

## PREDIKSI PELUANG JUARA PIALA DUNIA FIFA 2026 MENGGUNAKAN ALGORITMA RANDOM FOREST DAN SIMULASI MONTE CARLO BERDASARKAN DATA STATISTIK PERFORMA TIM NASIONAL

Mohamad Iqbal Ulumando

Institut Teknologi Alberth Foenay Kupang, Indonesia

[iqbal77ulumando@gmail.com](mailto:iqbal77ulumando@gmail.com)

### ABSTRACT

*The FIFA World Cup is the most prestigious international football competition, bringing together the best national teams from various countries. The 2026 FIFA World Cup introduces a new format featuring 48 teams, thereby increasing the complexity of predicting the eventual champion. This study aims to predict the winning probabilities for the 2026 FIFA World Cup by combining the Random Forest algorithm with Monte Carlo simulation, based on national team performance statistics. The dataset comprises the 48 participating teams, analyzed using eight variables: FIFA Ranking, Elo Rating, win percentage, average goals scored per match, average goals conceded per match, goal difference, squad market value, and past World Cup performance. The data underwent a preprocessing stage—specifically normalization—before being used to construct the Random Forest model. Model evaluation results yielded an accuracy of 84.50%, precision of 82.30%, recall of 81.70%, and an F1-score of 82.00%. Subsequently, a Monte Carlo simulation was conducted over 10,000 iterations to estimate the winning probability for each national team. The results indicate that Morocco has the highest probability of winning at 14.82%, followed by France (13.97%), Argentina (12.76%), Egypt (12.03%), and Brazil (10.88%). The findings demonstrate that the combination of the Random Forest algorithm and Monte Carlo simulation can provide objective, data-driven predictions regarding championship probabilities for international football competitions.*

### Keywords:

*Random Forest, Monte Carlo Simulation, 2026 FIFA World Cup, Winner Prediction, Machine Learning.*

### PENDAHULUAN

Sepak bola merupakan olahraga paling populer di dunia yang memiliki jutaan penggemar dan menjadi salah satu industri olahraga terbesar secara global. Salah satu kompetisi sepak bola paling bergengsi adalah Piala Dunia FIFA yang diselenggarakan setiap empat tahun sekali dan mempertemukan tim nasional terbaik dari berbagai negara (Aje Wicaksono & Maximianus Agus Prayudi, 2024). Piala Dunia tidak hanya menjadi ajang kompetisi olahraga, tetapi juga menarik perhatian masyarakat, akademisi, analis data, dan pelaku industri olahraga untuk melakukan berbagai analisis dan prediksi terkait performa tim peserta serta peluang mereka untuk meraih gelar juara (Taufik & Nuryananda, 2023).

Perkembangan teknologi informasi dan ketersediaan data statistik olahraga yang semakin lengkap telah mendorong pemanfaatan pendekatan berbasis data dalam pengambilan keputusan dan prediksi hasil pertandingan (Bafadal et al., 2024). Berbagai data seperti peringkat FIFA, jumlah kemenangan, produktivitas gol, tingkat kebobolan, performa pertandingan sebelumnya, hingga nilai pasar pemain dapat digunakan untuk menganalisis kekuatan relatif setiap tim nasional. Pemanfaatan data tersebut memungkinkan proses prediksi dilakukan secara lebih objektif dibandingkan dengan pendekatan konvensional yang hanya mengandalkan opini atau intuisi pengamat sepak bola (Pratama et al., 2023).

Dalam beberapa tahun terakhir, metode Machine Learning telah banyak digunakan untuk memprediksi hasil pertandingan olahraga. Algoritma seperti Logistic Regression, Decision Tree, Support Vector Machine, Neural Network, dan Random Forest telah menunjukkan kemampuan yang baik dalam mengidentifikasi pola tersembunyi dari data historis pertandingan (Ade Ryan Pratama et al., 2025). Di antara berbagai metode tersebut, Random Forest memiliki keunggulan dalam menangani data berdimensi tinggi, mengurangi risiko overfitting, serta mampu menghasilkan tingkat akurasi yang baik pada berbagai kasus klasifikasi dan prediksi (Hidayat et al., 2025). Meskipun demikian, sebagian besar penelitian terdahulu hanya berfokus pada prediksi hasil pertandingan individual, seperti kemenangan, kekalahan, atau hasil seri dalam satu pertandingan tertentu (Pratama et al., 2023).

Pendekatan tersebut belum sepenuhnya mampu menggambarkan peluang suatu tim untuk menjadi juara dalam sebuah turnamen besar yang terdiri dari banyak pertandingan dan berbagai kemungkinan skenario. Oleh karena itu, diperlukan pendekatan tambahan yang mampu mensimulasikan jalannya turnamen secara menyeluruh berdasarkan probabilitas yang dihasilkan oleh model prediksi (Efendi et al., 2024). Salah satu metode yang dapat digunakan untuk mengatasi permasalahan tersebut adalah Simulasi Monte Carlo (Anastasia Hutabarat et al., 2025). Metode ini memungkinkan proses simulasi dilakukan secara berulang dalam jumlah besar untuk menghasilkan estimasi

probabilitas berbagai kemungkinan hasil yang dapat terjadi selama turnamen berlangsung (Marpaung & Marpaung, 2025). Dengan mengombinasikan algoritma Random Forest sebagai model prediksi pertandingan dan Simulasi Monte Carlo sebagai mekanisme simulasi turnamen, peluang setiap tim nasional untuk menjadi juara dapat dihitung secara lebih komprehensif dan realistis (Hanif Al Azzam et al., 2025).

Piala Dunia FIFA 2026 menjadi menarik untuk diteliti karena merupakan edisi pertama yang menggunakan format baru dengan jumlah peserta sebanyak 48 tim nasional (FIFA, 2026a). Perubahan format kompetisi ini menyebabkan meningkatnya kompleksitas jalannya turnamen serta bertambahnya jumlah kemungkinan hasil yang dapat terjadi (Pratama et al., 2023). Kondisi tersebut menjadikan pendekatan berbasis Machine Learning dan simulasi probabilistik semakin relevan untuk digunakan dalam melakukan analisis peluang juara (Putra & Harahap, 2024).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membangun model prediksi peluang juara Piala Dunia FIFA 2026 menggunakan algoritma Random Forest dan Simulasi Monte Carlo berdasarkan statistik performa tim nasional. Hasil penelitian diharapkan dapat memberikan gambaran probabilitas keberhasilan masing-masing negara peserta serta menunjukkan efektivitas kombinasi metode Machine Learning dan simulasi probabilistik dalam melakukan prediksi pada kompetisi olahraga berskala internasional.

## TINJAUAN PUSTAKA

### Piala Dunia FIFA

Piala Dunia FIFA merupakan kompetisi sepak bola internasional yang diselenggarakan oleh Fédération Internationale de Football Association (FIFA) setiap empat tahun sekali (Ajie Wicaksono & Maximianus Agus Prayudi, 2024). Turnamen ini mempertemukan tim nasional terbaik dari berbagai negara yang berhasil lolos melalui proses kualifikasi pada masing-masing konfederasi sepak bola dunia. Sejak pertama kali diselenggarakan pada tahun 1930 di Uruguay, Piala Dunia FIFA telah berkembang menjadi ajang olahraga paling bergengsi di dunia dan menjadi salah satu peristiwa olahraga yang paling banyak disaksikan oleh masyarakat internasional (FIFA, 2026b).

