PREDIKSI KEMENANGAN ATLET PENCAK SILAT PADA KEJUARAAN DUNIA TAPAK SUCI 2025 MENGGUNAKAN *RANDOM FOREST*

SKRIPSI

Oleh:
AHMAD FADHLI DZILIKROM
NIM. 220605110078

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

PREDIKSI KEMENANGAN ATLET PENCAK SILAT PADA KEJUARAAN DUNIA TAPAK SUCI 2025 MENGGUNAKAN *RANDOM FOREST*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
AHMAD FADHLI DZILIKROM
NIM. 220605110078

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026

HALAMAN PERSETUJUAN

PREDIKSI KEMENANGAN ATLET PENCAK SILAT PADA KEJUARAAAN DUNIA TAPAK SUCI 2025 MENGGUNAKAN *RANDOM FOREST*

SKRIPSI

Oleh:
AHMAD FADHLI DZILIKROM
220605110078

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 11 Mei 2026

Pembimbing I,	Pembimbing II,
Dr. M. Amin Hariyadi, M.T	Okta Qomaruddin Aziz, M.Kom
NIP. 19670118 200501 1 001	NIP. 19911019 201903 1 013

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

PREDIKSI KEMENANGAN ATLET PENCAK SILAT PADA KEJUARAAN DUNIA TAPAK SUCI 2025 MENGGUNAKAN *RANDOM FOREST*

SKRIPSI

Oleh:
AHMAD FADHLI DZILIKROM
NIM. 220605110078

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal: 24 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: Shofin Nahwa Utama, M.T NIP. 19860703 202012 1 003	( )
Anggota Penguji I	: Ajib Hanani, M.T NIP. 19840731 202321 1 013	( )
Anggota Penguji II	: Dr. M. Amin Hariyadi M.T NIP. 19670118 200501 1 001	( )
Anggota Penguji III	: Okta Qomaruddin Aziz, M.Kom NIP. 19911019 201903 1 013	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: AHMAD FADHLI DZILIKROM
NIM	: 220605110078
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Prediksi Kemenangan Atlet Pencak Silat Pada Kejuaraan Dunia Tapak Suci 2025 Menggunakan *Random Forest*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 24 Juni 2026
Yang membuat pernyataan,

AHMAD FADHLI DZILIKROM
NIM. 220605110078

MOTTO

"Berbuatlah kebaikan di mana pun dan kepada siapa pun. Ingat selalu ya, Bang, Allah tidak akan memberikan ujian di luar kemampuan hamba-Nya, dan setelah kesulitan pasti ada kemudahan."

(Umi)

HALAMAN PERSEMBAHAN

Alhamdulillahi rabbil 'alamin, segala puji bagi Allah Swt. atas segala rahmat, nikmat, dan karunia-Nya. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad saw., beserta keluarga, sahabat, dan seluruh umatnya. Atas izin dan pertolongan Allah Swt., penulis diberikan kemudahan dan kekuatan hingga dapat menyelesaikan skripsi ini.

Skripsi ini penulis persembahkan kepada kedua orang tua tercinta sebagai wujud bakti, cinta, kasih sayang, dan rasa terima kasih atas segala doa, pengorbanan, serta dukungan yang telah diberikan. Kehadiran dan doa mereka menjadi alasan utama yang menguatkan penulis untuk terus berjuang hingga mampu menyelesaikan skripsi ini.

Kepada saudara-saudara tercinta dan seluruh keluarga besar, penulis mempersembahkan karya sederhana ini sebagai bentuk rasa syukur dan terima kasih atas perhatian, motivasi, dukungan, serta doa yang selalu menyertai perjalanan penulis dalam menempuh pendidikan hingga terselesaikannya skripsi ini.

Kepada teman-teman dan seluruh pihak yang telah turut membantu penulis dalam menyelesaikan skripsi ini, terima kasih atas doa, dukungan, dan segala bentuk bantuan yang diberikan sehingga skripsi ini dapat terselesaikan dengan baik.

KATA PENGANTAR

Alhamdulillāhirabbil ‘ālamīn, segala puji bagi Allah *Subḥānahu wa Ta‘ālā*. yang telah melimpahkan rahmat, taufik, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul *“Prediksi Kemenangan Atlet Pencak Silat Pada Kejuaraan Dunia Tapak Suci 2025 Menggunakan Random Forest”* dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad *Shollallohu ‘Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nurdiana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Bapak Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak Dr. M. Amin Hariyadi, M.T., selaku dosen pembimbing I dan Bapak Okta Qomaruddin Aziz, M.Kom., selaku dosen pembimbing II, yang telah memberikan bimbingan, arahan, masukan, serta motivasi kepada penulis selama proses penyusunan skripsi ini. Terima kasih atas waktu, kesabaran,

dan perhatian yang telah diberikan sehingga penelitian dan penulisan skripsi ini dapat terselesaikan dengan baik dan tepat waktu.

5. Bapak Shofin Nahwa Utama, M.T., selaku dosen penguji I dan Bapak Ajib Hanani, M.T., selaku dosen penguji II, yang telah memberikan evaluasi, saran, serta berbagai masukan yang berharga sehingga penelitian dan penulisan skripsi ini dapat tersusun dengan lebih baik. Terima kasih atas arahan dan perhatian yang diberikan dalam proses penyusunan skripsi.
6. Seluruh dosen Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, penulis menyampaikan ucapan terima kasih atas ilmu yang telah di berikan selama masa perkuliahan. Semoga segala ilmu dan kebaikan yang telah diberikan menjadi amal jariyah serta memberikan manfaat yang berkelanjutan bagi penulis di masa yang akan datang.
7. Almarhumah Ibu tercinta Sri Utami, yang menjadi alasan terbesar penulis untuk terus berjuang menyelesaikan pendidikan ini. Meskipun beliau telah berpulang ke rahmatullah di tengah proses penyusunan skripsi ini, kasih sayang, doa, dukungan, serta pengorbanan yang telah diberikan semasa hidup akan selalu menjadi kekuatan bagi penulis. Skripsi ini penulis persembahkan sebagai salah satu bentuk terima kasih dan bukti perjuangan yang tidak lepas dari peran beliau. Semoga Allah SWT memberikan tempat terbaik bagi beliau di sisi-Nya.
8. Ayah tercinta Imam Syafi'i, yang senantiasa menjadi teladan dalam kerja keras, keteguhan, dan tanggung jawab. Terima kasih atas doa, dukungan,

kasih sayang, serta pengorbanan yang telah diberikan selama ini. Di tengah berbagai proses dan tantangan yang dihadapi penulis, ayah selalu hadir memberikan semangat dan kekuatan untuk terus melangkah. Setiap nasihat dan perjuangan beliau menjadi motivasi bagi penulis untuk menyelesaikan skripsi ini dengan baik.

9. Mbak Nada Zahro Afifah, Mbak Azhar Shofia Husna dan Adik Muhammad Fahri Taqwa, yang selalu hadir memberikan semangat, doa, dan dukungan kepada penulis dalam setiap proses perjalanan pendidikan.
10. Seluruh keluarga besar penulis yang tidak dapat disebutkan satu per satu, terima kasih atas doa, dukungan, perhatian, dan semangat yang telah diberikan kepada penulis selama menempuh pendidikan.
11. Kak Adi Saputra beserta jajaran kader dan pelatih Tapak Suci, yang telah banyak membantu penulis dalam berbagai keperluan selama proses penelitian, mulai dari pengumpulan data, pemberian informasi, hingga dukungan selama penyusunan skripsi. Atas segala bantuan, kerja sama, dan dukungan yang diberikan, penulis menyampaikan terima kasih yang sebesar-besarnya sehingga penelitian ini dapat terlaksana dengan baik dan lancar.
12. Sahabat seperjuangan yang senantiasa membersamai penulis selama masa perkuliahan, yaitu Ki Ageng Nasrokh Mangkunegoro Putra, Muhammad Annas Darunnaja, Muhammad Andrean Hidayatulloh, Muhammad Alif Atallah, Muhammad Zidan Alvaro, dan Muhammad Haikal Azadin . Terima kasih atas dukungan, kebersamaan, serta berbagai pengalaman yang telah

dibagikan sehingga perjalanan akademik ini terasa lebih ringan dan dapat diselesaikan tepat waktu.

13. Ucapan terima kasih penulis sampaikan kepada keluarga besar Teknik Informatika angkatan 2022 Infinity, yang telah memberikan kebersamaan, dukungan, dan pengalaman berharga selama masa perkuliahan.
14. Ucapan terima kasih penulis sampaikan kepada, *coach* Fathurrahman, Tim Futsal Universitas, dan Tim Futsal Fakultas, yang telah membersamai perjalanan penulis selama masa perkuliahan. Terima kasih atas kebersamaan, dukungan, motivasi, serta pengalaman berharga yang telah diberikan. Berbagai momen dan pelajaran yang diperoleh selama berproses bersama menjadi penyemangat bagi penulis untuk terus berjuang hingga mampu menyelesaikan skripsi ini.
15. Temen-temen KKM Lokahita Adhikara yang sudah memberikan rasa kekeluargaan, dan kebersamaan. Walau hanya bertemu dalam waktu yang singkat, tapi silaturahmi masih terjaga sampai sekarang dan seterusnya.
16. Terima kasih juga kepada Natasya, yang selalu memberikan semangat, dukungan, inspirasi, dan motivasi kepada penulis. Terimakasih atas kebaikannya yang selalu menemani, mendukung, dan siap untuk selalu mendengarkan segala keluh kesah, terimakasih atas support yang diberikan selama ini.
17. Dan teruntuk Fadhli, terimakasih sudah berjuang dalam melewati hari-hari yang penuh rintangan dan tantangan, terimakasih sudah menjadi laki-laki yang hebat, mandiri dan selalu kuat dalam menghadapi semua persoalan

kehidupan di manapun, dan terimakasih sudah bisa menyelesaikan tugas akhir perkuliahan dengan baik.

18. Serta seluruh pihak yang terlibat secara langsung maupun tidak langsung yang tidak bisa disebutkan satu persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat berbagai keterbatasan yang memerlukan penyempurnaan di masa mendatang. Namun demikian, penulis meyakini bahwa proses belajar tidak memiliki batas akhir, sehingga setiap kritik dan saran yang membangun akan menjadi bekal berharga untuk terus berkembang. Bagi penulis, skripsi yang baik bukanlah skripsi yang sempurna, melainkan skripsi yang dapat diselesaikan dengan penuh tanggung jawab dan tepat pada waktunya. Oleh karena itu, penulis berharap skripsi ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang membutuhkan sebagai referensi untuk penelitian selanjutnya.

.

