0{,}20)^2 + (0{,}10 - 0{,}10)^2 + (1{,}30 - 1{,}90)^2}$$
$$\sqrt{+(0{,}30 - 0{,}80)^2 + (1{,}00 - 1{,}00)^2 + (2{,}30 - 2{,}30)^2} = 0{,}7874$$

$$d(3,1) = \sqrt{(0{,}40 - 0{,}70)^2 + (0{,}10 - 0{,}20)^2 + (0{,}50 - 0{,}10)^2 + (0{,}60 - 1{,}90)^2}$$
$$\sqrt{+(0{,}70 - 0{,}80)^2 + (-0{,}10 - 1{,}10)^2 + (1{,}30 - 2{,}30)^2} = 2{,}0976$$

……

$$d(36,1) = \sqrt{(0{,}00 - 0{,}70)^2 + (0{,}33 - 0{,}20)^2 + (0{,}67 - 0{,}10)^2 + (1{,}50 - 1{,}90)^2}$$
$$\sqrt{+(0{,}40 - 0{,}80)^2 + (1{,}10 - 1{,}10)^2 + (1{,}50 - 2{,}30)^2} = 0{,}9118$$

Penulisan seperti *d*(1,1) mengindikasikan perhitungan jarak antara data latih ke-1 pada Tabel 3 dengan data uji ke-1 pada Tabel 4, sedangkan *d*(2,1) menunjukkan jarak antara data latih ke-2 dengan data uji ke-1, dan seterusnya hingga seluruh data latih dibandingkan terhadap satu data uji. Setelah dilakukan perhitungan dengan menggunakan Persamaan 1, maka diperoleh hasil keseluruhan jarak *Euclidean* yang ditampilkan pada Tabel 5 sebagai berikut:

Tabel 5. Perhitungan Jarak *Euclidean*

Data ke	Jarak *Euclidean*
1	0,6164
2	0,7874
3	2,0976
4	2,1000
5	1,2035
…	…
…	…
36	0,9118

Tahap berikutnya ialah mencari lima data latih terdekat dengan menentukan kelas melalui voting mayoritas dari tetangga tersebut. Dalam penelitian ini digunakan *k* = 5, sehingga klasifikasi didasarkan pada lima data latih terdekat yang diukur menggunakan jarak *Euclidean*. Nilai jarak diurutkan dan dipilih lima yang terkecil sebagai tetangga terdekat.

Berdasarkan hasil perhitungan algoritma K-Nearest Neighbor (KNN) dengan *k* = 5, diperoleh lima tetangga terdekat yang seluruhnya berasal dari kelas Lolos tanpa adanya kemunculan kelas *play-off* maupun Tidak Lolos. Kondisi ini menunjukkan bahwa mayoritas kelas yang mendominasi adalah Lolos. Data uji ke-1 dengan nilai transformasi WinRatio = 0,7, DrawRatio = 0,2, LossRatio = 0,1, GoalAvg = 1,9, ConcedeAvg = 0,8, DiffAvg = 1,1, dan PointAvg = 2,3 diklasifikasikan ke dalam kategori tersebut berdasarkan kedekatan jarak *Euclidean*.

Implementasi Sistem

Setelah proses pelatihan algoritma K-Nearest Neighbor (K-NN) selesai, tahap implementasi dilakukan untuk menguji kemampuan model dalam memprediksi status kelolosan negara pada kualifikasi Piala Dunia zona AFC. Model yang telah dilatih dijalankan melalui Google Colab untuk mengolah data uji dan menghasilkan prediksi kategori lolos langsung, *play-off*, atau tidak lolos. Dari penerapan model yang digunakan di dapat hasil prediksi seperti yang terlihat pada Gambar 3 dibawah ini.

Hasil Prediksi:

Negara	Prediksi_Status
Iran	Lolos
Qatar	Tidak Lolos
Kyrgyzstan	Tidak Lolos
United Arab Emirates	Lolos
Uzbekistan	Lolos
Korea Utara	Tidak Lolos
Jordan	Play-off
Oman	Play-off
Kuwait	Tidak Lolos
Palestine	Tidak Lolos
Korea Selatan	Lolos
Iraq	Play-off
Japan	Lolos
Saudi Arabia	Play-off
Australia	Lolos
Indonesia	Tidak Lolos
China	Tidak Lolos
Bahrain	Tidak Lolos

Gambar 3. Tampilan Hasil Prediksi

Berdasarkan hasil prediksi pada Gambar 3 diperoleh bahwa negara yang lolos langsung adalah Iran, United Arab Emirates, Uzbekistan, Korea Selatan, Jepang, dan Australia. Negara yang masuk ke kategori *play-off* yaitu Jordan, Oman, Iraq, dan Saudi Arabia. Sementara itu, negara yang tidak lolos meliputi Qatar, Kyrgyzstan, Korea Utara, Kuwait, Palestine, Indonesia, China, dan Bahrain.