Piala Dunia FIFA 2026 merupakan edisi ke-23 dalam sejarah penyelenggaraan turnamen ini dan menjadi penyelenggaraan pertama yang menggunakan format 48 tim peserta (Britannica, 2026). Selain itu, Piala Dunia 2026 juga mencatat sejarah sebagai turnamen pertama yang diselenggarakan oleh tiga negara tuan rumah secara bersamaan, yaitu Amerika Serikat, Kanada, dan Meksiko. Ketiga negara tersebut memperoleh hak penyelenggaraan setelah memenangkan proses bidding FIFA dan secara otomatis lolos sebagai peserta turnamen (AlJazeera, 2026). Amerika Serikat menjadi negara dengan jumlah stadion dan kota penyelenggara terbanyak, Meksiko menjadi negara pertama yang tiga kali menjadi tuan rumah Piala Dunia setelah sebelumnya menjadi penyelenggara pada tahun 1970 dan 1986, sedangkan Kanada menjadi salah satu negara yang untuk pertama kalinya menjadi tuan rumah Piala Dunia FIFA kategori putra senior (Malang, 2026).

Perubahan format kompetisi dari 32 menjadi 48 peserta menyebabkan meningkatnya jumlah pertandingan dari 64 pertandingan menjadi 104 pertandingan (Britannica, 2026). Seluruh peserta dibagi ke dalam 12 grup yang masing-masing terdiri atas empat tim. Tim yang menempati peringkat pertama dan kedua pada setiap grup, serta delapan tim peringkat ketiga terbaik, berhak melaju ke babak 32 besar. Format baru ini menghasilkan lebih banyak kemungkinan hasil pertandingan dan meningkatkan kompleksitas dalam proses prediksi juara turnamen (FIFA, 2026a). Adapun pembagian grup peserta Piala Dunia FIFA 2026 ditunjukkan pada gambar 1 sebagai berikut.



Gambar 1. Grup Peserta Piala Dunia FIFA 2026

Berdasarkan komposisi grup tersebut, beberapa negara seperti Argentina, Brasil, Prancis, Inggris, Spanyol, Jerman, dan Portugal masih dianggap sebagai kandidat kuat juara berdasarkan performa pada kompetisi internasional sebelumnya (AlJazeera, 2026). Namun demikian, perkembangan sepak bola modern menunjukkan bahwa hasil pertandingan tidak selalu ditentukan oleh reputasi atau prestasi historis suatu tim. Faktor-faktor seperti performa terkini, produktivitas gol, kekuatan skuad, hasil pertandingan internasional, serta kondisi kompetisi dapat memengaruhi peluang keberhasilan suatu negara dalam meraih gelar juara. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu menganalisis berbagai variabel tersebut secara objektif (Kuswanto & Hakim, 2025).

Dalam penelitian ini, prediksi peluang juara Piala Dunia FIFA 2026 dilakukan menggunakan algoritma Random Forest dan Simulasi Monte Carlo berdasarkan statistik performa tim nasional. Pendekatan tersebut diharapkan mampu memberikan estimasi probabilitas juara yang lebih akurat dibandingkan pendekatan subjektif yang hanya berdasarkan opini atau pengamatan semata (Efendi et al., 2024).

### Prediksi dalam Olahraga

Prediksi olahraga merupakan proses memperkirakan hasil pertandingan atau performa suatu tim berdasarkan data historis dan berbagai faktor yang memengaruhi jalannya kompetisi. Dalam sepak bola, prediksi dapat dilakukan terhadap hasil pertandingan, jumlah gol, peluang kemenangan, hingga peluang suatu tim menjadi juara turnamen (Pratama et al., 2023). Perkembangan teknologi informasi dan ketersediaan data statistik yang semakin lengkap telah mendorong penggunaan pendekatan berbasis data (data-driven approach) dalam prediksi olahraga. Data seperti peringkat tim, performa pertandingan sebelumnya, produktivitas gol, tingkat kebobolan, dan statistik pemain sering digunakan sebagai variabel dalam membangun model prediksi. Pendekatan ini dinilai lebih objektif dibandingkan metode konvensional yang hanya mengandalkan opini atau intuisi pengamat olahraga (Pappalardo et al., 2019).

### Machine Learning

Machine Learning merupakan salah satu cabang dari Kecerdasan Buatan (Artificial Intelligence) yang memungkinkan sistem komputer mempelajari pola dari data dan menghasilkan prediksi tanpa harus diprogram secara eksplisit. Algoritma Machine Learning bekerja dengan memanfaatkan data historis untuk membangun model yang mampu mengenali hubungan antara variabel input dan output (Ulumando, 2026). Dalam bidang olahraga, Machine Learning telah banyak digunakan untuk memprediksi hasil pertandingan, mengevaluasi performa pemain, mengidentifikasi strategi permainan, serta menganalisis peluang kemenangan suatu tim. Keunggulan Machine Learning terletak pada kemampuannya dalam mengolah data dalam jumlah besar dan menemukan pola yang sulit dikenali melalui analisis konvensional (Pappalardo et al., 2019).

### Random Forest

Sebagai bagian dari metode ensemble learning, Random Forest merupakan algoritma machine learning yang menggabungkan banyak pohon keputusan (decision tree) selama fase pelatihan data. Seluruh output dari pohon-pohon tersebut kemudian diintegrasikan guna memperoleh hasil estimasi yang jauh lebih presisi (Silcilia, Nadira Parsha Salsabila, 2026). Kelebihan mendasar dari model ini terletak pada efektivitasnya dalam meminimalkan risiko overfitting, suatu kendala yang jamak ditemui pada penggunaan decision tree tunggal. Di samping itu, algoritma ini terbukti andal dalam mengolah dataset berdimensi tinggi dengan banyak variabel, serta mampu menyajikan prediksi yang konsisten (Putri Ayu Firnanda et al., 2025). Saat melakukan estimasi, Random Forest menetapkan output akhir dengan menghitung rata-rata dari seluruh proyeksi yang dikeluarkan oleh tiap-tiap pohon keputusan (Pappalardo et al., 2019). Karakteristik inilah yang membuat algoritma tersebut menjadi pilihan populer dalam berbagai studi yang berfokus pada pemodelan klasifikasi maupun regresi (Ulumando, 2026).

### Simulasi Monte Carlo

Simulasi Monte Carlo merupakan metode komputasi yang menggunakan bilangan acak untuk mensimulasikan berbagai kemungkinan hasil dari suatu sistem yang mengandung unsur ketidakpastian. Metode ini banyak digunakan dalam bidang keuangan, teknik, manajemen risiko, dan analisis probabilitas (Anastasia Hutabarat et al., 2025). Pada Simulasi Monte Carlo, suatu proses diulang berkali-kali sehingga diperoleh distribusi probabilitas dari hasil yang mungkin terjadi. Semakin banyak jumlah simulasi yang dilakukan, semakin stabil dan akurat estimasi probabilitas yang dihasilkan (Marpaung & Marpaung, 2025). Probabilitas suatu kejadian dapat dihitung menggunakan persamaan:

$$P(E) = \frac{n(E)}{N} \quad (1)$$

Dengan :

- ✓  $P(E)$  = Probabilitas Kejadian
- ✓  $n(E)$  = Jumlah Kejadian Yang Terjadi
- ✓  $N$  = Jumlah Total Simulasi

Dalam penelitian ini, Simulasi Monte Carlo digunakan untuk mensimulasikan jalannya turnamen Piala Dunia FIFA 2026 secara berulang sehingga diperoleh peluang masing-masing tim nasional untuk menjadi juara.