Malang, 24 Juni 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN ..... **iii**
HALAMAN PENGESAHAN ..... **iv**
PERNYATAAN KEASLIAN TULISAN ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR ..... **viii**
DAFTAR ISI ..... **xiii**
DAFTAR GAMBAR ..... **xv**
DAFTAR TABEL ..... **xvi**
ABSTRAK ..... **xvii**
ABSTRACT ..... **xviii**
البحث مستخلص ..... **xix**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 5
1.4 Tujuan Penelitian 5
1.5 Manfaat Penelitian 5
BAB II STUDI PUSTAKA **7**
2.1 Penelitian Terdahulu 7
2.2 Pencak Silat 13
2.3 Prediksi 14
2.4 Algoritma *Random Forest* 15
2.5 Evaluasi Model 18
BAB III METODOLOGI PENELITIAN **21**
3.1 Desain Penelitian 21
3.2 Pengumpulan Data 24
3.3 Desain Sistem 25
3.4 *Prepocessing* Data 30
3.4.1 Pembersihan Data 31
3.4.2 Pembersihan Data 31
3.4.3 Normalisasi Data 32
3.4.4 Pembagian Data (Split Data) 33
3.5 Algoritma *Random Forest* 34
3.6 Evaluasi Model 39
3.7 *K-Fold Cross Validation* 42
3.8 Skenario Pengujian 43
BAB IV HASIL DAN PEMBAHASAN **46**
4.1 Hasil 46
4.1.1 Hasil Pengolahan Data Penelitian 46
4.1.2 Hasil Prepocessing Data 50
4.1.3 Hasil Pembagian Data (Split Data) 53
4.1.4 Hasil Pelatihan Model *Random Forest* 55

4.1.5 Hasil Pelatihan Model *Random Forest* 56
4.1.6 Hasil Pengolahan Data Penelitian 58
4.1.7 Hasil *K-Fold Cross Validation* 63
4.2 Pembahasan 66
4.2.1 Analisis Pengaruh *Hyperparameter* 66
4.2.2 Analisis Pengaruh Split Data 68
4.2.3 Analisis *Feature Importance* dan Implikasinya 70
BAB V KESIMPULAN DAN SARAN 74
5.1 Kesimpulan 74
5.2 Saran 75
DAFTAR PUSTAKA

DAFTAR GAMBAR

Gambar 2. 1 Alur Metode Random Forest 18
Gambar 3. 1 Desain Penelitian 21
Gambar 3. 2 Flowchart Desain Sistem 26
Gambar 3. 3 Alur Kerja Algoritma *Random Forest* 35
Gambar 4. 1 Distribusi Kelas Hasil Pertandingan 49
Gambar 4. 2 Confusion matrix Hasil Evaluasi (Split 70:30) 60
Gambar 4. 3 Confusion matrix Hasil Evaluasi (Split 80:20) 61
Gambar 4. 4 Confusion matrix Hasil Evaluasi (Split 90:10) 61
Gambar 4. 5 Perbandingan Kinerja Tiga Model Skenario Pengujian 62
Gambar 4. 6 Perbandingan Mean Accuracy K-Fold Cross Validation 65

DAFTAR TABEL

Tabel 2. 1 Perbandingan Terdahulu 11
Tabel 3. 1 Deskripsi Atribut Penelitian..... 24
Tabel 3. 2 Contoh Hasil Pembersihan Data 31
Tabel 3. 3 Contoh Hasil Encoding Data..... 32
Tabel 3. 4 Contoh Hasil Normalisasi Data..... 33
Tabel 3. 5 Rancangan Skenario Pembagian Data 34
Tabel 3. 6 Konfigurasi *Hyperparameter* yang Diuji..... 44
Tabel 4. 1 Dataset Atlet Kejuaraan Dunia Tapak Suci 2025 46
Tabel 4. 2 Distribusi Data Penelitian 48
Tabel 4. 3 Hasil Data Cleaning 50
Tabel 4. 4 Hasil Encoding Variabel Target..... 51
Tabel 4. 5 Hasil Encoding Variabel Ordinal..... 51
Tabel 4. 6 Hasil Encoding Variabel Kategorikal 52
Tabel 4. 7 Hasil Encoding Variabel Kategorikal 52
Tabel 4. 8 Hasil Encoding Variabel Kategorikal 53
Tabel 4. 9 Hasil pembagian data pada tiga skenario split..... 54
Tabel 4. 10 Hasil Tuning hyperparameter Random Forest..... 57
Tabel 4. 11 Hasil Evaluasi Model Random Forest 58
Tabel 4. 12 Hasil Pengujian 10 *Fold* Cross Validation..... 63
Tabel 4. 13 Hasil Tuning hyperparameter Random Forest..... 67
Tabel 4. 14 Hasil pembagian data pada tiga skenario split..... 69
Tabel 4. 15 Perbandingan Nilai Accuracy pada Tiga Skenario Split..... 69
Tabel 4. 16 Analisis Feature importances..... 71

ABSTRAK

Dzilikrom, Ahmad Fadhli. 2026. **Prediksi Kemenangan Atlet Pencak Silat Pada Kejuaraan Dunia Tapak Suci 2025 Menggunakan *Random Forest***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, M.T. (II) Okta Qomaruddin Aziz, M.Kom.

.

Kata Kunci: Evaluasi Model, *Feature Importance*, *Machine Learning*, Pencak Silat, Prediksi Kemenangan Atlet, *Random Forest*

Pencak silat merupakan olahraga bela diri yang menuntut kesiapan fisik, pengalaman bertanding, intensitas latihan, serta kondisi psikologis atlet untuk mencapai hasil pertandingan yang optimal. Penelitian ini bertujuan untuk membangun model prediksi kemenangan atlet Pencak Silat pada Kejuaraan Dunia Tapak Suci 2025 menggunakan algoritma *Random Forest*, menilai performa model, serta mengidentifikasi variabel yang paling berpengaruh terhadap hasil prediksi kemenangan atlet. Data yang digunakan merupakan data primer hasil survei terhadap 226 atlet Tapak Suci dengan 15 atribut. Tahapan penelitian meliputi *preprocessing* data, pembagian data menggunakan tiga skenario split data (70:30, 80:20, dan 90:10), pelatihan model, *tuning hyperparameter* menggunakan *Grid Search Cross Validation*, evaluasi model menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*, serta analisis *feature importance*. Hasil penelitian menunjukkan bahwa seluruh skenario pembagian data menghasilkan performa yang tinggi. Pada skenario tersebut, model *Random Forest* menghasilkan nilai *accuracy* sebesar 95,65%, *precision* sebesar 95,99%, *recall* sebesar 95,65%, dan *F1-score* sebesar 95,64%. Skenario pembagian data 80:20 memperoleh *mean accuracy* sebesar 88,33% dan standar deviasi sebesar 5,80% berdasarkan 10-*Fold Cross Validation*. Analisis *feature importance* menunjukkan bahwa intensitas latihan harian, kemampuan mengontrol emosi saat bertanding, total perolehan medali, frekuensi latihan mingguan, dan tingkat motivasi untuk menang merupakan variabel yang paling berpengaruh terhadap hasil prediksi kemenangan atlet. Dengan demikian, algoritma *Random Forest* terbukti berhasil dalam memprediksi hasil atlet pencak silat serta mengidentifikasi faktor-faktor yang berpengaruh terhadap keberhasilan atlet dalam pertandingan.

ABSTRACT

Dzilikrom, Ahmad Fadhli. 2026. **Victory Prediction of Pencak Silat Athletes in the 2025 Tapak Suci World Championship Using Random Forest**. Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Supervisors: (I) Dr. M. Amin Hariyadi, M.T. (II) Okta Qomaruddin Aziz, M.Kom.

Key words: Athlete Victory Prediction, Feature Importance, Machine Learning, Model Evaluation, Pencak Silat, Random Forest.

Pencak silat is a martial art that demands physical readiness, competitive experience, training intensity, and the psychological condition of athletes to achieve optimal match results. This study aims to build a prediction model for the victory of Pencak Silat athletes in the 2025 Tapak Suci World Championship using the Random Forest algorithm, assess the model's performance, and identify the variables that most significantly influence the prediction results. The data used in this study consists of primary data obtained from a survey of 226 Tapak Suci athletes featuring 15 attributes. The research stages included data preprocessing, data splitting using three scenarios (70:30, 80:20, and 90:10), model training, hyperparameter tuning using Grid Search Cross Validation, model evaluation using accuracy, precision, recall, and F1-score, and feature importance analysis. The results indicated that all data-splitting scenarios yielded high performance. In these scenarios, the Random Forest model achieved an accuracy of 95.65%, precision of 95.99%, recall of 95.65%, and an F1-score of 95.64%. The 80:20 data-splitting scenario obtained a mean accuracy of 88.33% and a standard deviation of 5.80% based on 10-Fold Cross Validation. Furthermore, the feature importance analysis revealed that daily training intensity, the ability to control emotions during matches, total medal count, weekly training frequency, and the level of motivation to win were the most influential variables affecting the prediction of an athlete's victory. Thus, the Random Forest algorithm is proven successful in predicting the outcomes of pencak silat athletes and identifying the factors that contribute to their success in competitions.

البحث مستخلص

ذي الإكرام، أحمد فضلي. ٢٠٢٦. **التنبؤ بفوز لاعبي رياضة بنجاك سيلات في بطولة العالم لتاباك سوسي ٢٠٢٥ باستخدام خوارزمية الغابة العشوائية (Random Forest)**. رسالة جامعية. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. محمد أمين هاريادي، الماجستير في الهندسة. والمشرف الثاني: أوكتا قمر الدين عزيز، الماجستير في الحاسوب.

الكلمات المفتاحية: تقييم النموذج، أهمية الخصائص،تعلّم الآلة، بنجاك سيلات، التنبؤ بفوز اللاعبين، الغابة العشوائية.

تُعدّ رياضة بنجاك سيلات من الرياضات القتالية التي تتطلب الاستعداد البدني، والخبرة في المنافسات، وشدة التدريب، والحالة النفسية للاعب من أجل تحقيق أفضل النتائج في المباريات. يهدف هذا البحث إلى بناء نموذج للتنبؤ بفوز لاعبي بنجاك سيلات في بطولة العالم لتاباك سوسي ٢٠٢٥ باستخدام خوارزمية الغابة العشوائية(Random Forest) ، وتقييم أداء النموذج، بالإضافة إلى تحديد المتغيرات الأكثر تأثيرًا في نتائج التنبؤ بفوز اللاعبين. تُستخدم في هذا البحث بيانات أولية ناتجة عن استبيان شمل ٢٢٦ لاعبًا من لاعبي تاباك سوسي، وتحتوي البيانات على ١٥ سمة. وتشمل مراحل البحث المعالجة المسبقة للبيانات (Data Preprocessing)،وتقسيم البيانات باستخدام ثلاثة سيناريوهات لتقسيم البيانات (٧٠:٣٠، ٨٠:٢٠، و٩٠:١٠)، وتدريب النموذج، وضبط المعلمات الفائقة (Hyperparameter Tuning) باستخدام طريقة البحث الشبكي مع التحقق المتقاطع (Grid Search Cross Validation)، وتقييم النموذج باستخدام مقاييس الدقة (Accuracy) ،والدقة (Precision)،والاستدعاء(Recall) ، ودرجة F1 (F1-score) ، بالإضافة إلى تحليل أهمية الخصائص (Feature Importance).أظهرت نتائج البحث أن جميع سيناريوهات تقسيم البيانات حققت أداءً مرتفعًا. وفي أحد السيناريوهات، حقق نموذج الغابة العشوائية (Random Forest) قيمة دقة (Accuracy) بلغت ٩٥,٦٥٪، وقيمة الدقة (Precision) بلغت ٩٥,٩٩٪، وقيمة الاستدعاء (Recall) بلغت ٩٥,٦٥٪، ودرجة F1 (F1-score) بلغت ٩٥,٦٤٪. كما حصل سيناريو تقسيم البيانات بنسبة ٨٠:٢٠ على متوسط دقة (Mean Accuracy) قدره ٨٨,٣٣٪، وانحراف معياري قدره ٥,٨٠٪ بناءً على التحقق المتقاطع ذي العشر طيات (Fold Cross Validation-10).وأظهر تحليل أهمية الخصائص (Importance (Feature أن شدة التدريب اليومي، والقدرة على التحكم في المشاعر أثناء المباراة، وإجمالي عدد الميداليات المحصل عليها، وتكرار التدريب الأسبوعي، ومستوى الدافع نحو الفوز، هي المتغيرات الأكثر تأثيرًا في نتائج التنبؤ بفوز اللاعبين. وبناءً على ذلك، ثبت أن خوارزمية الغابة العشوائية (Random Forest) قادرة على التنبؤ بنتائج لاعبي بنجاك سيلات بنجاح، وكذلك تحديد العوامل المؤثرة في تحقيق نجاح اللاعبين أثناء المنافسات.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Olahraga merupakan sarana penting dalam membangun kualitas sumber daya manusia yang sehat, kuat, dan berkarakter. Melalui olahraga, seseorang tidak hanya melatih kemampuan fisik, tetapi juga mengembangkan kedisiplinan, ketekunan, dan mental juang. Salah satu cabang olahraga yang memiliki nilai budaya dan spiritual tinggi di Indonesia adalah pencak silat, yang telah berkembang menjadi warisan budaya sekaligus media pembentukan akhlak dan sportivitas. Di antara berbagai perguruan pencak silat yang ada, Tapak Suci Putera Muhammadiyah dikenal sebagai organisasi yang mengintegrasikan unsur iman, ilmu, dan prestasi dalam setiap pembinaan atletnya (Hadiana *et al*., 2022).

