

PENERAPAN RANDOM FOREST DALAM KLASIFIKASI FORMASI PERSEBAYA SURABAYA BERDASARKAN KARAKTERISTIK LAWAN PADA LIGA 1 INDONESIA MUSIM 2020–2025

Muhammad Nasrudin^{1*}, Soffa Zahara², Yanuarini Nur Sukmaningtyas³

^{1,2,3} Prodi Informatika, Universitas Islam Majapahit, Jl. Raya Jabon No.KM.0,7, Tambak Rejo, Gayaman, Kec. Mojoanyar, Mojokerto, Jawa Timur 61364.

Keywords:

Random Forest, klasifikasi formasi, machine learning, Persebaya Surabaya, Liga 1 Indonesia.

Correspondent Email:

mnasrudin2002@gmail.com

Abstrak. Perkembangan machine learning telah dimanfaatkan dalam analisis taktik sepak bola, salah satunya untuk mengklasifikasikan formasi berdasarkan karakteristik lawan. Penelitian ini bertujuan menerapkan algoritma Random Forest untuk mengklasifikasikan formasi Persebaya Surabaya pada kompetisi Liga 1 Indonesia musim 2020–2025. Dataset yang digunakan terdiri dari 172 pertandingan yang dikumpulkan dari platform Sofascore dengan memanfaatkan enam variabel karakteristik lawan sebagai fitur masukan. Model dibangun melalui tahapan pra-pemrosesan data, pembagian dataset dengan rasio 80:20, serta evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Random Forest memperoleh accuracy sebesar 87,88%, precision 0,90, recall 0,88, dan F1-score 0,89. Selain itu, persentase penguasaan bola menjadi fitur yang paling berpengaruh dalam proses klasifikasi. Hasil tersebut menunjukkan bahwa Random Forest mampu mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan dengan kinerja yang baik.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. Machine learning has been widely applied in football tactical analysis, including the classification of team formations based on opponent characteristics. This study aims to apply the Random Forest algorithm to classify Persebaya Surabaya's formations in the Indonesian Liga 1 competition during the 2020–2025 seasons. The dataset consisted of 172 matches collected from the Sofascore platform using six opponent-related characteristics as input features. The model was developed through data preprocessing, dataset splitting with an 80:20 ratio, and performance evaluation using a confusion matrix, accuracy, precision, recall, and F1-score. The experimental results showed that the Random Forest model achieved an accuracy of 87.88%, precision of 0.90, recall of 0.88, and F1-score of 0.89. In addition, ball possession percentage was identified as the most influential feature in the classification process. These findings indicate that the Random Forest algorithm is effective in classifying Persebaya Surabaya's formations based on opponent characteristics.

1. PENDAHULUAN

Sepak bola merupakan salah satu cabang olahraga yang mengalami perkembangan pesat dalam pemanfaatan teknologi informasi,

terutama melalui penerapan sports analytics sebagai dasar pengambilan keputusan berbasis data. Ketersediaan data pertandingan yang semakin lengkap memungkinkan klub dan

pelatih melakukan evaluasi performa tim, menganalisis karakteristik lawan, serta menyusun strategi pertandingan secara lebih objektif dibandingkan hanya mengandalkan pengamatan langsung [1]. Perkembangan artificial intelligence (AI) dan machine learning turut memperluas pemanfaatan analisis berbasis data dalam sepak bola karena mampu mengolah data historis untuk mengidentifikasi pola yang mendukung pengambilan keputusan taktis [2]. Berbagai penerapan machine learning telah menunjukkan manfaatnya dalam analisis performa pemain, evaluasi strategi permainan, hingga penyelesaian berbagai permasalahan klasifikasi pada sepak bola [3]. Salah satu keputusan strategis yang memiliki pengaruh besar terhadap jalannya pertandingan adalah pemilihan formasi, karena formasi menentukan keseimbangan permainan dan dapat berubah menyesuaikan karakteristik lawan maupun strategi yang diterapkan pelatih [4].

Salah satu algoritma yang banyak digunakan dalam permasalahan klasifikasi adalah Random Forest. Algoritma ini membangun sejumlah decision tree yang bekerja secara ensemble, sehingga mampu meningkatkan akurasi klasifikasi, mengurangi risiko overfitting, serta menangani data numerik dan kategorikal dengan baik [5]. Keunggulan tersebut telah dibuktikan pada berbagai penelitian. Fitri et al. berhasil menerapkan Random Forest pada sistem klasifikasi kecemasan berbasis web dengan akurasi sebesar 97,71% [6]. Sementara itu, Harahap et al. memanfaatkan algoritma yang sama untuk mendeteksi tautan phishing berbasis web dan menunjukkan performa klasifikasi yang baik [7]. Hasil-hasil tersebut menunjukkan bahwa Random Forest memiliki kemampuan yang andal untuk menyelesaikan berbagai permasalahan klasifikasi dan berpotensi diterapkan pada analisis data pertandingan sepak bola.

Penerapan machine learning dalam bidang sepak bola juga telah berkembang dengan berbagai tujuan penelitian. Triawan et al. menggunakan algoritma Naïve Bayes untuk mengklasifikasikan tipe penyerang berdasarkan statistik pemain [8]. Wisswani et al. mengembangkan klasifikasi tipe gelandang berdasarkan karakteristik statistik pemain [9], sedangkan Yanto menerapkan Naïve Bayes untuk memprediksi juara kompetisi sepak bola menggunakan data historis pertandingan [10].

Penelitian lain oleh Kusmakar et al. memanfaatkan machine learning untuk menganalisis performa tim sepak bola [11]. Di sisi lain, Kim mengembangkan metode identifikasi formasi menggunakan spatiotemporal tracking data [12], sementara Sotudeh et al. memanfaatkan artificial intelligence untuk identifikasi formasi dan analisis taktik pertandingan sepak bola [13]. Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada klasifikasi karakteristik pemain, prediksi hasil pertandingan, atau identifikasi formasi menggunakan tracking data. Penelitian yang mengkaji klasifikasi formasi berdasarkan karakteristik lawan, khususnya pada kompetisi Liga 1 Indonesia, masih sangat terbatas sehingga menjadi peluang untuk dilakukan penelitian lebih lanjut.