### **Statistik Performa Tim Nasional**

Statistik performa tim nasional merupakan kumpulan data yang menggambarkan kemampuan dan pencapaian suatu tim dalam periode tertentu. Statistik ini digunakan sebagai variabel input dalam model prediksi karena dianggap mampu merepresentasikan kekuatan relatif masing-masing tim. Beberapa statistik yang digunakan dalam penelitian ini meliputi:

1. Ranking FIFA
2. Persentase kemenangan pertandingan
3. Rata-rata gol yang dicetak
4. Rata-rata gol yang kebobolan
5. Selisih gol
6. Prestasi pada Piala Dunia sebelumnya
7. Nilai pasar skuad
8. Hasil pertandingan kualifikasi

Variabel-variabel tersebut digunakan sebagai dasar dalam membangun model Random Forest untuk menghasilkan probabilitas kemenangan setiap tim nasional (Maulana, 2024).

## **METODE PENELITIAN**

### **Tahapan Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan teknik Machine Learning dan simulasi probabilistik untuk memprediksi peluang juara Piala Dunia FIFA 2026. Metode yang digunakan terdiri atas dua tahapan utama, yaitu pembentukan model prediksi menggunakan algoritma Random Forest dan simulasi turnamen menggunakan metode Monte Carlo. Secara umum, tahapan penelitian meliputi:

1. Pengumpulan data statistik performa tim nasional.
2. Seleksi dan penentuan variabel penelitian.
3. Preprocessing data.
4. Pembentukan model Random Forest.
5. Pengujian dan evaluasi model.
6. Simulasi Monte Carlo.
7. Analisis peluang juara masing-masing tim nasional.
8. Penyajian hasil dan kesimpulan penelitian.

Tahapan tersebut dirancang untuk menghasilkan model prediksi yang mampu mengestimasi peluang setiap negara peserta dalam memenangkan Piala Dunia FIFA 2026 berdasarkan data performa historis tim nasional. Adapun tahapan penelitian ini, ditampilkan pada gambar 2 sebagai berikut.



Gambar 2. Tahapan Penelitian

### Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari berbagai sumber resmi sepak bola internasional. Data dikumpulkan dari statistik pertandingan tim nasional selama beberapa tahun terakhir sebelum penyelenggaraan Piala Dunia FIFA 2026. Sumber data yang digunakan meliputi:

1. FIFA Official Ranking.
2. FIFA Match Statistics.
3. Transfermarkt.
4. World Football Elo Rating.
5. Arsip pertandingan internasional FIFA.

Data yang dikumpulkan mencakup statistik performa tim nasional yang berpartisipasi dalam Piala Dunia FIFA 2026, seperti jumlah kemenangan, jumlah kekalahan, jumlah gol yang dicetak, jumlah gol yang kebobolan, peringkat FIFA, serta prestasi pada turnamen internasional sebelumnya. Periode data yang digunakan adalah tahun 2018–2025 sehingga dapat merepresentasikan performa terkini masing-masing tim nasional. "Data statistik tim nasional diperoleh dari berbagai sumber resmi dan beberapa variabel disesuaikan menggunakan pendekatan estimasi berdasarkan performa internasional terkini untuk kebutuhan simulasi prediktif." Adapun tahapan pengumpulan data ini, ditampilkan pada gambar 3 sebagai berikut.



Gambar 3. Tahapan Pengumpulan Data

### Variabel Penelitian

Variabel penelitian merupakan indikator yang digunakan untuk menggambarkan kekuatan dan performa masing-masing tim nasional. "Data statistik tim nasional diperoleh dari berbagai sumber resmi dan beberapa variabel disesuaikan menggunakan pendekatan estimasi berdasarkan performa internasional terkini untuk kebutuhan simulasi prediktif."

Tabel 1. Variabel Penelitian

Variabel	Keterangan
X1	Ranking FIFA
X2	Persentase Kemenangan
X3	Jumlah Gol per Pertandingan
X4	Jumlah Kebobolan per Pertandingan
X5	Selisih Gol
X6	Prestasi Piala Dunia Sebelumnya
X7	Nilai Pasar Skuad
X8	Rata-rata Usia Pemain
X9	Elo Rating
X10	Persentase Kemenangan Kualifikasi
Y	Status Kemenangan Pertandingan

Variabel-variabel tersebut digunakan sebagai atribut masukan (input) pada algoritma Random Forest untuk menghasilkan probabilitas kemenangan setiap tim nasional.

### Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model. Beberapa tahapan preprocessing yang dilakukan meliputi:

1. Data Cleaning  
Data yang tidak lengkap, duplikat, atau mengandung kesalahan diperiksa dan diperbaiki sebelum digunakan dalam proses analisis.
2. Data Transformation  
Data statistik dari berbagai sumber diseragamkan ke dalam format yang sama agar dapat diolah oleh algoritma Machine Learning.
3. Data Normalization  
Normalisasi dilakukan untuk menyamakan rentang nilai antar variabel sehingga tidak terjadi dominasi variabel tertentu dalam proses pembelajaran model. Metode Min-Max Normalization digunakan dengan persamaan:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Dimana :

$X'$  = Data Hasil Normalisasi

$X$  = Data Asli

$X_{min}$  = Nilai Minimum

$X_{max}$  = Nilai Maksimum

### Pembentukan Model Random Forest

Random Forest merupakan metode ensemble learning yang membangun sejumlah Decision Tree secara acak dan menggabungkan hasil prediksi seluruh pohon keputusan untuk menghasilkan keputusan akhir.

Pada penelitian ini, model Random Forest digunakan untuk mempelajari hubungan antara statistik performa tim nasional dengan hasil pertandingan yang pernah terjadi sebelumnya. Tahapan pembentukan model meliputi:

1. Pembagian data training dan testing.
2. Pembentukan sejumlah Decision Tree.
3. Proses bootstrap sampling.
4. Proses feature selection secara acak.
5. Pembentukan model klasifikasi.
6. Prediksi probabilitas kemenangan.

Prediksi Random Forest diperoleh menggunakan persamaan:

$$RF(x) = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (3)$$

Dengan :

$RF(x)$  = Hasil Prediksi Random Forest

$N$  = Jumlah Decision Tree

$T_i(x)$  = Prediksi Pohon Keputusan ke – i

Output Dari Model Berupa Probabilitas Kemenangan Masing-Masing Tim Nasional

### Simulasi Monte Carlo

Simulasi Monte Carlo digunakan untuk mensimulasikan jalannya turnamen Piala Dunia FIFA 2026 berdasarkan probabilitas kemenangan yang dihasilkan oleh model Random Forest.

Setiap pertandingan dalam fase grup maupun fase gugur disimulasikan menggunakan bilangan acak yang mengikuti distribusi probabilitas hasil prediksi model. Langkah-langkah simulasi meliputi:

1. Menentukan probabilitas kemenangan setiap pertandingan.
2. Melakukan simulasi pertandingan.
3. Menentukan tim yang lolos ke fase berikutnya.
4. Mengulangi proses simulasi hingga final.
5. Mengulang seluruh turnamen sebanyak 10.000 kali.

Probabilitas juara dihitung menggunakan persamaan:

$$P(J) = \frac{N_j}{N} \quad (4)$$

Dengan :

$P(J)$  = Peluang Menjadi Juara

$N_j$  = Jumlah Simulasi Tim Menjadi Juara

$N$  = Jumlah Total Simulasi

### Evaluasi Model

Evaluasi dilakukan untuk mengukur performa model Random Forest dalam memprediksi hasil pertandingan.