Pengujian

Pengujian evaluasi dilakukan untuk menilai kinerja model secara komprehensif dengan membandingkan hasil prediksi dan data aktual pada data latih menggunakan Google Colab. Metrik yang digunakan meliputi *accuracy score* untuk mengukur tingkat ketepatan, *confusion matrix* untuk melihat distribusi klasifikasi benar dan salah, serta *classification report* yang mencakup nilai presisi, *recall*, dan *F1-score*. Hasil evaluasi ini memberikan gambaran mengenai kemampuan model dalam mengklasifikasikan kategori kelolosan serta mengidentifikasi potensi kesalahan. Hasil pengujian disajikan pada Tabel 6 dan 7 berikut:

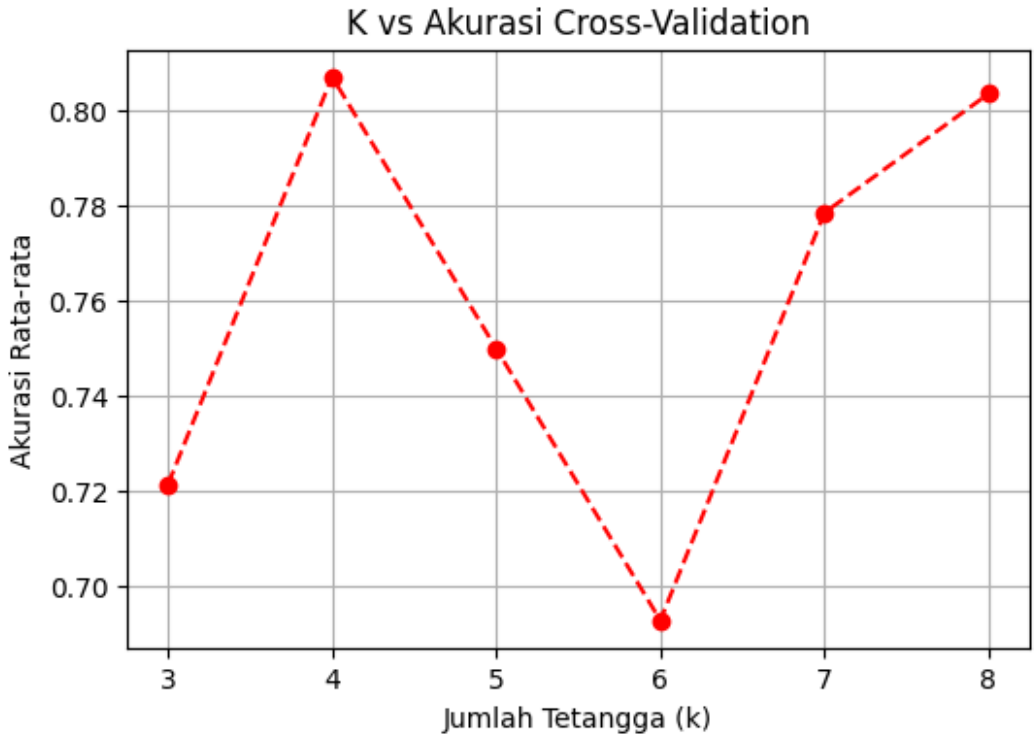
Tabel 6. Hasil Evaluasi KNN

Kategori Aktual \ Prediksi	Lolos	*Play-off*	Tidak Lolos
Lolos	13	1	0
Play-off	3	3	0
Tidak Lolos	1	1	14

Tabel 7. *Classification report* Hasil Evaluasi KNN

Kategori	*Precision*	*Recall*	*F1-score*	*Support*
Lolos	0.76	0.93	0.84	14
Play-off	0.60	0.50	0.55	6
Tidak Lolos	1.00	0.88	0.93	16
Accuracy	-	-	0.83	36
Macro Avg	0.79	0.77	0.77	36
Weighted Avg	0.84	0.83	0.83	36

Pada penelitian ini, pengujian nilai *k* dilakukan untuk memperoleh parameter terbaik pada algoritma KNN dengan membandingkan akurasi dari beberapa nilai *k* menggunakan Google Colab. Hasil pengujian ditampilkan pada Gambar 4, kemudian dipilih nilai *k* dengan performa paling optimal untuk klasifikasi.