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan penerapan algoritma Random Forest untuk mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan menggunakan data historis pertandingan Liga 1 Indonesia musim 2020–2025. Kebaruan penelitian ini terletak pada penggunaan karakteristik lawan sebagai dasar dalam proses klasifikasi formasi pada kompetisi sepak bola Indonesia, yang hingga saat ini masih jarang dibahas pada penelitian sebelumnya. Penelitian ini bertujuan membangun dan mengevaluasi model klasifikasi formasi Persebaya Surabaya menggunakan algoritma Random Forest sebagai pendukung analisis taktik pertandingan.

2. TINJAUAN PUSTAKA

2.1 Machine Learning

Machine learning merupakan salah satu cabang artificial intelligence yang memungkinkan komputer mempelajari pola dari data untuk membangun model yang dapat digunakan dalam proses prediksi maupun klasifikasi [2]. Berdasarkan metode pembelajarannya, machine learning dibedakan menjadi supervised learning, unsupervised learning, dan reinforcement learning [14]. Penelitian ini menggunakan pendekatan supervised learning karena data yang digunakan telah memiliki label berupa formasi Persebaya Surabaya, sehingga model dapat mempelajari

hubungan antara karakteristik lawan dengan formasi yang digunakan [15].

2.2 Random Forest

Random Forest merupakan algoritma ensemble learning yang membangun sekumpulan decision tree menggunakan teknik bootstrap sampling dan pemilihan atribut secara acak. Hasil prediksi dari setiap decision tree kemudian digabungkan melalui mekanisme majority voting untuk menghasilkan keputusan klasifikasi [5].

Random Forest memiliki keunggulan dalam mengurangi risiko overfitting, mampu menangani data numerik maupun kategorikal, serta menyediakan informasi mengenai tingkat kepentingan atribut (feature importance) [16]. Selain itu, algoritma ini telah banyak diterapkan pada berbagai penelitian klasifikasi dan menunjukkan performa yang baik [6], [7].

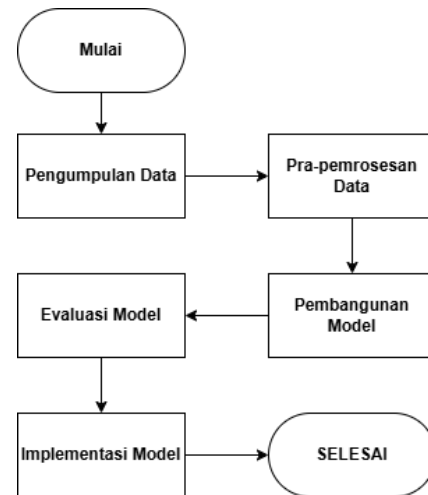
2.3 Klasifikasi

Klasifikasi merupakan salah satu teknik supervised learning yang bertujuan mengelompokkan data ke dalam kelas tertentu berdasarkan pola yang dipelajari dari data latih (training data) [17]. Model klasifikasi digunakan untuk mengenali hubungan antara atribut masukan dan label kelas sehingga dapat menentukan kelas dari data baru. Pada penelitian ini, proses klasifikasi digunakan untuk menentukan formasi Persebaya Surabaya berdasarkan karakteristik lawan yang dihadapi, sehingga setiap data pertandingan dapat dikelompokkan ke dalam salah satu kelas formasi yang telah ditentukan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma Random Forest untuk membangun model klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan pada pertandingan Liga 1 Indonesia musim 2020–2025. Penelitian diawali dengan pengumpulan data pertandingan yang diperoleh dari platform Sofascore, kemudian dilanjutkan dengan proses pra-pemrosesan data untuk mempersiapkan dataset sebelum digunakan dalam pembangunan model. Selanjutnya dilakukan pembangunan model menggunakan algoritma Random Forest yang kemudian dievaluasi

menggunakan beberapa metrik evaluasi klasifikasi. Model yang telah memenuhi kriteria performa selanjutnya diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit sehingga dapat dimanfaatkan untuk mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan. Tahapan penelitian yang dilakukan ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform statistik sepak bola Sofascore. Dataset mencakup seluruh pertandingan Persebaya Surabaya pada kompetisi Liga 1 Indonesia musim 2020–2025 dengan jumlah 172 pertandingan. Rentang waktu tersebut dipilih untuk memperoleh data historis yang representatif sehingga mampu menggambarkan variasi penggunaan formasi Persebaya Surabaya dalam menghadapi berbagai karakteristik lawan selama periode penelitian.

Proses pengumpulan data dilakukan dengan mencatat informasi pertandingan secara manual ke dalam Microsoft Excel sebagai dataset penelitian. Data yang dikumpulkan meliputi informasi mengenai lawan, lokasi pertandingan, penguasaan bola, posisi lawan di klasemen, rata-rata gol yang dicetak lawan, rata-rata gol yang kebobolan lawan, serta formasi yang digunakan oleh Persebaya Surabaya. Dataset yang telah disusun kemudian digunakan sebagai data awal penelitian sebelum memasuki tahap pra-pemrosesan untuk mempersiapkan

data agar sesuai dengan kebutuhan pembangunan model klasifikasi menggunakan algoritma Random Forest.

3.2. *Pra-pemrosesan Data*

Pra-pemrosesan data dilakukan untuk mempersiapkan dataset sebelum digunakan dalam pembangunan model klasifikasi menggunakan algoritma Random Forest. Tahap awal dilakukan dengan memverifikasi kelengkapan dan konsistensi data untuk memastikan tidak terdapat data yang kosong, duplikat, maupun kesalahan pencatatan. Selanjutnya diterapkan injeksi logika (rule-based) untuk menyesuaikan label formasi berdasarkan karakteristik pertandingan yang melibatkan penguasaan bola Persebaya Surabaya, rata-rata gol yang dicetak lawan, dan rata-rata gol yang kebobolan lawan. Data yang memenuhi aturan yang telah ditetapkan akan menggunakan label hasil injeksi, sedangkan data lainnya tetap menggunakan label formasi asli.

Setelah proses tersebut, dilakukan seleksi label dengan mempertahankan tiga formasi yang paling sering digunakan oleh Persebaya Surabaya, yaitu 4-3-3, 4-2-3-1, dan 4-4-2.