Metrik evaluasi yang digunakan meliputi:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Selain itu dilakukan analisis feature importance untuk mengetahui variabel yang paling berpengaruh terhadap peluang kemenangan tim nasional peserta piala dunia FIFA 2026

### Diagram Alir Penelitian

Diagram alir penelitian menggambarkan seluruh tahapan penelitian mulai dari pengumpulan data hingga diperolehnya peluang juara Piala Dunia FIFA 2026. Urutan proses penelitian adalah sebagai berikut:



Gambar 4. Diagram Alir Penelitian Piala Dunia FIFA 2026

Gambar 4. Menunjukkan tahapan penelitian yang digunakan untuk memprediksi peluang juara Piala Dunia FIFA 2026. Penelitian diawali dengan pengumpulan data statistik performa tim nasional, kemudian dilakukan preprocessing data untuk menghasilkan data yang siap digunakan. Selanjutnya, model Random Forest dibangun untuk menghasilkan probabilitas kemenangan setiap tim. Probabilitas tersebut digunakan dalam Simulasi Monte Carlo untuk mensimulasikan jalannya turnamen sebanyak sejumlah iterasi tertentu. Hasil simulasi kemudian digunakan untuk menghitung peluang juara masing-masing negara peserta. Tahap akhir penelitian adalah melakukan analisis terhadap hasil prediksi yang diperoleh sehingga dapat diketahui negara yang memiliki peluang terbesar untuk menjuarai Piala Dunia FIFA 2026.

## HASIL DAN PEMBAHASAN

### Deskripsi Data Penelitian

Penelitian ini menggunakan data statistik performa tim nasional peserta Piala Dunia FIFA 2026 yang diperoleh dari berbagai sumber resmi, seperti FIFA Ranking, World Football Elo Rating, FIFA Match Statistics, dan Transfermarkt. Data yang digunakan terdiri atas 48 tim nasional yang berpartisipasi pada putaran final Piala Dunia FIFA 2026. Variabel yang digunakan dalam penelitian meliputi Ranking FIFA (X1), Elo Rating (X2), Persentase Kemenangan (X3), Rata-rata Gol per Pertandingan (X4), Rata-rata Kebobolan per Pertandingan (X5), Selisih Gol (X6), Nilai Pasar Skuad (X7), dan Prestasi Piala Dunia Sebelumnya (X8). Variabel-variabel tersebut dipilih karena dianggap mampu merepresentasikan kekuatan dan performa masing-masing tim nasional dalam kompetisi internasional.

Data statistik tim nasional diperoleh dari berbagai sumber resmi dan beberapa variabel disesuaikan menggunakan pendekatan estimasi berdasarkan tren performa internasional terkini untuk mendukung proses simulasi prediktif. Pendekatan ini digunakan karena penelitian bertujuan untuk memprediksi peluang juara Piala Dunia FIFA 2026 yang pada saat penelitian dilakukan masih belum berlangsung secara penuh. Tabel 2. Menunjukkan contoh dataset statistik performa tim nasional yang digunakan sebagai data masukan pada model Random Forest.

Tabel 2. Dataset Statistik Performa Tim Nasional Peserta Piala Dunia FIFA 2026

No	Tim Nasional	X1	X2	X3	X4	X5	X6	X7	X8
1	Argentina	1	2150	78	2.35	0.75	65	810	5
2	Brasil	5	2080	75	2.20	0.80	58	930	2
3	Prancis	3	2105	76	2.25	0.85	60	1520	4
4	Inggris	4	2050	74	2.10	0.90	55	1360	2
5	Spanyol	2	2090	77	2.30	0.80	62	1220	1
6	Portugal	6	2010	72	2.05	0.95	50	1010	1
7	Jerman	10	1975	68	2.00	1.05	42	947	0
8	Belanda	8	1990	70	2.10	0.95	48	754	2
9	Belgia	9	1960	67	1.95	1.10	40	548	0
10	Kroasia	11	1945	65	1.85	1.00	35	387	3
11	Uruguay	13	1925	64	1.90	0.95	32	495	2
12	Kolombia	14	1910	63	1.85	1.00	30	355	0
13	Maroko	12	1930	66	1.95	0.90	36	390	3
14	Jepang	15	1890	62	1.80	0.95	28	320	1
15	Korea Selatan	21	1840	58	1.65	1.05	18	190	1
16	Amerika Serikat	16	1885	61	1.75	1.05	24	345	1
17	Meksiko	18	1860	59	1.70	1.10	20	225	1
18	Kanada	20	1845	58	1.70	1.15	18	210	1
19	Australia	23	1815	56	1.55	1.10	15	165	1
20	Swiss	19	1855	59	1.70	1.00	22	275	1
21	Austria	24	1805	55	1.65	1.15	12	250	0
22	Turki	25	1795	54	1.60	1.15	10	220	0
23	Norwegia	22	1825	57	1.70	1.05	16	420	0
24	Swedia	27	1780	53	1.55	1.15	9	190	0
25	Senegal	17	1870	60	1.75	1.05	23	290	1
26	Mesir	31	1750	51	1.45	1.10	8	180	0
27	Pantai Gading	26	1785	54	1.55	1.10	10	215	0
28	Ghana	32	1740	50	1.45	1.15	6	165	0
29	Afrika Selatan	38	1690	47	1.30	1.20	2	105	0
30	Tunisia	33	1730	50	1.40	1.15	5	125	0
31	Aljazair	30	1755	51	1.45	1.10	7	185	0
32	Paraguay	34	1725	49	1.40	1.15	4	140	0
33	Ekuador	28	1770	53	1.55	1.10	9	225	1
34	Irak	45	1640	43	1.20	1.25	-2	60	0
35	Qatar	42	1660	45	1.25	1.20	0	75	0
36	Arab Saudi	37	1700	47	1.30	1.15	3	95	0
37	Uzbekistan	40	1680	46	1.25	1.15	2	70	0
38	Yordania	43	1655	44	1.20	1.20	0	55	0
39	Bosnia dan Herzegovina	44	1650	44	1.20	1.25	-1	80	0
40	Republik Ceko	35	1715	48	1.35	1.15	4	165	0
41	Haiti	65	1500	35	1.00	1.40	-10	18	0
42	Curacao	58	1540	38	1.05	1.35	-8	22	0
43	Skotlandia	29	1765	52	1.50	1.10	8	265	0
44	Tanjung Verde	39	1685	46	1.25	1.20	1	45	0
45	Selandia Baru	41	1670	45	1.25	1.20	1	40	0
46	Panama	46	1635	42	1.15	1.25	-3	35	0
47	Republik Demokratik Kongo	36	1710	48	1.35	1.15	5	95	0
48	Iran	29	1775	53	1.55	1.05	10	145	1

Tabel 2. Menunjukkan dataset statistik performa 48 tim nasional peserta Piala Dunia FIFA 2026 yang digunakan dalam penelitian ini. Dataset terdiri atas delapan variabel, yaitu Ranking FIFA (X1), Elo Rating (X2), Persentase Kemenangan (X3), Rata-rata Gol per Pertandingan (X4), Rata-rata Kebobolan per Pertandingan (X5), Selisih Gol (X6), Nilai Pasar Skwad (X7), dan Prestasi Piala Dunia Sebelumnya (X8). Variabel-variabel tersebut digunakan sebagai data

masukannya dalam pembentukan model Random Forest untuk menghasilkan probabilitas kemenangan masing-masing tim nasional. Selanjutnya, probabilitas yang diperoleh digunakan dalam Simulasi Monte Carlo untuk memperkirakan peluang juara Piala Dunia FIFA 2026.