Dalam menghadapi Kejuaraan Dunia Tapak Suci 2025, para atlet dituntut memiliki performa optimal agar mampu bersaing di tingkat internasional. Proses penilaian performa atlet selama ini masih banyak bergantung pada intuisi dan pengalaman pelatih, padahal faktor penentu kemenangan bersifat multidimensional, mencakup aspek fisik, teknik, taktik, psikologis, serta pengalaman bertanding. Menurut (Yulianto *et al*., 2022), peningkatan kapasitas fisik seperti daya tahan anaerob, kekuatan otot, dan efisiensi teknik berpengaruh langsung terhadap performa bertanding atlet pencak silat. Namun, faktor-faktor tersebut sering kali sulit diukur secara objektif, sehingga diperlukan pendekatan berbasis data *(data-driven approach)* untuk menilai potensi kemenangan atlet secara lebih terukur.

Kemajuan teknologi di bidang kecerdasan buatan (AI) serta sains data telah memfasilitasi evaluasi kinerja olahragawan secara lebih empiris dan terstruktur. Metode komputasi yang diaplikasikan untuk menganalisis, mengukur, serta mengestimasi performa atlet maupun hasil kompetisi ini dikenal sebagai *sport analytics* (Bunker & Susnjak, 2022). Sejumlah studi terdahulu berhasil mengonfirmasi efisiensi *machine learning* dalam mengolah data kompetisi untuk menyusun model proyeksi berakurasi tinggi. Sebagai contoh, (Jianjun *et al*., 2025) lewat publikasinya di *Scientific Reports* memformulasikan *Predictive Athlete Performance Modeling*. Model yang menyatukan variabel fisiologis, psikologis, dan biomekanika atlet ini mampu membukukan tingkat akurasi hingga 91,7%. Temuan tersebut menegaskan bahwa instrumen berbasis *machine learning* memiliki kapabilitas krusial dalam mengoptimalkan evaluasi sekaligus merancang strategi pelatihan atlet secara terukur.

Dalam ranah analisis prediktif berbasis *machine learning*, *Random Forest* menjadi salah satu algoritma yang sangat diminati karena efektivitasnya yang tinggi. Berdasarkan studi (Salman *et al*., 2024), mekanisme kerja metode ini bertumpu pada penggabungan luaran dari berbagai *Decision Tree* guna mengoptimalkan akurasi serta stabilitas proyeksi akhir. Kelebihan utama *Random Forest* terletak pada fleksibilitasnya dalam memproses tipe data non-linear dan variabel yang heterogen, baik yang bersifat kategorikal maupun numerik. Di samping itu, risiko terjadinya *overfitting* dapat diminimalisasi melalui penerapan teknik *bootstrap aggregating* (*bagging*). Algoritma ini juga memfasilitasi penentuan signifikansi variabel lewat fitur *feature importance* guna melihat faktor

penentu pada hasil prediksi. Adapun efisiensi model ini diperkuat oleh riset (Sanjaykumar *et al*., 2024), di mana *Random Forest* sukses menembus tingkat akurasi hingga 91,73% saat memproyeksikan hasil akhir sebuah pertandingan (Nandekdkar *et al*., 2024).

Meskipun demikian, belum terdapat penelitian yang secara spesifik menerapkan algoritma *Random Forest* untuk memprediksi kemenangan atlet Pencak Silat Tapak Suci, khususnya dengan mempertimbangkan kombinasi faktor fisik, pengalaman, dan psikologis. Inilah yang menjadi celah penelitian *(research gap)* yang ingin diisi melalui studi ini. Dengan membangun model prediksi berbasis *Random Forest*, diharapkan proses penilaian dan pembinaan atlet dapat dilakukan secara objektif, akurat, dan berbasis data, bukan hanya mengandalkan penilaian subjektif pelatih.

Dari perspektif Islam, upaya manusia dalam melakukan analisis, prencanaan, dan prediksi berbasis ilmu merupakan bagian dari ikhtiar yang diajarkan oleh Al-Qur'an. Allah SWT berfirma dalam Q.S. *Al-Hasyr* ayat 18:

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوا اتَّقُوا اللّٰهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍۚ وَاتَّقُوا اللّٰهَۗ اِنَّ اللّٰهَ خَبِيْرٌۢ ۢبِمَا تَعْمَلُوْنَ

"Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan." (QS. Al-Hasyr:18)

Menurut Tafsir Ibnu Katsir (1998), ayat ini mengandung perintah agar setiap manusia melakukan muhasabah (evaluasi diri) dan menyiapkan bekal amal untuk menghadapi hari esok, yaitu hari akhirat. Setiap Individu diperintahkan untuk

memperhatikan seluruh perbuatannya baik yang besar maupun kecil karena semuanya akan dipertanggungjawabkan di hadapan Allah (Al-Sheikh, 2005). Tafsir ini menegaskan pentingnya kesadaran, perencanaan, dan tanggung jawab atas setiap tindakan yang dilakukan di dunia. Dalam konteks penelitian ini, pesan ayat tersebut sejalan dengan prinsip ilmiah dalam memprediksi dan mempersiapkan masa depan berdasarkan pengetahuan yang terukur. Penerapan algoritma *Random Forest* dapat dimaknai sebagai bentuk ikhtiar manusia untuk melakukan perencanaan dan evaluasi secara sistematis terhadap performa atlet, tanpa mengabaikan bahwa hasil akhir sepenuhnya berada dalam ketentuan Allah SWT.

Mengacu pada pemaparan tersebut, studi ini diorientasikan untuk mengonstruksi model estimasi kemenangan bagi pesilat Tapak Suci dalam Kejuaraan Dunia 2025 lewat pemanfaatan algoritma *Random Forest*. Luaran dari riset ini diproyeksikan mampu menyumbang kontribusi teoretis terkait implementasi kecerdasan buatan pada cabang olahraga beladiri. Di samping itu, penelitian ini diharapkan dapat menjadi rujukan bagi para akademisi maupun pelatih dalam merancang sistem evaluasi performa atlet yang berbasis data, akurat, serta objektif. Lebih jauh lagi, kajian ini merepresentasikan bentuk penyelarasan antara sains modern dengan esensi spiritualitas Islam, yang mana pendekatan ilmiah dalam membedah data dikombinasikan dengan kesadaran moral serta nilai-nilai keimanan.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, bagaimana membangun model prediksi kemenangan atlet Pencak Silat menggunakan algoritma *Random Forest*, serta

bagaimana kinerja model tersebut dalam mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil pertandingan ?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan agar pembahasan lebih terarah dan fokus pada tujuan penelitian, yaitu sebagai berikut:

1. Data yang digunakan merupakan data primer hasil survei atlet Tapak Suci, yang mencangkup faktor fisik (usia, tinggi, berat, latihan), teknik (kategori tanding dan gaya bertanding), serta psikologis (motivasi, kepercayaan diri, kontrol emosi, dan kondisi fisik saat bertanding).
2. Analisis difokuskan pada klasifikasi hasil pertandingan (menang atau kalah) berdasarkan variabel performa yang telah dikumpulkan, dan hasil penelitian ditunjukan untuk membantu pelatih Tapak Suci dalam mengevaluasi serta mempersiapkan strategi latihan berbasis data.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, penelitian ini bertujuan untuk membangun model prediksi kemenangan atlet Pencak Silat pada Kejuaraan Dunia Tapak Suci 2025 menggunakan algoritma *Random Forest* berdasarkan data performa pertandingan, serta menilai tingkat akurasi model dan menentukan variabel yang paling berpengaruh terhadap hasil prediksi kemenangan atlet.

1.5 Manfaat Penelitian

Penelitian ini diproyeksikan dapat memberikan kontribusi signifikan, baik pada tataran teoritis maupun aplikatif. Dari aspek akademis, studi ini berfokus pada

pengayaan khazanah keilmuan di sektor *data mining* dan *machine learning*, khususnya yang berkaitan dengan utilitas algoritma *Random Forest* untuk keperluan komputasi analitik serta proyeksi performa atlet. Sementara dari dimensi praktis, hasil riset ini dapat diimplementasikan oleh jajaran pelatih maupun tim Tapak Suci sebagai instrumen pendukung dalam meninjau capaian atlet secara objektif sekaligus merumuskan program latihan yang lebih terarah, efektif, dan berbasis pada data empiris.

BAB II

STUDI PUSTAKA

Uraian pada bab ini memuat kerangka teoritis yang melandasi jalannya riset, sekaligus memaparkan sejumlah studi terdahulu yang sejalan dengan fokus pembahasan. Penyusunan kajian literatur ini ditujukan untuk memperkokoh pijakan teoretis, memberikan peta rujukan mengenai pemetaan riset sejenis yang pernah dilakukan, serta mempermudah dalam mengidentifikasi celah penelitian (*research gap*) sebagai basis pengembangan studi ini.

2.1 Penelitian Terdahulu

Penelitian terdahulu diposisikan sebagai kerangka acuan utama dalam mematangkan pengembangan studi yang sedang berjalan ini. Melalui penelusuran terhadap literatur-literatur sebelum ini, dapat dipetakan secara jelas evolusi metodologi yang diimplementasikan, capaian akhir yang diperoleh, serta aspek diferensiasi atau unsur kebaruan (*novelty*) yang diangkat sebagai fokus inti dalam penelitian ini. Ada pun sejumlah studi relevan yang melandasi topik ini dipaparkan pada rincian berikut.

Penelitian oleh Nguyen, (2022) berjudul *"The Application of Machine learning and Deep Learning in Sport: Predicting NBA Players' Performance and Popularity"* menerapkan beberapa algoritma *machine learning* seperti *Random Forest* (RF), *Gradient Boosting* (GB), dan *Deep Neural Network* (DNN) untuk memprediksi performa serta popularitas pemain NBA. Hasilnya menunjukkan bahwa *Random Forest* dan *Gradient Boosting* memberikan akurasi sekitar 90% dan

lebih stabil pada data terstruktur, sedangkan DNN kurang optimal karena keterbatasan jumlah data. Penelitian ini relevan karena karakteristik dataset pencak silat juga bersifat numerik dan terstruktur.

Selanjutnya, Lampis *et al*., (2023) melalui studi *Predictions of European Basketball Match Results with Machine learning Algorithms* membandingkan *Logistic Regression, Random Forest,* dan *XGBoost* untuk memprediksi hasil pertandingan bola basket Eropa. Hasilnya, *XGBoost* memperoleh akurasi tertinggi (80%) diikuti *Random Forest* (78%), dengan keunggulan RF pada kestabilan hasil klasifikasi. Penelitian ini mendukung penggunaan *Random Forest* untuk prediksi hasil pertandingan individu seperti pencak silat.

Bunker & Susnjak, (2022) dalam kajian *The Application of Machine learning Techniques for Predicting Match Results in Team Sport: A Review* menganalisis berbagai algoritma seperti *Random Forest, SVM, Naïve Bayes*, dan *Gradient Boosting*. Mereka menyimpulkan bahwa *Random Forest* dan *Gradient Boosting* memiliki performa paling konsisten dengan akurasi rata-rata 70–80%. Temuan ini menjadi dasar teoretis penggunaan *Random Forest* pada penelitian olahraga.