Data yang memiliki label formasi di luar ketiga kelas tersebut tidak digunakan pada proses pembangunan model sehingga dataset yang digunakan hanya terdiri atas tiga kelas formasi. Dengan demikian, proses klasifikasi difokuskan pada formasi yang paling dominan digunakan selama periode penelitian.

Selanjutnya dilakukan seleksi fitur untuk menentukan variabel masukan (input) dan keluaran (output) yang digunakan dalam pembangunan model klasifikasi. Variabel yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Variabel yang Digunakan dalam Model

No	Variabel	Jenis Data	Peran
1	Lawan	Kategorikal	Input
2	Lokasi	Kategorikal	Input
3	Penguasaan Bola	Numerik	Input
4	Posisi Lawan di Klasemen	Numerik	Input
5	Rata-rata Gol Dicetak Lawan	Numerik	Input

6	Rata-rata Gol Kebobolan Lawan	Numerik	Input
7	Formasi	Kategorikal	Output

Tahap terakhir adalah transformasi data kategorikal menggunakan teknik Label Encoding pada variabel Lawan, Lokasi, dan Formasi menjadi representasi numerik agar dapat diproses oleh algoritma Random Forest. Dataset hasil pra-pemrosesan selanjutnya dipisahkan menjadi variabel masukan (X) dan variabel keluaran (y). Variabel X terdiri atas atribut Lawan, Lokasi, Penguasaan Bola, Posisi Lawan di Klasemen, Rata-rata Gol Dicetak Lawan, dan Rata-rata Gol Kebobolan Lawan, sedangkan variabel y merupakan atribut Formasi sebagai target klasifikasi. Dataset tersebut kemudian digunakan pada tahap pembangunan model klasifikasi.

3.3. *Pembangunan Model*

Dataset hasil pra-pemrosesan selanjutnya digunakan untuk membangun model klasifikasi menggunakan algoritma Random Forest. Sebelum proses pelatihan dilakukan, dataset dibagi menjadi dua bagian, yaitu 80% sebagai data latih (training set) dan 20% sebagai data uji (testing set). Data latih digunakan untuk mempelajari pola hubungan antara karakteristik lawan dengan formasi yang digunakan Persebaya Surabaya, sedangkan data uji digunakan pada tahap evaluasi untuk mengukur kemampuan model dalam mengklasifikasikan data yang belum pernah dipelajari sebelumnya.

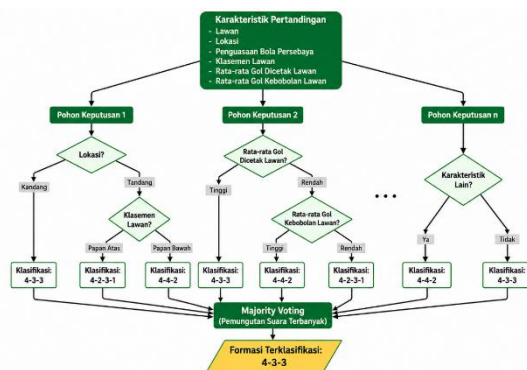
Pada penelitian ini, model Random Forest dibangun menggunakan parameter $n_estimators = 100$, yang menunjukkan jumlah decision tree yang dibentuk selama proses pelatihan. Setiap decision tree dikembangkan menggunakan sampel data yang dipilih secara acak melalui metode bootstrap sampling. Selain itu, pada setiap proses pemisahan simpul (node), algoritma juga memilih subset fitur secara acak sehingga setiap pohon keputusan memiliki struktur yang berbeda. Kombinasi kedua proses tersebut menghasilkan kumpulan decision tree yang beragam, sehingga mampu meningkatkan kemampuan generalisasi model serta mengurangi risiko overfitting.

Pada tahap klasifikasi, setiap decision tree menghasilkan prediksi formasi berdasarkan karakteristik pertandingan yang diberikan

sebagai masukan. Hasil prediksi dari seluruh decision tree kemudian digabungkan menggunakan mekanisme majority voting, yaitu kelas yang memperoleh jumlah suara terbanyak ditetapkan sebagai hasil klasifikasi akhir. Mekanisme tersebut dinyatakan pada Persamaan (1).

$$\hat{Y} \operatorname{argmax}_c \sum_{i=1}^B I(h_i(x) = c) \quad (1)$$

Dengan $\hat{Y}$ merupakan hasil klasifikasi akhir, c adalah kelas yang diprediksi B menyatakan jumlah decision tree, $h_i(x)$ merupakan hasil prediksi dari decision tree ke- i dan $I(h_i(x) = c)$ adalah fungsi indikator yang bernilai 1 apabila prediksi sesuai dengan kelas c dan bernilai 0 apabila tidak sesuai. Mekanisme pembangunan model klasifikasi menggunakan algoritma Random Forest pada penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Mekanisme Klasifikasi Menggunakan Algoritma Random Forest

3.4. Evaluasi Model

Setelah proses pelatihan selesai dilakukan, model dievaluasi menggunakan data uji (testing set) yang telah dipisahkan sebelumnya, yaitu sebesar 20% dari keseluruhan dataset. Evaluasi bertujuan untuk mengukur kemampuan model Random Forest dalam mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan pada data yang belum pernah dipelajari selama proses pelatihan. Hasil prediksi model kemudian dibandingkan dengan label aktual untuk mengetahui tingkat ketepatan klasifikasi yang dihasilkan.

Proses evaluasi dilakukan menggunakan Confusion Matrix, yaitu matriks yang

menggambarkan perbandingan antara hasil prediksi model dengan kondisi sebenarnya. Berdasarkan nilai pada Confusion Matrix, performa model diukur menggunakan empat metrik evaluasi, yaitu Accuracy, Precision, Recall, dan F1-score. Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data uji, Precision menunjukkan tingkat ketepatan model dalam memberikan prediksi pada suatu kelas, Recall mengukur kemampuan model dalam mengenali seluruh data yang termasuk ke dalam kelas yang benar, sedangkan F1-score merupakan rata-rata harmonis antara Precision dan Recall sehingga memberikan ukuran performa yang lebih seimbang. Persamaan yang digunakan untuk menghitung masing-masing metrik evaluasi ditunjukkan pada Persamaan (2) hingga Persamaan (5).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

3.5. Implementasi Model

Model klasifikasi yang telah melalui proses pelatihan dan evaluasi selanjutnya diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit. Implementasi ini bertujuan untuk menerapkan model Random Forest ke dalam sebuah sistem yang dapat digunakan secara interaktif, sehingga pengguna dapat melakukan proses klasifikasi formasi Persebaya Surabaya dengan memasukkan karakteristik lawan tanpa harus menjalankan proses pelatihan model secara langsung.