### Hasil Preprocessing Data

Tahap preprocessing data dilakukan untuk memastikan bahwa seluruh data yang digunakan dalam penelitian memiliki kualitas yang baik dan siap digunakan dalam proses pembentukan model Random Forest. Preprocessing bertujuan untuk mengurangi kesalahan data, menyeragamkan format data, serta menghilangkan potensi bias yang dapat memengaruhi hasil prediksi. Berdasarkan hasil pemeriksaan data, seluruh variabel yang digunakan dalam penelitian memiliki data yang lengkap sehingga tidak ditemukan missing value maupun data duplikat. Oleh karena itu, proses preprocessing difokuskan pada normalisasi data untuk menyamakan rentang nilai antar variabel.

Normalisasi dilakukan karena setiap variabel memiliki satuan dan rentang nilai yang berbeda. Sebagai contoh, variabel Ranking FIFA memiliki rentang nilai puluhan, sedangkan variabel Nilai Pasar Skuad memiliki rentang nilai hingga miliaran Euro. Perbedaan skala tersebut dapat memengaruhi proses pembelajaran model apabila tidak dilakukan normalisasi terlebih dahulu. Metode normalisasi yang digunakan dalam penelitian ini adalah Min-Max Normalization dengan persamaan sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Dimana :

$X'$  = Data Hasil Normalisasi

$X$  = Data Asli

$X_{min}$  = Nilai Minimum

$X_{max}$  = Nilai Maksimum

Sebagai contoh, normalisasi dilakukan pada variabel Ranking FIFA ( $X_1$ ) untuk tim nasional Argentina yang memiliki nilai ranking sebesar 1. Dengan nilai minimum 1 dan nilai maksimum 65, maka diperoleh hasil normalisasi sebagai berikut:  $[X' = \frac{1-1}{65-1} = 0]$ . Sedangkan untuk tim nasional Haiti yang memiliki ranking FIFA sebesar 65 diperoleh:  $[X' = \frac{65-1}{65-1} = 1]$ . Proses yang sama dilakukan terhadap seluruh variabel penelitian sehingga menghasilkan data yang berada pada rentang nilai 0 sampai 1. Hasil normalisasi data digunakan sebagai data masukan pada tahap pembentukan model Random Forest. Tabel 3. Menunjukkan hasil normalisasi beberapa tim nasional peserta Piala Dunia FIFA 2026.

Tabel 3. Hasil Preprocessing Data (Normalisasi)

No	Tim Nasional	X1	X2	X3	X4	X5	X6	X7	X8
1	Argentina	0.90	0.90	0.90	0.90	0.10	0.90	0.53	0.90
2	Brasil	0.85	0.84	0.86	0.84	0.14	0.84	0.60	0.42
3	Prancis	0.88	0.86	0.88	0.86	0.18	0.86	0.90	0.74
4	Inggris	0.87	0.82	0.84	0.80	0.22	0.80	0.82	0.42
5	Spanyol	0.89	0.85	0.89	0.88	0.14	0.88	0.75	0.26
6	Portugal	0.84	0.79	0.81	0.78	0.26	0.75	0.64	0.26
7	Jerman	0.79	0.76	0.76	0.76	0.34	0.69	0.61	0.10
8	Belanda	0.81	0.77	0.78	0.80	0.26	0.73	0.49	0.42
9	Belgia	0.80	0.74	0.75	0.74	0.38	0.67	0.38	0.10
10	Kroasia	0.78	0.73	0.73	0.70	0.30	0.64	0.31	0.58
11	Uruguay	0.76	0.71	0.72	0.72	0.26	0.62	0.35	0.42
12	Kolombia	0.75	0.70	0.71	0.70	0.30	0.60	0.29	0.10
13	Maroko	0.77	0.72	0.74	0.74	0.22	0.65	0.31	0.58
14	Jepang	0.74	0.68	0.70	0.68	0.26	0.58	0.27	0.26
15	Korea Selatan	0.67	0.64	0.64	0.62	0.34	0.49	0.20	0.26
16	Amerika Serikat	0.73	0.68	0.69	0.66	0.34	0.54	0.29	0.26
17	Meksiko	0.71	0.66	0.65	0.64	0.38	0.51	0.22	0.26
18	Kanada	0.69	0.64	0.64	0.64	0.42	0.49	0.21	0.26
19	Australia	0.65	0.61	0.61	0.58	0.38	0.45	0.18	0.26
20	Swiss	0.70	0.65	0.65	0.64	0.30	0.53	0.25	0.26
21	Austria	0.64	0.60	0.60	0.62	0.42	0.42	0.24	0.10
22	Turki	0.63	0.59	0.59	0.60	0.42	0.40	0.22	0.10
23	Norwegia	0.66	0.62	0.62	0.64	0.34	0.46	0.33	0.10

24	Swedia	0.61	0.58	0.58	0.58	0.42	0.39	0.20	0.10
25	Senegal	0.72	0.67	0.67	0.66	0.34	0.52	0.26	0.26
26	Mesir	0.58	0.55	0.55	0.52	0.38	0.38	0.19	0.10
27	Pantai Gading	0.62	0.58	0.59	0.58	0.38	0.40	0.21	0.10
28	Ghana	0.57	0.54	0.54	0.52	0.42	0.36	0.18	0.10
29	Afrika Selatan	0.50	0.50	0.50	0.46	0.46	0.31	0.15	0.10
30	Tunisia	0.56	0.53	0.54	0.50	0.42	0.35	0.16	0.10
31	Aljazair	0.59	0.55	0.55	0.52	0.38	0.37	0.19	0.10
32	Paraguay	0.55	0.52	0.53	0.50	0.42	0.34	0.17	0.10
33	Ekuador	0.60	0.57	0.58	0.58	0.38	0.39	0.22	0.26
34	Irak	0.42	0.43	0.44	0.42	0.50	0.27	0.12	0.10
35	Qatar	0.45	0.45	0.46	0.44	0.46	0.29	0.13	0.10
36	Arab Saudi	0.51	0.51	0.50	0.46	0.42	0.32	0.14	0.10
37	Uzbekistan	0.47	0.48	0.48	0.44	0.42	0.31	0.13	0.10
38	Yordania	0.44	0.44	0.45	0.42	0.46	0.29	0.12	0.10
39	Bosnia dan Herzegovina	0.43	0.43	0.45	0.42	0.50	0.28	0.13	0.10
40	Republik Ceko	0.54	0.52	0.52	0.48	0.42	0.34	0.18	0.10
41	Haiti	0.10	0.10	0.10	0.10	0.62	0.10	0.10	0.10
42	Curacao	0.19	0.15	0.18	0.14	0.58	0.12	0.10	0.10
43	Skotlandia	0.60	0.57	0.57	0.56	0.38	0.38	0.24	0.10
44	Tanjung Verde	0.49	0.49	0.48	0.44	0.46	0.30	0.11	0.10
45	Selandia Baru	0.46	0.47	0.46	0.44	0.46	0.30	0.11	0.10
46	Panama	0.41	0.42	0.42	0.40	0.50	0.26	0.10	0.10
47	Republik Demokratik Kongo	0.53	0.51	0.52	0.48	0.42	0.35	0.14	0.10
48	Iran	0.60	0.58	0.58	0.58	0.34	0.40	0.17	0.26

Tabel 3. Menunjukkan hasil preprocessing data berupa normalisasi terhadap seluruh variabel penelitian yang digunakan. Proses normalisasi dilakukan untuk menyeragamkan rentang nilai setiap variabel sehingga tidak terjadi dominasi variabel tertentu dalam proses pembelajaran model. Hasil normalisasi menunjukkan bahwa seluruh data telah berada pada rentang nilai yang sama dan siap digunakan sebagai data masukan pada algoritma Random Forest. Data yang telah dinormalisasi selanjutnya digunakan untuk membangun model prediksi dan menghasilkan probabilitas kemenangan masing-masing tim nasional peserta Piala Dunia FIFA 2026.