Sanjaykumar *et al*., (2024) dalam *Predicting Team Success in the Indian Premier League Cricket 2024 Season Using Random Forest Analysis* menggunakan *Random Forest* untuk memprediksi kesuksesan tim kriket berdasarkan data performa pemain. Model menghasilkan koefisien determinasi (R^2) sebesar 0.9173 atau akurasi 91,73%, menunjukkan kemampuan RF menangani hubungan variabel kompleks. Hasil ini memperkuat relevansi RF untuk data atlet pencak silat.

Penelitian oleh Nandekdkar *et al*., (2024) dengan judul *"Athlete Performance Prediction Using Random Forest"* juga menggunakan satu algoritma, yaitu *Random Forest*, untuk memprediksi performa atlet berdasarkan data biometrik seperti kecepatan, daya tahan, dan efisiensi gerak. Model ini mencapai akurasi sebesar 95%, menunjukkan kemampuan *Random Forest* dalam mengenali pola performa atlet secara individual. Penelitian ini sangat relevan karena pendekatannya serupa dengan penelitian ini, yaitu menggunakan indikator performa untuk memperkirakan hasil akhir pertandingan atau potensi kemenangan atlet.

Chen, (2024) dalam penelitiannya *"Factor and Correlation Analysis for Predicting Marathon Race Performance Using Machine learning Algorithms"* membandingkan *Random Forest* dan *Linear Regression* untuk menganalisis faktor performa pelari maraton. Hasilnya, RF memperoleh akurasi 89%, lebih tinggi dibandingkan LR (84%), karena mampu menangkap hubungan non-linear antarvariabel seperti usia dan waktu latihan. Kesamaan karakteristik data membuat penelitian ini relevan dengan pencak silat.

Penelitian oleh Ashraf *et al*., (2021) berjudul *"Predicting the Best Team Players of Pakistan Super League Using Machine learning Algorithms"* membandingkan beberapa algoritma, termasuk *Random Forest, SVM, Naïve Bayes,* dan *K-NN*. *Random Forest* mencapai akurasi tertinggi (100%) dalam mengklasifikasikan pemain terbaik berdasarkan performa statistik. Penelitian ini menunjukkan keunggulan RF dalam analisis performa individu di olahraga beregu.

Sementara itu, Ajgaonkar *et al*., (2021) melalui *Prediction of Winning Team Using Machine learning* menggunakan *Random Forest, SVM,* dan *Naïve Bayes* untuk memprediksi hasil pertandingan sepak bola. Meskipun *SVM* mencapai akurasi tertinggi (67%), *Random Forest* tetap unggul dalam kestabilan dan efisiensi. Hasil ini menegaskan bahwa efektivitas algoritma tergantung karakteristik dataset.

Menurut Zhou, (2022) dalam penelitiannya *"Sports Achievement Prediction and Influencing Factors Analysis Combined with Deep Learning Model"*, membandingkan *Deep Neural Network, Random Forest,* dan *SVM* untuk memprediksi prestasi atlet berdasarkan data fisik dan mental. *DNN* mencatat akurasi tertinggi (94%), namun RF (90%) lebih efisien untuk dataset berukuran sedang. Penelitian ini menunjukkan keseimbangan antara akurasi dan efisiensi yang menjadi alasan pemilihan RF dalam penelitian ini.

Terakhir penelitian oleh John & Damilola, (2021) berjudul *"Design and Implementation of a Predictive Model for Nigeria Local Football League"* membandingkan *Random Forest* dan *Decision Tree* untuk memprediksi hasil pertandingan sepak bola Nigeria. Hasil menunjukkan RF memiliki akurasi lebih tinggi (90%) dibandingkan DT (85%), membuktikan kemampuannya menghasilkan prediksi yang stabil dan akurat. Penelitian ini mendukung penggunaan RF untuk prediksi kemenangan berdasarkan data historis pertandingan.

Berdasarkan beberapa penelitian yang telah diuraikan di atas, ringkasan mengenai metode, hasil, dan relevansi penelitian terdahulu terhadap penelitian ini dapat dilihat pada Tabel 2.1 berikut.

Tabel 2. 1 Perbandingan Terdahulu

No	Peneliti	Input	Output	Metode	Hasil
1.	(Sanjaykumar *et al*., 2024)	Data performa pemain dan hasil pertandingan Indian Premier League (IPL) 2024	Prediksi kesuksesan tim kriket berdasarkan performa pemain	*Random Forest*	*Random Forest* terbukti efektif untuk data olahraga dengan banyak variabel yang saling berpengaruh, dengan nilai akurasi sebesar 91,73%.
2.	(Nandekdkar *et al*., 2024)	Data biometrik atlet seperti kecepatan, daya tahan, dan efisiensi gerak	Prediksi performa atlet individu	*Random Forest*	Algoritma *Random Forest* menunjukkan kemampuan tinggi dalam mengenali pola performa atlet dari data numerik, dengan tingkat akurasi sebesar 95%.
3.	(Chen, 2024)	Data pelari maraton (usia, waktu latihan, kecepatan rata-rata)	Prediksi hasil lomba maraton	*Random Forest* dan *Linear Regression*	Hasil penelitian menunjukkan bahwa *Random Forest* memberikan akurasi 89%, lebih tinggi dibanding *Linear Regression* sebesar 84%.
4.	(Lampis *et al*., 2023)	Data hasil pertandingan bola basket di liga Eropa	Prediksi hasil pertandingan (menang/kalah)	*Logistic Regression, Random Forest, XGBoost*	Hasil menunjukkan bahwa *XGBoost* mencapai akurasi 80%, sedangkan *Random Forest* sebesar 78%.
5.	(Bunker & Susnjak, 2022)	Dataset hasil pertandingan berbagai olahraga beregu	Kajian akurasi berbagai algoritma ML dalam prediksi olahraga	*Random Forest, SVM, Gradient Boosting, Naïve Bayes, KNN*	Kajian menunjukkan bahwa *Random Forest* dan *Gradient Boosting* memberikan hasil terbaik di antara metode lain, dengan rentang akurasi 70–80%.
6.	(Zhou, 2022)	Data performa fisik dan mental atlet	Prediksi pencapaian olahraga berdasarkan faktor performa	*Deep Neural Network (DNN), Random Forest, SVM*	*Deep Neural Network* memperoleh hasil terbaik, diikuti *Random Forest* dan SVM, dengan akurasi masing-masing 94%, 90%, dan 88%.
7.	(H. N. Nguyen, 2022)	Data statistik performa dan popularitas pemain NBA	Prediksi performa serta popularitas pemain	*Random Forest, Gradient Boosting, Deep Neural Network*	*Random Forest* dan *Gradient Boosting* menghasilkan performa lebih baik dibanding *DNN*, dengan akurasi keduanya mencapai 90%.

No	Peneliti	Input	Output	Metode	Hasil
8.	(Ashraf *et al*., 2021)	Data performa pemain kriket *(batting dan bowling)*	Klasifikasi pemain terbaik	*Random Forest, SVM, Naïve Bayes, KNN, Linear Regression*	*Random Forest* menunjukkan performa tertinggi dibanding metode lain, dengan akurasi sebesar 100%.
9.	(Ajgaonkar *et al*., 2021)	Data statistik pertandingan sepak bola *English Premier League* (EPL)	Prediksi hasil pertandingan (menang/seri /kalah)	*Random Forest, SVM, Naïve Bayes*	SVM menunjukkan hasil terbaik dibanding *Random Forest* dan *Naïve Bayes,* dengan akurasi masing-masing 67%, 60%, dan 56%.
10.	(John & Damilola, 2021)	Data historis pertandingan *Nigeria Local Football League*	Prediksi hasil pertandingan	*Random Forest dan Decision Tree*	*Random Forest* memberikan hasil lebih stabil dan akurat dibanding *Decision Tree ,* dengan akurasi masing-masing 90% dan 85%.

Menilik pemaparan dari sepuluh riset terdahulu tersebut, dapat ditarik kesimpulan bahwa Random Forest tergolong sebagai salah satu teknik machine learning yang paling mendominasi dan teruji memberikan kinerja optimal pada berbagai pemodelan prediktif di dunia olahraga. Keunggulan algoritma ini terletak pada kapabilitasnya dalam mengeksekusi data yang rumit dengan multi-variabel, di samping konsistensinya dalam menyajikan luaran klasifikasi maupun regresi berpresisi tinggi yang umumnya berada pada rentang 80% sampai 100%.

Kendati beberapa studi mengindikasikan bahwa arsitektur lain semacam Support Vector Machine maupun Deep Neural Network mampu menyajikan akurasi sedikit lebih unggul pada situasi spesifik, Random Forest tetap dinilai lebih superior dalam aspek efisiensi pemrosesan data, kemudahan interpretasi model, serta ketahanan terhadap gejala overfitting. Atas dasar pertimbangan tersebut, metodologi ini dipandang sebagai instrumen yang paling relevan untuk diimplementasikan pada studi ini, yang mana fokus utamanya adalah

memproyeksikan kemenangan atlet pencak silat dengan mengandalkan data capaian performa beserta indikator penentu terkait lainnya.

2.2 Pencak Silat

Sebagai bagian dari kekayaan budaya nasional, pencak silat dipahami sebagai perpaduan antara aktivitas olahraga dan seni bela diri asli Indonesia. (Sasmita *et al*., 2023) menjelaskan bahwa frasa *pencak* menitikberatkan pada estetika atau keindahan dari gerakan dasar pertahanan diri, sementara kata *silat* lebih berorientasi pada pengaplikasian taktik bela diri demi proteksi diri. Melalui kombinasi tersebut, pencak silat tidak sekadar berfokus pada ketangkasan fisik semata, melainkan turut berperan penting dalam penanaman nilai-nilai kedisiplinan, penempaan karakter, serta kontrol diri.

Dari aspek institusional, Ikatan Pencak Silat Indonesia (IPSI) telah dibentuk sejak tahun 1948 dan beroperasi di bawah naungan Federasi Pencak Silat Antarbangsa (PERSILAT). Pada ranah kompetisi, kejuaraan pencak silat secara garis besar dikelompokkan ke dalam dua divisi utama, yakni kategori seni dan tanding. Berdasarkan studi (Razak, 2024), divisi tanding mengadu dua atlet yang menempati sudut biru dan merah dalam laga sepanjang tiga babak, di mana masing-masing babak berlangsung selama dua menit. Sistem poin pada kategori ini dihitung berdasarkan keabsahan serta efektivitas serangan yang dilancarkan, meliputi bantingan, sapuan, tendangan, maupun pukulan. Sebaliknya, poin dapat dikurangi atau bahkan berujung pada diskualifikasi apabila atlet melakukan pelanggaran ataupun perilaku yang mencederai nilai sportivitas.

Kemenangan dalam pertandingan pencak silat tidak hanya dipengaruhi oleh kemampuan teknik, tetapi juga oleh aspek fisik, taktis, dan psikologis. (Patah *et al*., 2021) menyebutkan bahwa kebugaran jasmani dan kekuatan otot merupakan fondasi utama performa atlet. (Subekti *et al*., 2021) menambahkan bahwa strategi pertahanan dan serangan balik yang efisien menjadi kunci kemenangan, sedangkan (Alisti *et al*., 2025) menekankan pentingnya kesiapan mental, kontrol emosi, dan rasa percaya diri dalam menjaga stabilitas performa. Dengan demikian, pencak silat menuntut keseimbangan antara kekuatan fisik, strategi, dan ketenangan mental. Pemahaman terhadap faktor-faktor tersebut penting dalam penelitian ini karena akan dijadikan dasar untuk membangun model prediksi kemenangan atlet menggunakan algoritma *machine learning*.