Aplikasi menerima beberapa variabel masukan yang digunakan dalam proses klasifikasi, yaitu lawan, lokasi pertandingan, penguasaan bola, posisi lawan di klasemen, rata-rata gol yang dicetak lawan, dan rata-rata

gol yang kebobolan lawan. Data masukan tersebut diproses melalui tahapan pra-pemrosesan yang sama seperti pada proses pelatihan, kemudian diteruskan ke model Random Forest yang telah dilatih sebelumnya untuk menghasilkan prediksi formasi.

Hasil klasifikasi ditampilkan dalam bentuk rekomendasi formasi yang diprediksi sesuai dengan karakteristik lawan yang diberikan, yaitu 4-3-3, 4-2-3-1, atau 4-4-2. Dengan demikian, implementasi berbasis web tidak hanya berfungsi sebagai media untuk menguji penerapan model klasifikasi, tetapi juga mempermudah pengguna dalam memanfaatkan model secara praktis melalui antarmuka yang sederhana tanpa memerlukan pengetahuan teknis mengenai algoritma Random Forest.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pra-pemrosesan Data

Tahap pra-pemrosesan data menghasilkan dataset yang telah siap digunakan dalam proses pembangunan model klasifikasi menggunakan algoritma Random Forest. Berdasarkan hasil verifikasi, dataset yang terdiri atas 172 pertandingan tidak memiliki data yang kosong maupun duplikat sehingga seluruh data dapat digunakan pada tahap selanjutnya. Penerapan injeksi logika (rule-based) berhasil menyesuaikan label formasi berdasarkan karakteristik pertandingan, kemudian dilakukan seleksi label dengan mempertahankan tiga formasi yang paling sering digunakan oleh Persebaya Surabaya, yaitu 4-3-3, 4-2-3-1, dan 4-4-2. Hasil seleksi label menunjukkan bahwa sebanyak 165 data memenuhi kriteria untuk digunakan pada proses pembangunan model, sedangkan data dengan formasi di luar ketiga kelas tersebut tidak disertakan pada tahap selanjutnya. Selanjutnya, dilakukan seleksi fitur sehingga diperoleh enam variabel masukan (input), yaitu Lawan, Lokasi, Penguasaan Bola, Posisi Lawan di Klasemen, Rata-rata Gol Dicetak Lawan, dan Rata-rata Gol Kebobolan Lawan, serta satu variabel keluaran (output), yaitu Formasi.

Sebelum dilakukan proses Label Encoding, variabel yang bersifat kategorikal masih berupa nilai teks, seperti nama lawan, lokasi pertandingan, dan formasi yang digunakan. Contoh dataset sebelum proses Label Encoding ditunjukkan pada Tabel 2.

Tabel 2. Dataset Sebelum Label Encoding

No	Lawan	Lokasi	Formasi
1	Persik Kediri	Home	4-3-3
2	Persipura Jayapura	Home	4-3-3
3	Borneo FC	Away	4-3-3
4	Persikabo 1973	Home	4-3-3
5	PSM Makassar	Away	4-3-3
6	Bhayangkara FC	Home	4-2-3-1
7	PSS Sleman	Away	4-2-3-1
8	PSIS Semarang	Home	4-2-3-1
9	Persipura Jayapura	Away	4-3-3
10	Persija Jakarta	Home	4-3-3

Selanjutnya, variabel kategorikal, yaitu Lawan, Lokasi, dan Formasi, ditransformasikan ke dalam bentuk numerik menggunakan teknik Label Encoding. Transformasi ini bertujuan agar data kategorikal dapat diproses oleh algoritma Random Forest. Hasil transformasi tersebut ditunjukkan pada Tabel 3.

Tabel 3. Dataset Setelah Label Encoding

No	Lawan	Lokasi	Formasi
1	17	1	6
2	19	1	6
3	4	0	6
4	18	1	6
5	19	0	6
6	3	1	4
7	12	0	4
8	10	1	4
9	19	0	6
10	13	1	6

Hasil Label Encoding menunjukkan bahwa seluruh variabel kategorikal telah berhasil ditransformasikan menjadi representasi numerik tanpa mengubah informasi yang terkandung pada setiap kategori. Dengan demikian, seluruh data telah memenuhi format yang diperlukan oleh algoritma Random Forest. Dataset hasil pra-pemrosesan selanjutnya dibagi menjadi data latih (training data) dan data uji (testing data) sebelum digunakan pada tahap pembangunan model klasifikasi.

4.2 Hasil Pembagian Dataset

Setelah melalui tahap pra-pemrosesan, dataset yang berjumlah 165 data selanjutnya

dibagi menjadi data latih (training data) dan data uji (testing data). Pembagian dataset dilakukan menggunakan rasio 80:20, di mana 80% data digunakan sebagai data latih untuk membangun model klasifikasi, sedangkan 20% sisanya digunakan sebagai data uji untuk mengevaluasi kinerja model terhadap data yang belum pernah dipelajari sebelumnya. Hasil pembagian dataset ditunjukkan pada Tabel 4.

Tabel 4. Pembagian Dataset

Dataset	Jumlah Data	Persentase
Data Latih (<i>Training Data</i>)	132	80%
Data Uji (<i>Testing Data</i>)	33	20%
Total	165	100%

Berdasarkan hasil pembagian tersebut, sebanyak 132 data digunakan sebagai data latih untuk membangun model Random Forest, sedangkan 33 data digunakan sebagai data uji untuk mengukur kemampuan model dalam melakukan klasifikasi terhadap data baru. Penggunaan rasio 80:20 memberikan jumlah data latih yang cukup untuk mempelajari pola hubungan antara karakteristik lawan dan formasi yang digunakan Persebaya Surabaya, sekaligus menyediakan data uji yang representatif untuk mengevaluasi kemampuan generalisasi model. Dataset yang telah dibagi tersebut selanjutnya digunakan pada tahap pembangunan model klasifikasi menggunakan algoritma Random Forest.