Gambar 4. Tampilan Pengujian Akurasi *k*

SIMPULAN

Berdasarkan hasil penelitian, implementasi algoritma K-Nearest Neighbor (KNN) dengan parameter *k* terbaik sebesar 4 melalui *cross-validation* diterapkan pada 54 data yang terdiri atas 36 data latih dan 18 data uji, dengan tahapan pra-pemrosesan berupa pembersihan data dan transformasi fitur menjadi rasio performa. Evaluasi menunjukkan akurasi sebesar 83,33%, dengan rincian pada kelas Lolos diperoleh *precision* 0,76, *recall* 0,93, dan *f1-score* 0,84; pada kelas Play-off diperoleh *precision* 0,60, *recall* 0,50, dan *f1-score* 0,55; serta pada kelas Tidak Lolos diperoleh *precision* 1,00, *recall* 0,88, dan *f1-score* 0,93. Nilai rata-rata makro untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,79, 0,77, dan 0,77, sedangkan rata-rata berbobot sebesar 0,84, 0,83, dan 0,83. Hasil prediksi menunjukkan bahwa Iran, Uni Emirat Arab, Uzbekistan, Korea Selatan, Jepang, dan Australia termasuk kategori lolos langsung; Yordania, Oman, Irak, dan Arab Saudi berada pada kategori play-off; serta Qatar, Kirgizstan, Korea Utara, Kuwait, Palestina, Indonesia, Tiongkok, dan Bahrain tergolong tidak lolos.

DAFTAR PUSTAKA

Amalia, A., Zaidiah, A. & Isnainiyah, I.N. (2022). Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*. 7(2): 496-507.

Bhowmick, K. & Sarvaiya, V. (2021). A Comparative Study of The Different Classification Algorithms on Football Analytics. *International Journal of Advanced Research (IJAR)*. 9(8): 392-407.

Damanik, M.H.U. & Gunawan, D.I. (2024). Penerapan Metode KNN Dalam Data Mining Prediksi Penjualan Produk Pada CV. Chavasindo Berbasis Web. *Jurnal JUREKSI (Jurnal Rekayasa Sistem)*. 2(3): 2123-2135.

Dewi, S.P., Nurwati, N. & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*. 3(4): 639-648.

Dip, A.D., Rahman, N. & Ahmed, M. (2024). Predicting Football Match Results: An Analysis of Feature Selection and Machine Learning Techniques Using A Curated Dataset. *Prosiding 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON)*.: 1-6.

Evitania, C.G. (2023). Implementation Of The K-Nearest Neighbor Algorithm to Predict Air Pollution. *Information Technology and Systems*. 1(1): 45-54.

Karim, A.A., Prasetyo, M.A. & Saputro, M.R. (2023). Perbandingan Metode Random Forest, K-Nearest Neighbor, Dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia. *Prosiding Seminar Nasional Teknologi Dan Sains***.** 2: 377-382.

Maskuri, M.N., Sukerti, K. & Bhakti, R.M.H. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Penyakit Stroke. *Jurnal Ilmiah Intech: Information Technology Journal Of UMUS*. 4(1): 130-140.

Mulyadi, M., Muttaqin, Z., Sormin, M.A., Ribut, R. & Atok, A. (2025). *Pendekatan Kualitatif, Kuantitatif Dan Mixed Methods: Menyusun Metodologi Penelitian Yang Komprehensif*. Indramayu: PT. Adab Indonesia.

Munif, M., Mustain, M. & Yahya, K. (2022). Analysis of the K-Nearest Neighbor Algorithm to Determine the Prediction of Tofu Production. *Applied Technology and Computing Science Journal*.: 50-64.

Rahmadinia, R., Lubis, L., Priansyah, A., R.W.N., Y. & Meutia, T. (2023). Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Mahasiswa Akuntansi Samudra (JMAS)*. 4(4): 223-235.

Rismala, R., Ali, I. & Rinaldi, A.R. (2023). Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris. *JATI (Jurnal Mahasiswa Teknik Informatika)*. 7(1): 585-590.

Urva, G., Desyanti, D., Albanna, I., Sungkar, M.S., Gunawan, I.M.A.O., Adhicandra, I., Ramadhan, S., Rahardian, R.L., Herlawati, H., Handayanto, R.T., Ariana, A.A.G.B., Hartatik, H., Atika, P.D. & Junaidi, S. (2023). *Penerapan Data Mining Di Berbagai Bidang (Konsep, Metode, Dan Studi Kasus)*. Jambi: PT. Sonpedia Publishing Indonesia.

Zein, S. & Gunawan, G. (2022). Prediksi Hasil FIFA World Cup Qatar 2022 Menggunakan Machine Learning Dengan Python. *Jurnal Riset Matematika*. 2(2): 153-162.