2.3 Prediksi

Prediksi dirumuskan sebagai sebuah mekanisme untuk mengestimasi suatu peristiwa atau luaran di masa depan dengan mengacu pada rekaman data serta tren historis yang ada. Di dalam lingkup sains data, proses peramalan ini mengintegrasikan metode statistika beserta algoritma machine learning guna mengidentifikasi korelasi antar-variabel, yang pada gilirannya mampu memberikan proyeksi terhadap situasi-situasi baru (Kampezidou *et al*., 2024). Sementara itu, (KC *et al*., 2021) memaparkan bahwa kerangka prediksi umumnya mengadopsi asas supervised learning, yakni suatu kondisi di mana model diberikan pelatihan lewat kombinasi data masukan (input) dan keluaran (output) agar dapat menangkap pola yang signifikan sekaligus menyajikan hasil klasifikasi maupun estimasi nilai yang berpresisi tinggi.

Dalam bidang olahraga, teknik prediksi banyak digunakan untuk memperkirakan hasil pertandingan, performa atlet, maupun strategi permainan berdasarkan data statistik dan performa fisik. (J. M. Nguyen *et al*., 2022) menyatakan bahwa penerapan *machine learning* memungkinkan proses analisis dilakukan secara objektif dan cepat dibandingkan penilaian manual. Model seperti *Random Forest*, *Support Vector Machine* (SVM), dan *Gradient Boosting* terbukti mampu mengolah data kompleks yang melibatkan berbagai faktor seperti kekuatan, kecepatan, pengalaman bertanding, serta kondisi mental atlet. Dengan demikian, peluang kemenangan dapat dinilai secara ilmiah dan berbasis data, bukan sekadar intuisi pelatih.

Pada penelitian ini, konsep prediksi diterapkan untuk memperkirakan kemungkinan kemenangan atlet Pencak Silat Tapak Suci pada Kejuaraan Dunia 2025 berdasarkan data performa pertandingan. Pendekatan ini tidak hanya menghasilkan klasifikasi akhir berupa menang atau kalah, tetapi juga membantu mengidentifikasi variabel yang paling berpengaruh terhadap hasil pertandingan. Dengan penerapan algoritma *machine learning*, penelitian ini diharapkan dapat memberikan pandangan baru dalam proses evaluasi performa dan penyusunan strategi latihan atlet secara objektif dan berbasis data.

2.4 Algoritma *Random Forest*

Evolusi di sektor teknologi *machine learning* telah menghadirkan beragam variasi algoritma yang memiliki kapabilitas mumpuni untuk mengeksekusi analisis data secara efisien serta berakurasi tinggi. Di antara sekian banyak opsi untuk menyelesaikan problem prediktif maupun klasifikasi, *Random Forest* memegang

posisi sebagai salah satu arsitektur yang paling kerap diandalkan dan dikategorikan ke dalam rumpun *ensemble learning methods*. Berdasarkan kajian (Salman *et al*., 2024), prinsip dasar yang diusung oleh algoritma ini melibatkan pengintegrasian kumpulan model sederhana (*weak learners*) guna membangun sebuah model terpadu yang jauh lebih tangguh (*strong learner*), yang pada gilirannya mampu mengoptimalkan stabilitas sekaligus ketepatan hasil proyeksi akhir.

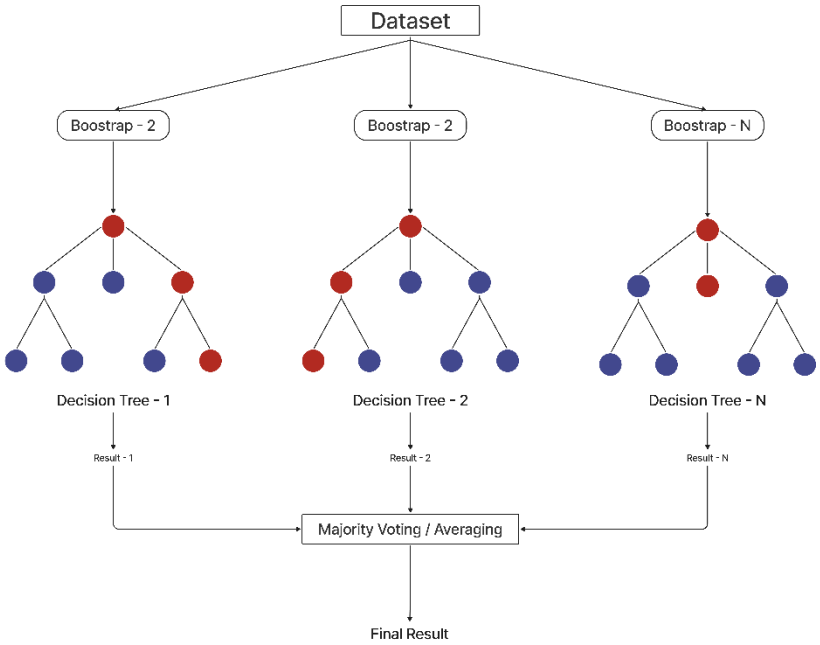
Sebagaimana dipaparkan oleh (J. M. Nguyen *et al.*, 2022), *Random Forest* dipahami sebagai bentuk pengembangan dari model *Decision Tree* yang beroperasi melalui pembentukan sekumpulan pohon keputusan terpisah secara acak memanfaatkan variasi data latih serta subset fitur yang berbeda. Masing-masing pohon tersebut nantinya bakal memproduksi estimasi mandiri, yang mana keputusan final dikombinasikan lewat mekanisme *majority voting* untuk persoalan klasifikasi atau menerapkan fungsi *averaging* apabila berupa kasus regresi. Strategi ini membekali model dengan kapabilitas generalisasi yang optimal sekaligus resistensi yang kokoh terhadap gejala *overfitting*, mengingat kegagalan atau galat pada satu entitas pohon dapat ditoleransi oleh kontribusi pohon lainnya. Di samping itu, perancangan tiap pohon mengadopsi teknik *bootstrap aggregating* (*bagging*), sebuah metode di mana sampel data latih diambil acak dengan sistem pengembalian sehingga memicu karakteristik yang unik pada setiap pohon yang terbentuk.

Nilai plus utama dari *Random Forest* bersumber pada efektivitasnya dalam mengeksekusi himpunan data bervolume besar, mengelola atribut numerik sekaligus kategorikal, serta mengidentifikasi korelasi non-linear antar-variabel tanpa terikat pada asumsi distribusi data tertentu (Mienye & Jere, 2024). Di samping

itu, metode ini dibekali fitur penilai signifikansi (*feature importance*) guna mendeteksi variabel penentu yang memegang andil paling krusial terhadap akurasi prediksi. Karakteristik tersebut menjadikan *Random Forest* tidak sekadar diandalkan untuk kebutuhan klasifikasi, melainkan juga sangat aplikatif untuk membedah serta menginterpretasikan faktor-faktor pemicu suatu fenomena. Eksistensi model ini dalam menembus tingkat presisi yang tinggi pun telah divalidasi oleh berbagai studi terdahulu di lintas sektor, mulai dari ranah olahraga, medis, hingga industri finansial (Sanjaykumar *et al*., 2024).

Meskipun memiliki banyak keunggulan, *Random Forest* juga memiliki keterbatasan, terutama dalam hal interpretasi hasil dan efisiensi komputasi. Jumlah pohon yang besar dapat meningkatkan waktu pelatihan dan kebutuhan memori, serta membuat hasil model sulit dijelaskan secara intuitif karena merupakan kombinasi dari banyak pohon keputusan. Namun, kelemahan tersebut dapat diminimalkan dengan pengaturan parameter yang tepat, seperti jumlah pohon (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*), sehingga diperoleh keseimbangan antara akurasi dan efisiensi komputasi (J. M. Nguyen *et al*., 2022).

Secara umum, *Random Forest* bekerja dengan membentuk banyak pohon keputusan yang independen menggunakan kombinasi acak data dan fitur, kemudian menggabungkan hasil prediksinya melalui proses *voting* mayoritas. Pendekatan ini memungkinkan algoritma menghasilkan prediksi yang stabil, akurat, dan tahan terhadap *noise* dalam data. Gambar 2.1 berikut menggambarkan alur umum proses kerja algoritma *Random Forest.*



Gambar 2. 1 Alur Metode *Random Forest*

Berdasarkan Gambar 2.1, dapat dijelaskan bahwa setiap *Decision Tree* dalam *Random Forest* dibangun secara independen menggunakan subset data dan fitur yang berbeda. Hasil prediksi dari masing-masing pohon kemudian digabungkan melalui proses *majority voting* untuk menentukan keputusan akhir. Pendekatan ini memastikan bahwa kesalahan pada satu pohon dapat dikompensasi oleh pohon lainnya, sehingga model secara keseluruhan memiliki tingkat akurasi dan kestabilan yang lebih tinggi dibandingkan dengan metode *Decision Tree* tunggal.

2.5 Evaluasi Model

Tahap pengujian atau evaluasi model memegang peranan krusial dalam siklus *machine learning* untuk menakar sejauh mana efektivitas sistem dalam memproyeksikan data uji. Melalui pengujian ini, peneliti dapat mengukur tingkat

keandalan, akurasi, serta kapabilitas generalisasi model saat dihadapkan pada himpunan data baru yang belum pernah dipelajari sebelumnya (Miller *et al*., 2024). Absennya proses kalkulasi evaluasi yang presisi berisiko memicu terjadinya *overfitting*, yakni suatu anomali di mana model menunjukkan performa yang sangat optimal pada fase latihan (*training data*) namun justru lumpuh dan gagal menyajikan prediksi yang valid ketika mengeksekusi data uji (Gori *et al*., 2024).

Pada rumpun permasalahan klasifikasi seperti halnya studi ini, agenda validasi dijalankan dengan memanfaatkan sederet teknik pengujian serta metrik yang diselaraskan terhadap karakter dataset sekaligus jenis algoritma yang diimplementasikan. (Aqmar *et al*., 2025) menyebutkan beberapa parameter ukur yang lazim dipakai meliputi *F1-score*, *Recall*, Presisi, Akurasi, dan *Confusion matrix*, sementara pendekatan semacam *K-Fold Cross Validation* kerap dipilih untuk menguji konsistensi kinerja model. Di samping itu, (Jubeile Mark Baladjay *et al*., 2023) menambahkan bahwa luaran dari fase evaluasi ini tidak sekadar menjadi indikator keberhasilan sistem, melainkan berperan sebagai pijakan utama dalam mengeksekusi *model tuning* serta memacu optimalisasi algoritma. Oleh sebab itu, tahapan penilaian ini memegang andil yang sangat strategis dalam menjamin bahwa arsitektur yang diaplikasikan seperti *Random Forest* dalam riset ini mampu menyajikan hasil proyeksi yang akurat, konsisten, serta optimal.

Secara keseluruhan, penilaian performa model dikerjakan dengan mengomparasikan nilai proyeksi sistem terhadap nilai riil (*actual value*) pada himpunan data uji menggunakan parameter ukur tertentu. Prosedur ini memfasilitasi peneliti untuk memetakan sejauh mana kompetensi model dalam

menangkap tren pola data sekaligus melacak galat atau kekeliruan klasifikasi yang muncul. Terkait hal tersebut, (Miller *et al*., 2024) menggarisbawahi bahwa bersandar pada satu metrik tunggal belum memadai untuk meninjau kapabilitas model secara utuh. Oleh karenanya, integrasi dari multi-metrik evaluasi kerap diaplikasikan guna memperoleh representasi performa yang lebih objektif dan menyeluruh, khususnya saat berhadapan dengan karakteristik dataset yang memiliki distribusi kelas tidak merata (*imbalanced dataset*).