4.3 Hasil Pembangunan Model Random Forest

Setelah dataset dibagi menjadi data latih dan data uji, tahap selanjutnya adalah pembangunan model klasifikasi menggunakan algoritma Random Forest. Proses pelatihan model dilakukan menggunakan 132 data latih dengan memanfaatkan enam variabel masukan (input), yaitu Lawan, Lokasi, Penguasaan Bola, Posisi Lawan di Klasemen, Rata-rata Gol Diketuk Lawan, dan Rata-rata Gol Kebobolan Lawan, serta satu variabel keluaran (output), yaitu Formasi. Pada penelitian ini, model dibangun menggunakan parameter $n_estimators = 100$, yang berarti model terdiri atas 100 pohon keputusan (decision tree). Konfigurasi

parameter yang digunakan dalam pembangunan model ditunjukkan pada Tabel 5.

Tabel 5. Konfigurasi Pembangunan Model Random Forest

Parameter	Nilai
Algoritma	<i>Random Forest Classifier</i>
Jumlah Data Latih	132
Jumlah Data Uji	33
Jumlah Variabel Masukan	6
Variabel Keluaran	Formasi
$n_estimators$	100

Berdasarkan konfigurasi tersebut, proses pelatihan berhasil menghasilkan model klasifikasi yang mampu mempelajari hubungan antara karakteristik lawan dan formasi yang digunakan oleh Persebaya Surabaya. Pada algoritma Random Forest, setiap decision tree dibangun menggunakan sampel data hasil bootstrap dan subset fitur yang dipilih secara acak. Selanjutnya, hasil prediksi dari seluruh decision tree digabungkan menggunakan mekanisme majority voting untuk menentukan kelas formasi akhir. Pendekatan ini membantu meningkatkan stabilitas model dan mengurangi risiko overfitting, sehingga model memiliki kemampuan yang lebih baik dalam melakukan klasifikasi terhadap data baru. Model yang telah berhasil dibangun selanjutnya dievaluasi menggunakan data uji untuk mengetahui tingkat kinerjanya.

4.4 Hasil Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur kinerja algoritma Random Forest dalam mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan. Evaluasi dilakukan menggunakan 33 data uji (testing data) yang tidak digunakan pada proses pelatihan model. Pengujian terhadap data yang belum pernah dipelajari bertujuan untuk mengetahui kemampuan model dalam melakukan klasifikasi pada data baru serta mengukur kemampuan generalisasi model. Pada penelitian ini, evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil evaluasi model ditunjukkan pada Gambar 3.

```

***
=== HASIL EVALUASI MODEL FINAL ===
Akurasi Model: 87.88%

=== LAPORAN KLASIFIKASI ===

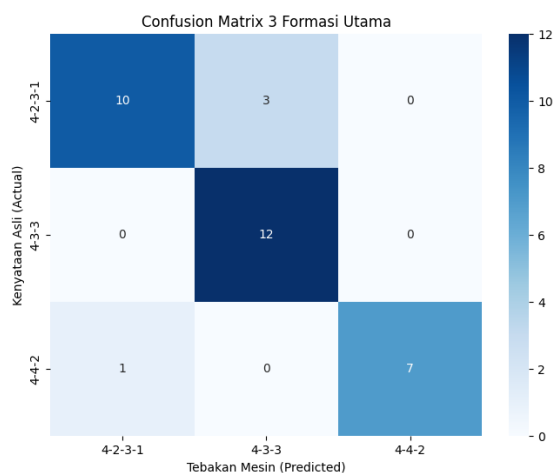
```

	precision	recall	f1-score	support
4-2-3-1	0.91	0.77	0.83	13
4-3-3	0.80	1.00	0.89	12
4-4-2	1.00	0.88	0.93	8
accuracy			0.88	33
macro avg	0.90	0.88	0.89	33
weighted avg	0.89	0.88	0.88	33

Gambar 3. Classification Report

Berdasarkan hasil evaluasi, model Random Forest memperoleh nilai accuracy sebesar 87,88%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan 29 dari 33 data uji dengan benar. Selain itu, hasil classification report menunjukkan bahwa setiap kelas formasi memiliki performa yang baik. Formasi 4-2-3-1 memperoleh nilai precision sebesar 0,91, recall sebesar 0,77, dan F1-score sebesar 0,83. Formasi 4-3-3 memperoleh precision sebesar 0,80, recall sebesar 1,00, dan F1-score sebesar 0,89. Sementara itu, formasi 4-4-2 menunjukkan performa terbaik dengan nilai precision sebesar 1,00, recall sebesar 0,88, dan F1-score sebesar 0,93.

Untuk memberikan gambaran yang lebih rinci mengenai hasil klasifikasi, digunakan confusion matrix yang menunjukkan jumlah data yang berhasil diprediksi dengan benar maupun data yang mengalami salah klasifikasi pada setiap kelas formasi. Hasil confusion matrix ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix

Berdasarkan confusion matrix, model berhasil mengklasifikasikan 10 dari 13 data formasi 4-2-3-1 dengan benar, sedangkan 3 data

lainnya diprediksi sebagai formasi 4-3-3. Pada kelas 4-3-3, seluruh 12 data berhasil diklasifikasikan dengan benar tanpa adanya kesalahan prediksi, sehingga menghasilkan nilai recall sebesar 1,00. Sementara itu, pada kelas 4-4-2, model berhasil mengklasifikasikan 7 dari 8 data dengan benar, sedangkan 1 data diprediksi sebagai formasi 4-2-3-1.