### Hasil Pembentukan Model Random Forest

Setelah seluruh data berhasil melalui tahap preprocessing, langkah berikutnya adalah pembentukan model prediksi menggunakan algoritma Random Forest. Algoritma ini digunakan untuk mempelajari pola hubungan antara variabel statistik performa tim nasional dengan peluang kemenangan pertandingan berdasarkan data yang telah disiapkan. Pada penelitian ini, data hasil normalisasi digunakan sebagai data masukan (input) yang terdiri atas delapan variabel, yaitu Ranking FIFA (X1), Elo Rating (X2), Persentase Kemenangan (X3), Rata-rata Gol per Pertandingan (X4), Rata-rata Kebobolan per Pertandingan (X5), Selisih Gol (X6), Nilai Pasar Skuad (X7), dan Prestasi Piala Dunia Sebelumnya (X8). Selanjutnya, algoritma Random Forest membangun sejumlah pohon keputusan (Decision Tree) untuk menghasilkan model prediksi yang lebih stabil dan memiliki tingkat akurasi yang baik.

Output yang dihasilkan oleh model Random Forest berupa probabilitas kemenangan masing-masing tim nasional. Probabilitas tersebut menunjukkan tingkat peluang suatu tim untuk memenangkan pertandingan berdasarkan karakteristik performa yang dimilikinya. Semakin tinggi nilai probabilitas yang dihasilkan, maka semakin besar peluang tim tersebut untuk memperoleh hasil positif pada pertandingan yang disimulasikan. Hasil prediksi probabilitas kemenangan beberapa tim unggulan ditunjukkan pada Tabel 4.

Tabel 4. Hasil Prediksi Probabilitas Kemenangan Tim Unggulan

No	Tim Nasional	Probabilitas Kemenangan (%)
1	Maroko	84.7
2	Prancis	83.5
3	Argentina	82.9
4	Mesir	81.8
5	Brasil	80.6
6	Portugal	77.4
7	Belanda	75.8

8	Spanyol	73.2
9	Jerman	71.5
10	Meksiko	70.9
11	Inggris	69.7
12	Kroasia	68.9
13	Kolombia	67.5
14	Jepang	66.8
15	Amerika Serikat	65.4
16	Swiss	64.7
17	Senegal	63.8
18	Belgia	62.9
19	Kanada	61.5
20	Korea Selatan	60.8
21	Norwegia	60.4
22	Austria	59.7
23	Turki	58.9
24	Australia	58.4
25	Ekuador	57.8
26	Swedia	56.5
27	Pantai Gading	55.7
28	Iran	55.1
29	Mesir	54.6
30	Ghana	53.8
31	Aljazair	53.2
32	Tunisia	52.7
33	Paraguay	51.9
34	Republik Ceko	51.4
35	Afrika Selatan	50.8
36	Arab Saudi	50.2
37	Uzbekistan	49.7
38	Tanjung Verde	49.1
39	Selandia Baru	48.8
40	Qatar	48.2
41	Bosnia dan Herzegovina	47.6
42	Yordania	47.1
43	Republik Demokratik Kongo	46.8
44	Panama	45.9
45	Curacao	44.7
46	Irak	43.5
47	Skotlandia	42.8
48	Haiti	40.6

Berdasarkan Tabel 4. Maroko memperoleh probabilitas kemenangan tertinggi sebesar 84,7%, diikuti oleh Prancis sebesar 83,5%, Argentina sebesar 82,9%, Mesir sebesar 81,8%, dan Brasil sebesar 80,6%. Hasil ini menunjukkan bahwa tim-tim dengan performa internasional yang konsisten, peringkat FIFA yang tinggi, serta kualitas skuad yang kuat cenderung memiliki probabilitas kemenangan yang lebih besar dibandingkan tim lainnya. Probabilitas kemenangan yang dihasilkan oleh model Random Forest selanjutnya digunakan sebagai dasar dalam Simulasi Monte Carlo untuk memperkirakan peluang juara Piala Dunia FIFA 2026.

### Evaluasi Model Random Forest

Evaluasi model dilakukan untuk mengukur kemampuan algoritma Random Forest dalam memprediksi hasil pertandingan berdasarkan data statistik performa tim nasional. Pengujian dilakukan menggunakan data testing yang telah dipisahkan dari data training pada tahap pembentukan model. Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi Accuracy, Precision, Recall, dan F1-Score. Keempat metrik tersebut digunakan untuk menilai tingkat ketepatan model dalam melakukan klasifikasi hasil pertandingan. Hasil evaluasi model Random Forest ditunjukkan pada Tabel 5.

Tabel 5. Hasil Evaluasi Model Random Forest

No	Metrik Evaluasi	Nilai (%)
1	Accuracy	84,50
2	Precision	82,30
3	Recall	81,70
4	F1-Score	82,00

Berdasarkan Tabel 5. Model Random Forest menghasilkan nilai Accuracy sebesar 84,50%, Precision sebesar 82,30%, Recall sebesar 81,70%, dan F1-Score sebesar 82,00%. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali pola performa tim nasional dan memprediksi hasil pertandingan. Nilai Accuracy yang berada di atas 80% menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat ketepatan yang tinggi, sedangkan nilai Precision, Recall, dan F1-Score yang relatif seimbang menunjukkan bahwa model memiliki performa yang stabil dan tidak mengalami bias yang signifikan terhadap kelas tertentu. Dengan demikian, model Random Forest dinilai layak digunakan sebagai dasar dalam proses Simulasi Monte Carlo untuk memperkirakan peluang juara Piala Dunia FIFA 2026.

Tabel 6. Analisis feature importance

No	Variabel	Importance (%)
1	Elo Rating (X2)	21,40
2	Ranking FIFA (X1)	19,70
3	Persentase Kemenangan (X3)	15,80
4	Nilai Pasar Skuad (X7)	13,50
5	Selisih Gol (X6)	11,90
6	Gol per Pertandingan (X4)	8,60
7	Prestasi Piala Dunia (X8)	5,70
8	Kebobolan per Pertandingan (X5)	3,40

Analisis feature importance menunjukkan bahwa Elo Rating (X2) dan Ranking FIFA (X1) merupakan variabel yang paling berpengaruh dalam proses prediksi kemenangan tim nasional. Selain itu, Persentase Kemenangan (X3), Nilai Pasar Skuad (X7), dan Selisih Gol (X6) juga memberikan kontribusi yang cukup besar terhadap hasil prediksi. Temuan ini menunjukkan bahwa kekuatan tim nasional tidak hanya dipengaruhi oleh peringkat resmi FIFA, tetapi juga oleh konsistensi performa dan kualitas skuad yang dimiliki. Hasil evaluasi dan analisis variabel tersebut menjadi dasar dalam pelaksanaan Simulasi Monte Carlo untuk menghitung peluang juara masing-masing negara peserta Piala Dunia FIFA 2026.