Dalam penelitian ini, evaluasi model digunakan untuk mengukur kemampuan algoritma *Random Forest* dalam memprediksi kemenangan atlet Pencak Silat Tapak Suci berdasarkan variabel performa yang diperoleh dari hasil kuesioner dan data pertandingan. Hasil evaluasi diharapkan dapat menunjukkan tingkat akurasi dan keandalan model dalam mengklasifikasikan hasil pertandingan ke dalam kategori menang atau kalah, serta memberikan dasar ilmiah bagi pelatih dan akademisi dalam proses analisis performa atlet berbasis data.

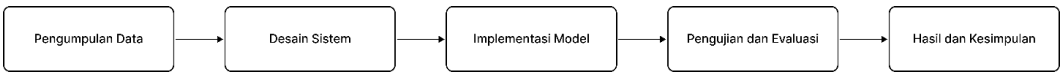
BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Rancangan penelitian ini menguraikan secara terstruktur mengenai rangkaian fase yang ditempuh dalam proses studi untuk memformulasikan model estimasi kemenangan pesilat Tapak Suci pada Kejuaraan Dunia 2025 lewat penerapan algoritma *Random Forest*. Kerangka pendekatan yang dipilih mengacu pada metode kuantitatif berbasis *machine learning*, didasarkan pada kompetensinya dalam memproses variabel berkategori maupun angka (numerik) demi melahirkan klasifikasi model yang konsisten, objektif, serta berpresisi tinggi.

Siklus dalam desain riset ini diformulasikan secara sistematis agar tiap fasenya saling terintegrasi dan berkonvergensi pada pencapaian target penelitian. Alur kerja ini diinisiasi dari perolehan atau pengumpulan data hingga berlanjut pada pemaknaan hasil serta perumusan simpulan akhir. Ada pun gambaran skema alur riset secara makro direpresentasikan melalui Gambar 3.1 di bawah ini:



Gambar 3. 1 Desain Penelitian

Penjelasan tahapan desain penelitian:

1. Pengumpulan Data

   Tahap ini merupakan proses awal dalam penelitian yang bertujuan memperoleh data primer dari atlet Pencak Silat Tapak Suci melalui penyebaran kuesioner survei secara langsung pada peserta Kejuaraan Dunia

2025. Data yang dikumpulkan mencakup karakteristik dan performa atlet, seperti usia, jenis kelamin, tinggi badan, berat badan, frekuensi latihan, durasi latihan, pengalaman bertanding, jumlah kemenangan, kekalahan, serta hasil pertandingan (menang atau kalah). Data tersebut berfungsi sebagai dasar dalam proses pembentukan dan pelatihan model prediksi.

2. Desain Sistem

   Pada tahap ini dilakukan perancangan alur kerja sistem prediksi kemenangan atlet, yang mencakup proses input data, *preprocessing,* penerapan model *Random Forest,* hingga output berupa hasil prediksi dan analisis variabel yang berpengaruh terhadap peluang kemenangan. Desain sistem ini menggambarkan hubungan antarproses secara konseptual agar alur prediksi dapat dijalankan secara efisien dan terukur.

3. Implementasi Model

   Apabila arsitektur sistem telah ditetapkan, langkah selanjutnya berfokus pada pengaplikasian algoritma *Random Forest* dengan mengandalkan dataset yang tuntas melewati tahapan pra-pemrosesan (*preprocessing*). Fase eksekusi ini mengintegrasikan proses pembelajaran (*training*) pada sampel data latih sekaligus pengujian (*testing*) memanfaatkaan sampel data uji. Optimalisasi model kemudian dilakukan melalui penyesuaian atau tuning konfigurasi variabel, seperti kedalaman maksimum pohon (*max depth*) serta kuantitas *estimators*, guna menghasilkan performa proyeksi yang paling ideal.

4. Evaluasi dan Pengujian

   Fase ini diarahkan untuk mengukur sejauh mana kapabilitas model dalam menghasilkan proyeksi yang tepat. Untuk meninjau sejauh mana keberhasilan maupun kegagalan sistem dalam melakukan klasifikasi, digunakan alat ukur *Confusion matrix*, di samping kalkulasi parameter performa lainnya seperti *F1-score*, *Recall*, *Precision*, serta *Accuracy*. Luaran dari pengujian tersebut selanjutnya dijadikan acuan untuk merefleksikan efektivitas model serta memastikan apakah kinerjanya sudah mencapai ambang batas akurasi yang ditargetkan.

5. Hasil dan Kesimpulan

   Tahap terakhir adalah menganalisis hasil pengujian model dan menginterpretasikan faktor-faktor yang paling berpengaruh terhadap kemenangan atlet. Dari hasil tersebut, diperoleh kesimpulan mengenai performa algoritma *Random Forest* dalam memprediksi hasil pertandingan serta rekomendasi penerapannya dalam mendukung proses pembinaan atlet secara berbasis data dan teknologi.

Sebagai keseluruhan, desain penelitian ini berperan sebagai kerangka metodologis yang mengintegrasikan antara pengumpulan data lapangan, penerapan algoritma machine learning, dan analisis hasil berbasis evaluasi kuantitatif. Dengan rancangan yang sistematis ini, diharapkan penelitian mampu menghasilkan model prediksi yang akurat serta memberikan kontribusi nyata bagi pengembangan strategi pelatihan dan peningkatan prestasi atlet Tapak Suci di tingkat internasional.

3.2 Pengumpulan Data

Desain Data dalam penelitian ini diperoleh melalui penyebaran kuesioner survei kepada atlet Tapak Suci yang berpartisipasi dalam Kejuaraan Dunia Tapak Suci 2025. Kuesioner disusun untuk mengumpulkan data karakteristik atlet, riwayat pertandingan, pola latihan, serta faktor psikologis yang berpotensi memengaruhi hasil pertandingan. Penyusunan atribut dan struktur data mengacu pada pendekatan penelitian di bidang olahraga bela diri yang memanfaatkan data individu atlet dan performa pertandingan sebagai masukan model prediksi berbasis machine learning, sebagaimana diterapkan pada penelitian *Intelligent Decision-Making System for Martial Art* (Li *et al*., 2021).

Pengumpulan data dilakukan secara langsung di lokasi kejuaraan dengan dukungan pelatih dan panitia pertandingan guna memastikan keaslian dan keakuratan data yang diperoleh. Data yang digunakan dalam penelitian ini merupakan data responden yang memenuhi kriteria kelengkapan dan relevansi terhadap kebutuhan analisis, sehingga layak digunakan sebagai masukan dalam proses pemodelan dan evaluasi sistem prediksi hasil pertandingan.

Tabel 3. 1 Deskripsi Atribut Penelitian

No	Nama Atribut	Jenis Data	Deskripsi
1.	Jenis kelamin	Kategorikal	Jenis kelamin atlet (Laki-laki/Perempuan).
2.	Usia saat ini	Numerik *(integer)*	Usia atlet saat kejuaraan (dalam tahun).
3.	Tinggi badan	Numerik *(float)*	Tinggi badan atlet (dalam sentimeter)
4.	Berat badan	Numerik *(float)*	Berat badan atlet yang diukur dalam satuan kilogram (kg).
5.	Berapa jumlah pertandingan resmi yang pernah anda ikuti sejauh ini?	Numerik *(integer)*	Jumlah total pertandingan resmi yang pernah diikuti oleh atlet.
6.	Dari seluruh pertandingan yang anda ikuti, berapa	Numerik *(integer)*	Jumlah total kemenangan yang diperoleh atlet dari seluruh pertandingan yang diikuti.

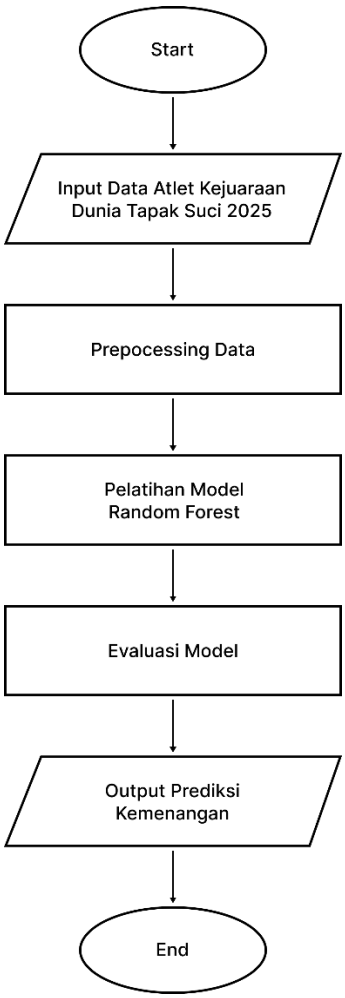
No	Nama Atribut	Jenis Data	Deskripsi
	jumlah kemenangan yang telah anda raih?		
7.	Dari seluruh pertandingan yang ada ikuti, berapa jumlah kekalahan yang anda alami?	Numerik *(integer)*	Jumlah total kekalahan yang dialami atlet dari seluruh pertandingan yang diikuti.
8.	Total keseluruhan medali (emas, perak, perunggu) yang pernah anda dapatkan selama mengikuti kejuaraan?	Numerik *(integer)*	Jumlah total medali (emas, perak, dan perunggu) yang diperoleh atlet selama mengikuti kejuaraan.
9.	Berapa kali anda berlatih dalam satu hari selama masa persiapan kejuaraan?	Numerik *(integer)*	Jumlah sesi latihan yang dilakukan atlet dalam satu hari selama masa persiapan kejuaraan.
10.	Berapa kali anda biasanya berlatih dalam seminggu?	Numerik *(integer)*	Jumlah sesi latihan yang dilakukan atlet dalam satu minggu.
11.	Berapa lama rata-rata durasi latihan anda dalam satu sesi?	Numerik *(float)*	Rata-rata durasi latihan dalam satu sesi yang diukur dalam satuan jam.
12.	Sejak tahun berapa anda mulai aktif bertanding di kejuaraan pencak silat?	Numerik *(integer)*	Tahun pertama atlet mulai aktif mengikuti kejuaraan pencak silat.
13.	Seberapa tinggi tingkat kepercayaan diri anda saat bertanding?	Kategorikal	Tingkat kepercayaan diri atlet saat bertanding (misalnya: rendah, sedang, tinggi).
14.	Seberapa besar motivasi anda untuk menang dalam pertandingan ini?	Kategorikal	Tingkat motivasi atlet untuk meraih kemenangan dalam pertandingan (misalnya: rendah, sedang, tinggi).
15.	Sejauh mana anda mampu mengontrol emosi selama pertandingan berlangsung?	Kategorikal	Kemampuan atlet dalam mengontrol emosi selama pertandingan berlangsung (misalnya: rendah, sedang, tinggi).

3.3 Desain Sistem

Desain sistem pada penelitian ini menjelaskan alur kerja model prediksi kemenangan atlet Pencak Silat Tapak Suci menggunakan algoritma *Random Forest.* Sistem ini berfungsi untuk mengubah data hasil kuesioner atlet menjadi

informasi prediktif berupa hasil pertandingan (menang atau kalah) berdasarkan faktor-faktor fisik, latihan, pengalaman, dan psikologis.

Tujuan dari perancangan sistem ini adalah memberikan gambaran menyeluruh mengenai proses pengolahan data hingga menghasilkan prediksi akhir, sehingga setiap tahapan dalam penelitian dapat dipahami secara sistematis dan terukur.



Gambar 3. 2 Flowchart Desain Sistem

Gambar 3.2 menunjukkan alur sistem prediksi kemenangan atlet Tapak Suci pada Kejuaraan Dunia 2025. Secara umum, sistem dibagi menjadi tiga komponen utama, yaitu *Input*, *Process*, dan *Output*.

1. Input

   Tahap ini merupakan sumber data utama yang diperoleh dari hasil kuesioner atlet Tapak Suci peserta kejuaraan. Berdasarkan hasil pengolahan data, jumlah dataset yang digunakan sebanyak 226 data atlet dengan 15 fitur yang merepresentasikan karakteristik atlet.