Perbedaan nilai precision, recall, dan F1-score pada masing-masing kelas menunjukkan bahwa kemampuan model dalam mengenali setiap formasi tidak sepenuhnya sama. Formasi 4-4-2 memperoleh nilai precision sebesar 1,00, yang menunjukkan bahwa seluruh prediksi model terhadap kelas tersebut merupakan prediksi yang benar tanpa menghasilkan false positive. Di sisi lain, formasi 4-3-3 memperoleh nilai recall sebesar 1,00, yang menunjukkan bahwa seluruh data aktual pada kelas tersebut berhasil dikenali oleh model. Sementara itu, formasi 4-2-3-1 memiliki nilai recall yang lebih rendah, yaitu 0,77, karena terdapat beberapa data yang memiliki karakteristik serupa dengan formasi 4-3-3, sehingga model masih mengalami kesalahan dalam membedakan kedua kelas tersebut.

Hasil evaluasi ini menunjukkan bahwa enam variabel masukan yang digunakan, yaitu Lawan, Lokasi, Penguasaan Bola, Posisi Lawan di Klasemen, Rata-rata Gol Dicetak Lawan, dan Rata-rata Gol Kebobolan Lawan, mampu memberikan informasi yang cukup bagi algoritma Random Forest untuk mempelajari pola penggunaan formasi. Mekanisme pembentukan banyak decision tree yang dikombinasikan melalui proses majority voting juga berkontribusi dalam meningkatkan stabilitas model sehingga mampu menghasilkan prediksi yang konsisten pada sebagian besar data uji.

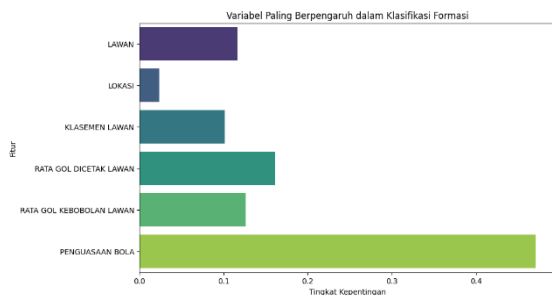
Secara keseluruhan, nilai accuracy sebesar 87,88% menunjukkan bahwa algoritma Random Forest memiliki kemampuan yang baik dalam mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan. Tingginya nilai precision, recall, dan F1-score pada sebagian besar kelas menunjukkan bahwa model mampu mengenali pola penggunaan formasi dengan baik, meskipun masih terdapat beberapa kesalahan klasifikasi pada formasi yang memiliki karakteristik pertandingan yang serupa. Dengan demikian, model yang dibangun dinilai memiliki performa

yang baik dan layak untuk diimplementasikan pada sistem klasifikasi formasi berbasis web.

4.5 Hasil Analisis Feature Importance

Analisis feature importance dilakukan untuk mengetahui tingkat kontribusi masing-masing variabel masukan dalam proses klasifikasi formasi Persebaya Surabaya menggunakan algoritma Random Forest. Semakin besar nilai feature importance, semakin besar pula pengaruh variabel tersebut dalam proses pembentukan decision tree dan penentuan hasil klasifikasi.

Untuk mempermudah interpretasi hasil analisis, nilai feature importance pada setiap variabel divisualisasikan dalam bentuk grafik sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Grafik Feature Importance

Berdasarkan grafik pada Gambar 5, terlihat bahwa setiap variabel memiliki tingkat kepentingan yang berbeda dalam proses klasifikasi. Variabel Penguasaan Bola memiliki kontribusi paling tinggi, sedangkan variabel Lokasi memiliki kontribusi paling rendah. Nilai feature importance masing-masing variabel disajikan secara rinci pada Tabel 6.

Tabel 6. Nilai Feature Importance

Variabel	Nilai <i>Feature Importance</i>
Penguasaan Bola	0,471
Rata-rata Gol Dicitak Lawan	0,161
Rata-rata Gol Kebobolan Lawan	0,126
Lawan	0,116
Posisi Lawan di Klasemen	0,101
Lokasi	0,023

Berdasarkan Tabel 6, variabel Penguasaan Bola memiliki nilai feature importance tertinggi, yaitu 0,471, sehingga menjadi variabel yang paling berpengaruh dalam proses klasifikasi formasi. Hasil ini menunjukkan bahwa tingkat penguasaan bola merupakan indikator utama yang dipelajari oleh model dalam membedakan penggunaan formasi Persebaya Surabaya. Variabel Rata-rata Gol Dicitak Lawan (0,161) dan Rata-rata Gol Kebobolan Lawan (0,126) juga memberikan kontribusi yang cukup besar, menunjukkan bahwa performa menyerang dan bertahan lawan turut memengaruhi hasil klasifikasi. Sementara itu, variabel Lawan (0,116) dan Posisi Lawan di Klasemen (0,101) memiliki kontribusi sedang, sedangkan variabel Lokasi (0,023) memberikan pengaruh paling kecil dibandingkan variabel lainnya.

Secara keseluruhan, hasil analisis feature importance menunjukkan bahwa model Random Forest lebih banyak memanfaatkan variabel yang berkaitan dengan performa pertandingan dibandingkan variabel kontekstual, seperti lokasi pertandingan. Dominannya variabel Penguasaan Bola menunjukkan bahwa aspek penguasaan permainan menjadi faktor utama dalam proses klasifikasi formasi Persebaya Surabaya. Temuan ini juga mendukung hasil evaluasi model yang memperoleh tingkat akurasi sebesar 87,88%, sehingga menunjukkan bahwa kombinasi enam variabel yang digunakan telah mampu merepresentasikan karakteristik pertandingan dengan baik.

4.6 Implementasi Sistem Berbasis Web

Setelah model Random Forest berhasil dibangun dan melalui tahap evaluasi, langkah selanjutnya adalah mengimplementasikan model tersebut ke dalam sistem berbasis web menggunakan Streamlit. Implementasi ini bertujuan agar model yang telah dikembangkan dapat digunakan secara langsung oleh pengguna melalui antarmuka yang sederhana dan mudah dipahami. Dengan demikian, pengguna dapat melakukan klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan tanpa harus menjalankan kode program atau memahami proses komputasi yang terjadi di dalam sistem.

Tampilan awal aplikasi ditunjukkan pada Gambar 6. Halaman ini merupakan tampilan utama yang akan muncul ketika aplikasi dijalankan. Pada halaman tersebut ditampilkan judul aplikasi, deskripsi singkat mengenai tujuan sistem, serta menu yang mengarahkan pengguna untuk memulai proses klasifikasi. Tampilan antarmuka dirancang secara sederhana agar pengguna dapat memahami fungsi aplikasi dan mengoperasikannya dengan mudah.