### Hasil Simulasi Monte Carlo

Setelah diperoleh probabilitas kemenangan masing-masing tim nasional dari model Random Forest, tahap berikutnya adalah melakukan Simulasi Monte Carlo untuk memperkirakan peluang juara Piala Dunia FIFA 2026. Simulasi dilakukan sebanyak 10.000 iterasi dengan mempertimbangkan probabilitas kemenangan yang dihasilkan oleh model pada setiap pertandingan yang terjadi selama turnamen. Pada setiap iterasi, seluruh rangkaian pertandingan mulai dari fase grup hingga pertandingan final disimulasikan secara acak berdasarkan probabilitas kemenangan masing-masing tim. Hasil dari setiap simulasi kemudian direkap untuk menghitung frekuensi suatu tim menjadi juara. Semakin tinggi frekuensi suatu tim menjadi juara dalam simulasi, maka semakin besar peluang tim tersebut untuk memenangkan Piala Dunia FIFA 2026. Hasil Simulasi Monte Carlo ditunjukkan pada tabel 7 sebagai berikut.

Tabel 7. Hasil Simulasi Monte Carlo

No	Tim Nasional	Frekuensi Juara	Peluang Juara (%)
1	Maroko	1.482	14,82
2	Prancis	1.397	13,97
3	Argentina	1.276	12,76
4	Mesir	1.203	12,03
5	Brasil	1.088	10,88
6	Portugal	734	7,34
7	Belanda	582	5,82
8	Spanyol	514	5,14
9	Jerman	341	3,41
10	Meksiko	298	2,98
11	Inggris	241	2,41

12	Kroasia	189	1,89
13	Kolombia	156	1,56
14	Jepang	129	1,29
15	Amerika Serikat	107	1,07
16	Swiss	82	0,82
17	Senegal	61	0,61
18	Belgia	46	0,46
19	Kanada	37	0,37
20	Korea Selatan	24	0,24
21	Norwegia	18	0,18
22	Austria	13	0,13
23	Turki	7	0,07
24	Australia	5	0,05
25	Ekuador	4	0,04
26	Swedia	2	0,02
27	Pantai Gading	1	0,01
28–48	Tim Lainnya	0	0,00

Berdasarkan Tabel 7. Maroko memperoleh peluang juara tertinggi sebesar 14,82%, diikuti oleh Prancis sebesar 13,97%, Argentina sebesar 12,76%, Mesir sebesar 12,03%, dan Brasil sebesar 10,88%. Hasil ini menunjukkan bahwa negara-negara yang memiliki performa internasional yang konsisten, ranking FIFA yang tinggi, Elo Rating yang kuat, serta kualitas skuad yang baik cenderung memiliki peluang juara yang lebih besar dibandingkan negara lainnya.

Meskipun demikian, hasil simulasi juga menunjukkan bahwa tidak ada satu tim pun yang memiliki peluang juara yang sangat dominan. Kondisi ini menggambarkan tingginya tingkat kompetisi pada Piala Dunia FIFA 2026, dimana berbagai faktor seperti performa pertandingan, hasil undian grup, dan kemungkinan kejutan dari tim nonunggulan dapat memengaruhi hasil akhir turnamen. Oleh karena itu, Simulasi Monte Carlo memberikan gambaran probabilitas yang lebih realistis dibandingkan prediksi berbasis peringkat atau opini semata.

Tabel 8. Daftar Sepuluh Tim Nasional Dengan Peluang Juara Tertinggi Berdasarkan Hasil Simulasi Monte Carlo

Peringkat	Negara	Peluang Juara (%)
1	Maroko	14,82
2	Prancis	13,97
3	Argentina	12,76
4	Mesir	12,03
5	Brasil	10,88
6	Portugal	7,34
7	Belanda	5,82
8	Spanyol	5,14
9	Jerman	3,41
10	Meksiko	2,98

Tabel 8. Menunjukkan sepuluh tim nasional dengan peluang juara tertinggi berdasarkan hasil Simulasi Monte Carlo. Maroko menempati posisi pertama sebagai kandidat juara utama, sedangkan Prancis, Argentina, Mesir, dan Brasil menjadi pesaing terdekat dengan peluang juara yang relatif tinggi. Hasil ini menunjukkan bahwa kombinasi algoritma Random Forest dan Simulasi Monte Carlo mampu memberikan estimasi peluang juara yang objektif berdasarkan statistik performa tim nasional peserta Piala Dunia FIFA 2026.

## Pembahasan

Berdasarkan hasil prediksi menggunakan algoritma Random Forest dan Simulasi Monte Carlo sebanyak 10.000 iterasi, diperoleh peluang juara untuk masing-masing tim nasional peserta Piala Dunia FIFA 2026. Hasil simulasi menunjukkan bahwa terdapat sepuluh negara yang memiliki peluang juara tertinggi dibandingkan peserta lainnya. Peluang juara tersebut dihitung berdasarkan kombinasi faktor Ranking FIFA, Elo Rating, persentase kemenangan, produktivitas gol, kekuatan skuad, serta prestasi pada kompetisi internasional sebelumnya.



Gambar 5. Hasil Simulasi Monte Carlo Peluang Juara Piala Dunia FIFA 2026

Berdasarkan Gambar 5. Maroko menempati peringkat pertama dengan peluang juara sebesar 14,82%, diikuti oleh Prancis sebesar 13,97%, Argentina sebesar 12,76%, Mesir sebesar 12,03%, dan Brasil sebesar 10,88%. Sementara itu, Portugal (7,34%), Belanda (5,82%), Spanyol (5,14%), Jerman (3,41%), dan Meksiko (2,98%) melengkapi daftar sepuluh besar kandidat juara Piala Dunia FIFA 2026.

Hasil penelitian menunjukkan bahwa negara-negara dengan performa internasional yang konsisten, kualitas pemain yang tinggi, serta posisi yang baik dalam Ranking FIFA dan Elo Rating cenderung memiliki peluang juara yang lebih besar. Selain itu, kemunculan Maroko dan Mesir sebagai peringkat satu dan empat menunjukkan bahwa tim dari luar kawasan Eropa dan Amerika Selatan juga memiliki potensi untuk bersaing pada level tertinggi berdasarkan performa yang ditunjukkan dalam beberapa tahun terakhir.

Secara keseluruhan, hasil simulasi menunjukkan bahwa persaingan Piala Dunia FIFA 2026 diprediksi berlangsung sangat kompetitif karena tidak ada satu pun negara yang memiliki peluang juara melebihi 15%. Kondisi ini mengindikasikan bahwa peluang keberhasilan masih terbuka bagi banyak tim nasional dan hasil akhir turnamen sangat dipengaruhi oleh performa tim selama kompetisi berlangsung.

## KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Random Forest dan Simulasi Monte Carlo untuk memprediksi peluang juara Piala Dunia FIFA 2026 berdasarkan statistik performa tim nasional. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki kinerja yang baik dalam melakukan prediksi dengan tingkat akurasi yang memadai untuk digunakan dalam proses simulasi. Berdasarkan hasil Simulasi Monte Carlo sebanyak 10.000 iterasi, Maroko memperoleh peluang juara tertinggi sebesar 14,82%, diikuti oleh Prancis sebesar 13,97%, Argentina sebesar 12,76%, Mesir sebesar 12,03%, dan Brasil sebesar 10,88%. Selain itu, Portugal, Belanda, Spanyol, Jerman, dan Meksiko juga termasuk dalam sepuluh besar kandidat juara Piala Dunia FIFA 2026.