   Data yang digunakan dalam penelitian ini dikelompokkan menjadi empat kategori utama, yaitu:

   a. Data fisik: usia, tinggi badan, berat badan, dan jenis kelamin.
      1) Data teknik dan pengalaman: jumlah pertandingan, jumlah kemenangan, jumlah kekalahan, total medali, dan tahun mulai aktif bertanding.
      2) Data psikologis: tingkat motivasi, kepercayaan diri, dan kemampuan mengontrol emosi.
      3) Data latihan: frekuensi latihan per hari, frekuensi latihan per minggu, dan durasi latihan.

   Data ini kemudian disusun dalam format tabular dan digunakan sebagai dataset utama pada tahap pemrosesan.

2. Process

Tahap Tahap process merupakan inti dari sistem prediksi yang meliputi preprocessing data, pelatihan model, *tuning hyperparameter*, dan evaluasi model menggunakan algoritma *Random Forest*.

a. Prepocessing Data

Tahap ini bertujuan menyiapkan data agar dapat diproses oleh model *machine learning* dengan Langkah-langkah berikut:

1) Pemisahan data: memisahkan fitur (X) dan target (Y), di mana target adalah hasil pertandingan.
2) Pembersihan data, memastikan tidak terdapat data kosong atau duplikat.
3) Encoding data mengubah atribut kategorikal menjadi numerik.
4) Normalisasi data menggunakan metode *Min-Max Scalling* untuk menyamakan skala data.

Hasil preprocessing menunjukkan bahwa dataset terdiri dari 226 data dengan 15 fitur numerik dan tanpa missing value, sehingga siap digunakan untuk pemodelan.

b. Pembagian Data (Split Data)

Dataset dibagi menjadi data latih (*training* data) dan data uji (*testing* data) menggunakan tiga skenario pembagian data, yaitu 70%:30%, 80%:20%, dan 90%:10%. Penggunaan beberapa skenario pembagian data bertujuan untuk mengevaluasi konsistensi dan kestabilan performa model *Random Forest* terhadap variasi proporsi data latih dan data uji.

Seluruh proses pembagian data dilakukan menggunakan metode *stratified sampling* sehingga distribusi kelas target tetap terjaga secara proporsional pada data latih maupun data uji.

c. Modelling *(Random Forest)*

Model dibangun menggunakan algoritma *Random Forest*, yang bekerja dengan membentuk beberapa *decision tree* dan menghasilkan prediksi melalui mekanisme *majority voting*.

d. Prepocessing Data

Dilakukan pengujian kombinasi parameter:

1) *n*_estimator: 100, 200, 300
2) *max_depth*: 10, 20, None

Penerapan metode *Grid Search Cross Validation* (GridSearchCV) diandalkan untuk mengeksekusi tahapan penalaan atau optimasi *hyperparameter*. Melalui pendekatan ini, seluruh variasi kombinasi *hyperparameter* yang telah ditetapkan bakal diuji secara menyeluruh demi memetakan susunan konfigurasi yang mampu menyajikan performa paling superior. Lantaran tiap pasang parameter dinilai memanfaatkan skema *cross validation*, penentuan nilai parameter final menjadi tidak bias dan tidak sekadar bertumpu pada satu kali pembagian dataset (*data splitting*). Luaran optimal dari rangkaian proses inilah yang kemudian diintegrasikan sebagai pondasi parameter terbaik dalam mengonstruksi model *Random Forest*.

e. Evaluasi Model

Evaluasi model dilakukan dengan empat metrik evaluasi seperti *Accuracy, Precision, Recall,* dan *F1-Score*.

3. Output

Tahap output merupakan hasil akhir dari keseluruhan sistem prediksi, adapun keluaran yang dihasilkan berupa:

1) Prediksi hasil pertandingan atlet Tapak Suci, dengan dua kemungkinan nilai yaitu menang (1) atau kalah (0).
2) Identifikasi fitur paling berpengaruh terhadap hasil prediksi, seperti frekuensi latihan, pengalaman bertanding, atau tingkat motivasi.

Informasi ini dapat digunakan oleh pelatih dan tim Tapak Suci sebagai dasar pertimbangan dalam menentukan strategis latihan, pembinaan atlet, dan analisis perfroma secara objektif serta berbasis data.

3.4 *Prepocessing* Data

Eksekusi pra-pemrosesan (*preprocessing*) data diorientasikan untuk mengondisikan dataset hasil kuesioner agar berada dalam status siap pakai pada fase pembelajaran model *machine learning*. Agenda ini diselenggarakan guna menjamin bahwa seluruh rekaman data telah bersih, selaras (konsisten), sekaligus dikonversi ke dalam representasi angka (numerik) yang kompatibel untuk dieksekusi oleh algoritma *Random Forest*. Secara garis besar, pengerjaan pra-pemrosesan ini mencakup empat pilar tahapan inti, yakni pembersihan data (*data*

cleaning), transformasi atau *encoding* data, penyeragaman (normalisasi) data, serta pengelompokan data (*split data*).

3.4.1 Pembersihan Data

Rangkaian prosedur normalisasi diselenggarakan demi menyelaraskan proporsi skala pada tiap fitur numerik. Hal ini dilakukan supaya sebaran nilainya berada dalam cakupan yang setara, sehingga meminimalisasi risiko adanya parameter tertentu yang mendominasi jalannya fase pembelajaran model. Terkait hal tersebut, teknik penyeragaman yang diimplementasikan pada riset ini mengadopsi metode *Min–Max Scaling*. Merujuk pada Allorerung dkk. (2024), pendekatan ini dirancang untuk mengonversi nilai dataset ke dalam batasan interval antara 0 hingga 1. Ada pun formulasi matematis dari *Min-Max Scaling* dijabarkan melalui persamaan di bawah ini:

Tabel 3. 2 Contoh Hasil Pembersihan Data

Nama Atlet	Tinggi Badan (awal)	Tinggi Badan (hasil)	Berat Badan (awal)	Berat Badan (hasil)
Nasrullah	166 cm	166	60 kg	60
Prima Armanda	176 cm	176	88 kg	88
Erika Indriyani	158 cm	158	44 kg	44

3.4.2 Pembersihan Data

Beberapa atribut dalam dataset penelitian berbentuk kategorikal sehingga perlu dikonversi menjadi nilai numerik agar dapat diproses oleh algoritma machine learning. Proses pengubahan data dilakukan menggunakan beberapa teknik encoding yang disesuaikan dengan karakteristik masing-masing atribut.

1. Label *Encoding* digunakan untuk atribut yang bersifat biner atau memiliki dua kategori, seperti Jenis Kelamin, dengan pengkodean Laki-

laki = 1 dan Perempuan = 0, serta Hasil Pertandingan, dengan Menang = 1 dan Kalah = 0.

2. Sementara itu, *Ordinal Encoding* digunakan untuk atribut kategorikal yang memiliki urutan logis, seperti tingkat motivasi, kepercayaan diri, dan kemampuan mengontrol emosi, yang dinyatakan dalam skala 1 sampai 5. Nilai yang lebih tinggi menunjukkan tingkat yang lebih besar pada masing-masing atribut tersebut.

Atribut numerik seperti usia, tinggi badan, berat badan, jumlah kemenangan, jumlah kekalahan, serta durasi latihan tidak mengalami perubahan nilai, melainkan hanya dilakukan standarisasi format agar konsisten dalam satuan pengukuran. Variabel target (label *output*) dalam penelitian ini adalah Hasil Pertandingan, yang terdiri dari dua kelas, yaitu Menang (1) dan Kalah (0).

Tabel 3. 3 Contoh Hasil Encoding Data

Jenis Kelamin	Hasil Pertandingan	Motivasi	Kepercayaan Diri	Kontrol Emosi
Laki-laki (1)	Menang (1)	5	5	5
Perempuan (0)	Kalah (0)	4	4	4

3.4.3 Normalisasi Data

Beberapa Tahapan normalisasi dilakukan untuk menyesuaikan skala antarfitur numerik agar memiliki rentang nilai yang sebanding, sehingga tidak ada variabel yang mendominasi proses pembelajaran model. Metode normalisasi yang digunakan dalam penelitian ini adalah *Min–Max Scaling*, yang bertujuan untuk mentransformasikan nilai data ke dalam rentang 0 hingga 1 (Allorerung *et al*., 2024) *Min-Max Scalling* dituliskan sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

Di mana:
X = nilai asli data
X_{min} = nilai minimum pada kolom tersebut
X_{max} = nilai maksimum pada kolom tersebut
X' = hasil normalisasi dalam rentang 0-1

Proses ini membantu model membaca setiap fitur secara proporsional dan mempercepat proses selama pelatihan tanpa mengubah distribusi data asli.

Tabel 3. 4 Contoh Hasil Normalisasi Data

Usia	Tinggi Badan	Berat Badan	Jumlah Pertandingan	Total Medali
0.71	0.63	0.45	0.50	0.60
0.85	0.82	0.70	0.75	0.80
0.57	0.48	0.40	0.35	0.50

3.4.4 Pembagian Data (Split Data)

Setelah melewati rangkaian pembersihan, transformasi (encoding), serta penyeragaman skala (normalisasi), dataset tersebut berikutnya dipisahkan ke dalam dua kelompok utama, yakni data latih (training data) beserta data uji (testing data). Pemisahan himpunan data ini diselenggarakan dengan maksud untuk memberikan pembelajaran pada sistem sekaligus mengevaluasi kompetensi model ketika melakukan estimasi terhadap data baru yang belum pernah dipelajari sebelumnya.

Dalam riset ini, eksperimen dijalankan menggunakan tiga skema pemisahan dataset guna membedah dampak rasio perbandingan antara data latih dan data uji terhadap tingkat keberhasilan model. Skema pembagian yang diterapkan meliputi komposisi data latih sebanding dengan data uji masing-masing sebesar 70:30, dilanjutkan dengan perbandingan 80:20, serta rancangan rasio terakhir pada porsi 90:10.

Pemisahan dataset dieksekusi secara acak memanfaatkan fungsi train_test_split yang dikombinasikan dengan metode stratified sampling. Pendekatan ini diterapkan guna menjamin proporsi penyebaran label target, baik untuk kategori "menang" maupun "kalah", tetap proporsional dan selaras di dalam himpunan data latih maupun data uji. Selanjutnya, subset data latih dialokasikan untuk memfasilitasi fase pembelajaran sistem menggunakan algoritma Random Forest, sementara subset data uji difungsikan sebagai instrumen untuk menakar dan meninjau keandalan performa model pada tahapan pengujian akhir.