Gambar 6. Halaman Awal Aplikasi

Selanjutnya, pengguna diarahkan menuju halaman input yang ditunjukkan pada Gambar 7. Pada halaman ini, pengguna diminta untuk mengisi enam variabel masukan yang digunakan dalam proses klasifikasi, yaitu Lawan, Lokasi pertandingan, Posisi lawan di klasemen, Penguasaan bola, Rata-rata gol yang dicetak lawan, dan Rata-rata gol yang kebocoran lawan. Keenam variabel tersebut merupakan fitur yang digunakan pada tahap pelatihan model Random Forest. Setelah seluruh data diisi, sistem akan menerima masukan tersebut untuk diproses pada tahap klasifikasi.



Gambar 7. Halaman Input Data

Proses klasifikasi dilakukan ketika pengguna menekan tombol Klasifikasi. Selanjutnya, sistem akan memproses data masukan menggunakan model Random Forest yang telah dilatih sebelumnya. Berdasarkan pola yang dipelajari selama proses pelatihan, model akan menghasilkan klasifikasi formasi yang sesuai dengan karakteristik lawan.

Hasil proses tersebut kemudian ditampilkan pada halaman hasil sebagaimana ditunjukkan pada Gambar 8.

Gambar 8. Halaman Hasil Klasifikasi



Halaman hasil menampilkan formasi Persebaya Surabaya yang berhasil diklasifikasikan berdasarkan data yang telah dimasukkan oleh pengguna. Proses klasifikasi berlangsung secara otomatis sehingga hasil dapat ditampilkan dalam waktu yang relatif singkat. Hal ini menunjukkan bahwa model Random Forest telah berhasil diintegrasikan ke dalam aplikasi berbasis web menggunakan Streamlit dan dapat dimanfaatkan untuk melakukan proses klasifikasi tanpa perlu melakukan pelatihan model kembali setiap kali aplikasi digunakan.

Secara keseluruhan, implementasi sistem berbasis web menggunakan Streamlit berhasil menyediakan media yang sederhana dan mudah digunakan dalam melakukan klasifikasi formasi Persebaya Surabaya. Pengguna hanya perlu memasukkan enam variabel masukan sesuai karakteristik lawan, kemudian sistem akan memproses data tersebut menggunakan model Random Forest dan menampilkan hasil klasifikasi secara otomatis. Dengan demikian, hasil penelitian ini tidak hanya menghasilkan model klasifikasi yang memiliki performa yang baik, tetapi juga menghasilkan sebuah aplikasi berbasis web yang mampu mengimplementasikan model tersebut sehingga dapat digunakan sebagai sarana pendukung dalam proses klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan.

4.7 Hasil Pengujian Black Box

Pengujian Black Box dilakukan untuk memastikan bahwa seluruh fitur pada aplikasi berbasis web dapat berfungsi sesuai dengan kebutuhan fungsional yang telah dirancang. Pengujian ini berfokus pada pengujian fungsi sistem berdasarkan masukan (input) yang

diberikan pengguna dan keluaran (output) yang dihasilkan oleh aplikasi, tanpa memperhatikan proses internal atau kode program yang digunakan. Setiap fitur diuji menggunakan beberapa skenario untuk mengetahui apakah sistem mampu menjalankan fungsinya dengan benar, mulai dari menampilkan halaman aplikasi, menerima data karakteristik lawan, memproses data menggunakan model Random Forest, hingga menampilkan hasil klasifikasi kepada pengguna. Hasil pengujian terhadap setiap fitur aplikasi disajikan pada Tabel 7.

Tabel 7. Hasil Pengujian Blackbox

Fitur	Skenario	Hasil	Status
Halaman input	Membuka aplikasi	Halaman tampil	Berhasil
Input data	Mengisi data	Data diterima	Berhasil
Klasifikasi	Menekan tombol klasifikasi	Proses klasifikasi berhasil	Berhasil
Hasil	Menampilkan hasil	Formasi berhasil ditampilkan	Berhasil

Berdasarkan hasil pengujian pada Tabel 7, seluruh fitur pada aplikasi berhasil berfungsi sesuai dengan hasil yang diharapkan. Halaman input dapat ditampilkan dengan baik ketika aplikasi dijalankan, kemudian sistem mampu menerima seluruh data karakteristik lawan yang dimasukkan oleh pengguna. Setelah tombol klasifikasi ditekan, model Random Forest berhasil memproses data masukan dan menghasilkan klasifikasi formasi sesuai dengan pola yang telah dipelajari selama proses pelatihan. Hasil ini menunjukkan bahwa setiap tahapan pada aplikasi, mulai dari proses input hingga pemrosesan data, telah berjalan dengan baik sesuai dengan rancangan sistem yang telah dibuat.

Selain itu, hasil klasifikasi dapat ditampilkan dengan baik pada halaman hasil tanpa ditemukan kesalahan fungsional selama proses pengujian. Seluruh proses berlangsung secara otomatis sehingga pengguna dapat memperoleh hasil klasifikasi dalam waktu yang singkat setelah data dimasukkan. Hal ini menunjukkan bahwa integrasi antara antarmuka aplikasi berbasis Streamlit dengan model Random Forest telah berjalan dengan baik sehingga seluruh alur penggunaan aplikasi, mulai dari

memasukkan data hingga memperoleh hasil klasifikasi, dapat dilakukan secara lancar dan konsisten.

Secara keseluruhan, hasil pengujian Black Box menunjukkan bahwa seluruh fungsi utama aplikasi telah berjalan sesuai dengan kebutuhan fungsional. Tidak ditemukan kegagalan pada setiap skenario pengujian yang dilakukan, sehingga aplikasi mampu menjalankan seluruh proses sesuai dengan tujuan implementasinya. Sistem berhasil menerima masukan dari pengguna, melakukan proses klasifikasi menggunakan model Random Forest, dan menampilkan hasil klasifikasi tanpa mengalami kesalahan fungsional. Dengan demikian, aplikasi yang dikembangkan telah memenuhi tujuan implementasi sebagai sistem berbasis web untuk melakukan klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan serta siap digunakan sebagai media untuk membantu proses klasifikasi secara praktis dan mudah diakses oleh pengguna. Hasil pengujian ini menunjukkan bahwa integrasi antara model Random Forest dan antarmuka aplikasi berjalan sesuai dengan rancangan sistem sehingga aplikasi layak digunakan sebagai media klasifikasi formasi.