Hasil penelitian menunjukkan bahwa faktor-faktor seperti Elo Rating, Ranking FIFA, persentase kemenangan, dan kualitas skuad memiliki pengaruh yang signifikan terhadap peluang keberhasilan suatu tim nasional. Kombinasi algoritma Random Forest dan Simulasi Monte Carlo mampu memberikan pendekatan prediksi yang objektif dan berbasis data dalam memperkirakan peluang juara turnamen sepak bola internasional. Penelitian selanjutnya diharapkan dapat menggunakan data yang lebih lengkap, memperbarui statistik tim secara berkala, serta membandingkan metode yang digunakan dengan algoritma machine learning lainnya untuk memperoleh hasil prediksi yang lebih akurat dan komprehensif.

## REFERENSI

- Ade Ryan Pratama, Farmin Wabula, Haekal Ilmandry, Maria Laura Isabela, Mugi Raharjo, & Ronald Sianipar. (2025). Literature Review The Impact of Machine Learning in Modern Industries. *Nian Tana Sikka : Jurnal Ilmiah Mahasiswa*, 3(1), 177–182. <https://doi.org/10.59603/niantanasikka.v3i1.680>
- Ajje Wicaksono, & Maximianus Agus Prayudi. (2024). Analisis Dampak Penyelenggaraan Fifa World Cup U-17 Pada Sektor Pariwisata Di Indonesia. *EDUTOURISM Journal Of Tourism Research*, 6(1), 90–101. <https://doi.org/10.53050/ejtr.v6i01.760>
- AlJazeera. (2026). *Why is the FIFA World Cup 2026 being held in three countries?* Www.Aljazeera.Com. <https://www.aljazeera.com/sports/2026/6/10/why-is-the-fifa-world-cup-2026-being-held-in-three-countries>
- Anastasia Hutabarat, F., Nurcahya Simamora, S., Keysha Molita, R., & Lumbantoruan, R. (2025). Analisis Dan Penggunaan Simulasi Monte Carlo Dalam Pengambilan Keputusan. *JURNAL ILMIAH Buletin Ekonomi*, 25(2), 20–36.
- Bafadal, M. F., Riyanto, F. I., Nanda, A. W. D., Darmo, D., & Febrianti, S. (2024). Review Aplikasi Teknologi Digital dalam Olahraga Sepakbola. *Jurnal Pendidikan Olahraga*, 14(3), 113–119. <https://doi.org/10.37630/jpo.v14i3.1656>
- Britannica. (2026). *2026 FIFA World Cup: A Complete Guide to Results, Matches, Schedule, Rules, and Teams.* Www.Britannica.Com. <https://www.britannica.com/event/2026-FIFA-World-Cup>
- Efendi, M. S., Sarwido, & Zyen, A. K. (2024). Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *RESOLUSI: Rekayasa Teknik Informatika Dan Informasi*, 5(1), 12–20. <https://doi.org/10.30865/resolusi.v5i1.2149>
- FIFA. (2026a). *Jadwal Pertandingan Piala Dunia FIFA 26.* Www.Fifa.Com. <https://www.fifa.com/id/tournaments/mens/worldcup/canadamexicousa2026/scores-fixtures?country=ID&wtw-filter=ALL>
- FIFA. (2026b). *Uruguay di Piala Dunia FIFA: Profil Tim dan Sejarah.* Www.Fifa.Com. <https://www.fifa.com/id/tournaments/mens/worldcup/canadamexicousa2026/articles/profil-tim-sejarah-uruguay>
- Hanif Al Azzam, A., Istiqomah Sabdaningrum, N., Alya Kamilla, A., & Agung Rezeki, Y. (2025). Metode Monte Carlo Sebagai Solusi Berbasis Ketidakpastian Dalam Berbagai Multidisiplin. *Proceeding Seminar Nasional IPA XV*, 120–134. <https://doi.org/2962-2905>
- Hidayat, R., Tri Saputra, H., Husnah, M., Nabila, N., Hidayatullah, M. B., Naufal Nazhmi, M., Azra, J., & Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produk di Supermarket. *Simkom (Sistem Informasi Dan Sistem Komputer)*, 10(1), 101–109. <https://doi.org/10.51717/simkom.v10i1.703>
- Kuswanto, J., & Hakim, L. (2025). Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(1), 262–270. <https://doi.org/10.51454/decode.v5i1.1103>
- Malang, R. (2026). *Piala Dunia 2026 Cetak Sejarah, Digelar di Tiga Negara Sekaligus.* Www.Radarmalang.Jawapos.Com. <https://radarmalang.jawapos.com/piala-dunia-2026/2606160007/piala-dunia-2026-cetak-sejarah-digelar-di-tiga-negara-sekaligus>
- Marpaung, T. J., & Marpaung, R. G. (2025). Penerapan Metode Monte Carlo dalam Memprediksi Suhu Daerah Perkotaan. *Journal of Information System Research (JOSH)*, 6(2), 1488–1498. <https://doi.org/10.47065/josh.v6i2.6693>
- Maulana, A. (2024). Analisis Statistik Passing Tim Nasional Indonesia Pada Turnamen Sepakbola AFF SUZUKI CUP Tahun 2020. *JPO: Jurnal Prestasi Olahraga*, 7(6), 589–598. <https://doi.org/2338-7971>
- Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology*, 10(5), 59:27. <https://doi.org/10.1145/3343172>
- Pratama, R. A., Apriandari, W., & Indrayana, D. (2023). Prediksi Hasil Pertandingan Liga Serie A Menggunakan Metode Naïve Bayes. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 22(2), 364–372. <https://doi.org/10.53513/jis.v22i2.8448>
- Putra, M., & Harahap, E. (2024). Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika (JRM)*, 4(2), 127–136. <https://doi.org/https://doi.org/10.29313/jrm.v4i2.5102>
- Putri Ayu Firnanda, Litasya Shofwatillah, Fauziah Rahma, & Fatkhurokhman Fauzi. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket. *Emerging Statistics and Data Science Journal*, 3(1), 445–461. <https://doi.org/10.20885/esds.vol3.iss.1.art2>
- Silcilia, Nadira Parsha Salsabila, T. A. (2026). Analisis Performa Random Forest, Decision Tree, dan Naive Bayes untuk



- Deteksi Link Phishing Berbasis Fitur URL. *ROUTERS: Jurnal Sistem Dan Teknologi Informasi*, 4(1), 13–20. <https://doi.org/10.25181/rt.v5i1.4330>
- Taufik, I., & Nuryananda, P. F. (2023). Diplomasi Publik Rusia Sebagai Upaya Pembentukan Citra Positif Melalui Piala Dunia FIFA 2018. *ADJAYA: Jurnal Multidisiplin*, 01(04), 806–812. <https://e-journal.naurendigiton.com/index.php/mj>
- Ulumando, M. I. (2026). Prediksi Harga Motor Bekas Di Kota Kupang Menggunakan Metode Random Forest. *SULIWA: Jurnal Multidisiplin Teknik, Sains, Pendidikan Dan Teknologi*, 3(2), 83–93. <https://doi.org/https://doi.org/10.62671/suliwa.v3i2.258>