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mulai dari tahap pengumpulan data, pra-pemrosesan data, pembangunan model, evaluasi model, hingga implementasi sistem berbasis web, dapat disimpulkan bahwa:

1. Algoritma Random Forest berhasil diterapkan untuk mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan pada Liga 1 Indonesia musim 2020–2025 menggunakan enam variabel masukan, yaitu Lawan, Lokasi, Posisi Lawan di Klasemen, Penguasaan Bola, Rata-rata Gol Dicetak Lawan, dan Rata-rata Gol Kebobolan Lawan.
2. Model Random Forest berhasil memberikan hasil klasifikasi dengan tingkat akurasi sebesar 87,88% pada data pengujian. Hasil analisis feature importance menunjukkan bahwa variabel Penguasaan Bola merupakan faktor yang paling berpengaruh dalam proses klasifikasi, diikuti oleh Rata-rata Gol Dicetak Lawan, Rata-rata Gol

Kebobolan Lawan, Lawan, Posisi Lawan di Klasemen, dan Lokasi.

3. Model yang telah dibangun berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit. Hasil pengujian Black Box menunjukkan bahwa seluruh fitur aplikasi telah berjalan sesuai dengan kebutuhan fungsional, mulai dari proses input data, klasifikasi menggunakan model Random Forest, hingga penampilan hasil klasifikasi.

Secara keseluruhan, penelitian ini berhasil mencapai tujuan penelitian, yaitu menerapkan algoritma Random Forest untuk mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan serta mengimplementasikannya ke dalam aplikasi berbasis web. Sistem yang dikembangkan mampu memberikan hasil klasifikasi dengan baik dan dapat digunakan sebagai media untuk membantu proses klasifikasi formasi secara lebih cepat dan mudah.

Berikut adalah beberapa saran untuk pengembangan penelitian selanjutnya:

1. Menambahkan jumlah data pertandingan dan memperluas cakupan data agar model dapat mempelajari pola permainan yang lebih beragam sehingga berpotensi meningkatkan performa klasifikasi.
2. Menambahkan variabel lain yang berpotensi memengaruhi pemilihan formasi, seperti kondisi pemain, riwayat pertemuan antar tim, statistik pertandingan, atau faktor taktikal lainnya sehingga hasil klasifikasi menjadi lebih komprehensif.
3. Membandingkan performa algoritma Random Forest dengan algoritma klasifikasi lainnya, seperti XGBoost, Support Vector Machine (SVM), atau Gradient Boosting, untuk mengetahui metode yang memberikan kinerja terbaik pada kasus klasifikasi formasi sepak bola.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis menyampaikan terima kasih kepada dosen pembimbing atas

bimbingan dan arahan yang diberikan selama penelitian, kepada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Majapahit atas dukungannya, serta kepada keluarga, teman-teman, dan kekasih tercinta atas doa, motivasi, dan semangat yang diberikan selama proses penyusunan penelitian ini.

DAFTAR PUSTAKA

- [1] Hidayat, "Analitik prediktif sepakbola : model machine learning bri liga 1 indonesia Soccer predictive analytics : bri liga 1 indonesia machine learning models," vol. 23, no. 4, pp. 386–399, 2024.
- [2] J. E. Teixeira, E. Maio, and P. Afonso, "Mapping football tactical behavior and collective dynamics with artificial intelligence : a systematic review," no. May, 2025, doi: 10.3389/fspor.2025.1569155.
- [3] K. Fujii, *Machine Learning in Sports*. 2024.
- [4] S. Plakias, T. Tsatalas, V. Armatas, and D. Tsaopoulos, "Tactical Situations and Playing Styles as Key Performance Indicators in Soccer," 2024.
- [5] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.
- [6] R. R. Fitri *et al.*, "PENGUNAAN RANDOM FOREST DALAM SISTEM," vol. 13, no. 3.
- [7] A. D. Harahap, D. Juardi, A. Susilo, Y. Irawan, U. S. Karawang, and T. Timur, "PHISHING MENGGUNAKAN ALGORITMA RANDOM," vol. 12, no. 3, 2024.
- [8] E. Triawan, N. Suarna, and A. R. Dikananda, "KLASIFIKASI TIPE PENYERANG SEPAK BOLA LIGA INGGRIS BERDASARKAN DATA STATISTIK PEMAIN MENGGUNAKAN METODE NAIVE BAYES," vol. 8, no. 2, pp. 1809–1814, 2024.
- [9] N. W. Wisswani, T. R. Tulili, M. F. Andrijasa, and M. R. S. N., "KLASIFIKASI TIPE GELANDANG SEPAK BOLA BERDASARKAN," vol. 2, no. July, pp. 46–50, 2020.
- [10] A. Yanto, "Implementasi Metode Klasifikasi Naïve Bayes Untuk Memprediksi Juara La Liga," vol. 5, no. May, pp. 128–139, 2024, doi: 10.35957/jtsi.v5i2.8028.
- [11] S. Kusmakar, S. Shelyag, D. A. N. Dwyer, P. Gastin, and M. Angelova, "Machine Learning Enabled Team Performance Analysis in the Dynamical Environment of Soccer," vol. 8, 2020, doi: 10.1109/ACCESS.2020.2992025.
- [12] H. Kim, "SoccerCPD : Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data," 2023, doi: 10.1145/3534678.3539150.

- [13] H. Sotudeh, "formation identi fi cation in," vol. 1, no. February, pp. 1–15, 2025, doi: 10.3389/fspor.2024.1512386.
- [14] A. A. Permana *et al.*, *Machine learning*. 2023.
- [15] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning," 2021.
- [16] Genuer & Poggi, *Random Forests with R*. 2020.
- [17] J. Han, *DATA MINING Concepts and Techniques*. 2012.