Tabel 3. 5 Rancangan Skenario Pembagian Data

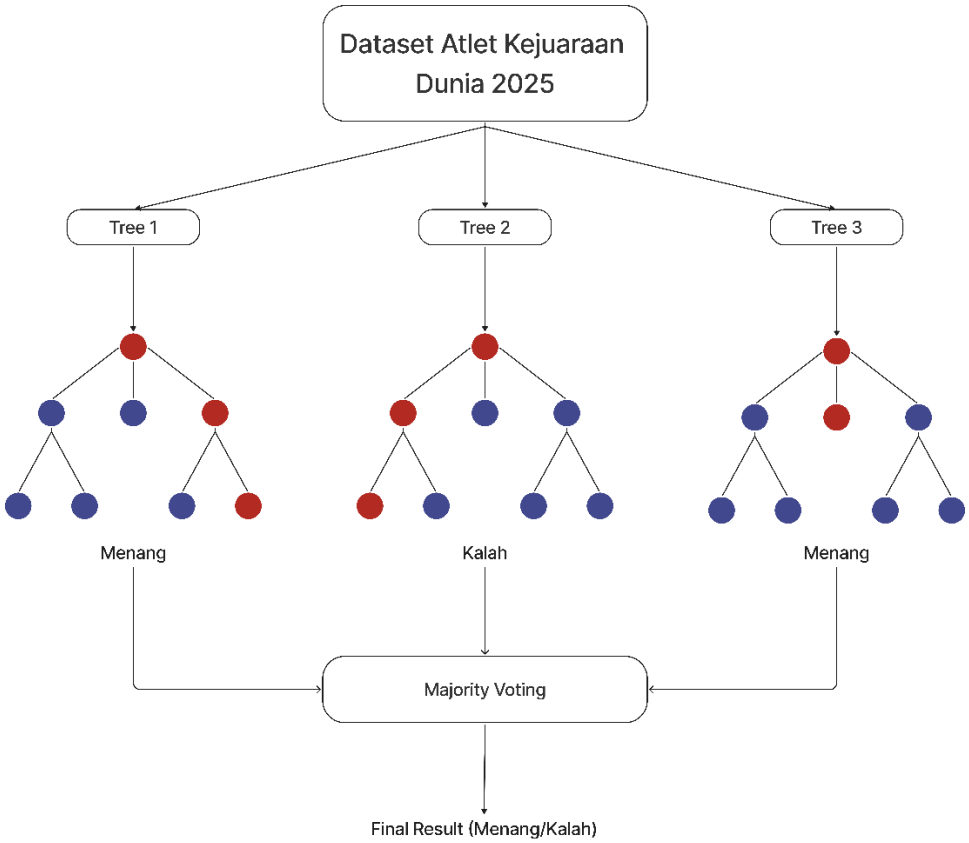
No	Skenario Split	Data Latih	Data Uji
1.	70:30	70%	30%
2.	80:20	80%	20%
3.	90:10	90%	10%

3.5 Algoritma *Random Forest*

Dalam riset ini, algoritma *Random Forest* diandalkan sebagai instrumen metodologi utama guna mengonstruksi model estimasi hasil kemenangan bagi pesilat Tapak Suci. Sebagai bagian dari rumpun *ensemble learning*, arsitektur ini bekerja melalui pengintegrasian luaran dari sekumpulan pohon keputusan (*Decision Tree*). Berdasarkan studi (Yu *et al.*, 2020), mekanisme penggabungan ini diterapkan demi mewujudkan hasil proyeksi akhir yang lebih konsisten, berpresisi tinggi, serta tangguh dalam mengantisipasi risiko *overfitting*.

Metode ini dipilih karena mampu mengolah data dengan tipe beragam (numerik maupun kategorikal), memiliki ketahanan terhadap noise, serta dapat mengidentifikasi fitur-fitur yang paling berpengaruh terhadap hasil prediksi (*feature importance,*). Dalam penelitian ini, model *Random Forest* digunakan

untuk memprediksi kemungkinan kemenangan atlet Tapak Suci berdasarkan faktor fisik, latihan, pengalaman, dan psikologis yang diperoleh dari hasil kuesioner atlet peserta Kejuaraan Dunia Tapak Suci 2025.



Gambar 3. 3 Alur Kerja Algoritma *Random Forest*

Gambar 3.3 menjelaskan alur kerja algoritma *Random Forest* yang digunakan dalam penelitian ini. Secara umum, proses pembentukan model terdiri dari beberapa tahapan sebagai berikut:

1. Dataset Atlet Tapak Suci

   Tahap awal dimulai dengan penggunaan dataset hasil kuesioner yang berisi berbagai atribut seperti usia, tinggi badan, berat badan, frekuensi latihan, durasi latihan, jumlah pertandingan, kemenangan, kekalahan, total medali, serta aspek

psikologis seperti motivasi, kepercayaan diri, dan kontrol emosi. Data ini berfungsi sebagai masukan utama dalam proses pembentukan model prediksi kemenangan atlet.

Dari hasil pengumpulan data, terdapat 15 fitur input (x) yang digunakan, dan 1 fitur output (y) yaitu hasil pertandingan (menang atau kalah). Fitur seperti nama atlet dan kontingen tidak disertakan karena tidak relavan terhadap hasil klasifikasi dan berpotensi menyebabkan bias dalam model.

2. *Boostrap* Sampling

   Tahapan berikutnya adalah boostrap sampling, yaitu proses pengambilan sample data secara acak dengan pengembalian. Dari dataset utama berisi 226 data atlet, setiap *Decision Tree* akan dilatih menggunakan subset data acak yang memiliki ukuran sama dengan dataset asli, namun isi datanya berbeda karena hasil pengacakan.

   Misalnya, satu pohon dapat memuat beberapa atlet yang muncul dua kali, sementara beberapa atlet lain tidak terpilih sama sekali. Proses ini menciptakan variasi antar pohon sehingga model menjadi lebih stabil dan mampu melakukan generalisasi terhadap data baru (Bunker & Susnjak, 2022).

3. *Random Feature Selection*

   Setelah subset data terbentuk, tahap selanjutnya adalah pemilihan fitur secara acak *(random feature selection).* Pada tahap ini, algoritma tidak menggunakan semua fitur sekaligus, melainkan hanya sebagian fitur yang dipilih secara acak di setiap *node* dalam *Decision Tree* .

Jumlah fitur acak yang dipertimbangkan pada setiap node dihitung menggunakan rumus (Pongthao *et al*., 2025):

$$(m = \sqrt{M}) \tag{3.2}$$

Keterangan:
M = jumlah total fitur input (x)
m = jumlah fitur acak yang dipilih pada setiap node.

Berdasarkan perhitungan tersebut, diperoleh $m = \sqrt{14} = 3{,}74$, sehingga pada setiap node pohon, algoritma mempertimbangkan sekitar 3 hingga 4 fitur acak sebagai kandidat pemisah terbaik.

Langkah ini membuat setiap pohon dalam forest memiliki kombinasi fitur yang berbeda, mengurangi kolerasi antar pohon, dan meningkatkan keragaman model secara keseluruhan.

4. Pembentukan *Decision Tree*

Setiap subset data hasil *bootstrap* sampling digunakan untuk membentuk satu *Decision Tree* dengan algoritma *Classification* and *Regression Tree* (CART). Proses pembentukan pohon diawali dari node awal *(root node),* yaitu simpul pertama tempat seluruh data atlet dimasukkan untuk dihitung tingkat ketidakmurniannya *(impurity).*

Ukuran ketidakmurnian ini dihitung menggunakan *Gini Index,* dengan rumus (Sindar *et al*., 2025):

$$Gini\ (t) = 1 - \sum_{i=1}^{n} p_i^2 \tag{3.3}$$

Keterangan:
p_i = Proporsi data dari kelas ke-i pada suatu node

Nilai Gini yang semakin kecil menunjukkan bahwa data di node tersebut semakin "murni" dalam satu kelas (Menang atau Kalah). Pemisahan *(split)* dilakukan secara rekursif, yaitu setiap node yang masih mengandung campuran kelas akan terus dibagi berdasarkan fitur acak yang memberikan nilai Gini terkecil.

Sebagai contoh sederhana, misalkan pada node awal terdapat 10 data atlet dengan 6 "Menang" dan 4 "Kalah", Maka:

$$Gini\ (t) = 1 - (p_m enang^2 + p_k alah^2) = 1 - (0.6^2 + 0.4^2)$$

$$= 1 - (0.36 + 0.16) = 0.48$$

Nilai 0,48 ini menunjukkan bahwa node masih belum murni (karena campur Menang dan Kalah), sehingga model akan mencari fitur terbaik, misalnya "Jumlah Kemenangan" atau "Motivasi" yang dapat memisahkan data menjadi dua kelompok dengan nilai Gini lebih kecil. Proses ini berlanjut hingga pohon mencapai batas kedalaman maksimum *(max depth)* atau semua node telah murni.

5. Prediksi Setiap Pohon

   Setelah seluruh *Decision Tree* terbentuk, setiap pohon memberikan hasil prediksi terhadap data uji berdasarkan pola yang telah dipelajarinya dari data pelatihan. Misalnya, untuk satu data atlet baru, *Tree* 1 memprediksi "Menang", *Tree* 2 memprediksi "Kalah", dan *Tree* 3 memprediksi "Menang". Setiap pohon dapat memberikan keputusan yang berbeda karena masing-masing dibangun dari subset data dan fitur yang berbeda.

6. *Majority Voting* dan Hasil Akhir

   Semua hasil prediksi dari setiap pohon kemudian digabungkan menggunakan mekanisme majority voting, yaitu pengambilan keputusan berdasarkan suara terbanyak. Kelas yang paling banyak dipilih akan menjadi hasil akhir model.

   Berdasarkan contoh sebelumnya, jika dua dari tiga pohon memprediksi "Menang", maka keputusan akhir model juga adalah "Menang". Mekanisme ini membuat hasil prediksi lebih akurat dan konsisten dibandingkan dengan model *Decision Tree* tunggal.

7. *Feature Importance*

   Selain menghasilkan prediksi, algoritma *Random Forest* juga menghitung tingkat kepentingan setiap fitur terhadap hasil klasifikasi. Nilai *feature importance,* menunjukkan seberapa besar kontribusi suatu fitur dalam mengurangi ketidakmurnian *(impurity reduction)* di seluruh pohon keputusan dalam *forest.*

   Dalam konteks penelitian ini, fitur seperti frekuensi latihan per minggu, jumlah kemenangan, dan tingkat motivasi diperkirakan memiliki pengaruh paling besar terhadap peluang kemenangan atlet Tapak Suci. Analisis *feature importance,* nantinya digunakan untuk membantu pelatih dalam memahami faktor apa yang paling menentukan keberhasilan atlet pada kejuaraan dunia.

3.6 Evaluasi Model

Pengujian terhadap arsitektur sistem dieksekusi lintas skema rasio pemisahan dataset, mencakup proporsi 70:30, 80:20, hingga 90:10. Langkah ini ditujukan demi mengamati ketahanan serta stabilitas capaian model saat

dihadapkan pada fluktuasi volume data latih maupun data uji. Peninjauan kapabilitas itu sendiri merupakan fase krusial di dalam rancang bangun model klasifikasi guna menakar ketepatan hasil estimasi sistem terhadap fakta riil di lapangan. Guna mengakomodasi kebutuhan tersebut, instrumen *Confusion matrix* diaplikasikan sebagai salah satu pendekatan validasi utama pada penelitian ini.

Confusion matrix dipahami sebagai sebuah matriks atau tabel dua dimensi yang difungsikan dalam menakar efektivitas kinerja model klasifikasi lewat komparasi antara nilai hasil estimasi sistem dan label riil *(ground truth)*. Pada permasalahan klasifikasi biner, skema ini mengadopsi struktur dasar berukuran 2×2 yang mengintegrasikan empat elemen pembentuk utama, meliputi:

1. *True Positive* (TP) adalah jumlah data positif yang berhasil diprediksi positif oleh model (misalnya atlet yang menang dan diprediksi menang).
2. *True Negative* (TN) adalah jumlah data negatif yang diprediksi negatif (atlet yang kalah dan diprediksi kalah).
3. *False Positive* (FP) adalah data negatif yang salah diprediksi positif (atlet kalah namun diprediksi menang).
4. *False Negative* (FN) adalah data positif yang salah diprediksi negatif (atlet menang namun diprediksi kalah).

Dari keempat komponen tersebut, dapat diturunkan berbagai metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi, antara lain *Accuracy, Precision, Recall,* dan *F1-score.*

1. *Accuracy*

   Akurasi mengukur tingkat keseluruhan kebenaran prediksi model dengan menghitung perbandingan antara jumlah prediksi benar dan jumlah seluruh data yang diuji (Sujon *et al*., 2025).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \; x \; 100\% \tag{3.4}$$

   Akurasi memberikan gambaran umum mengenai kemampuan model dalam mengklasifikasikan data secara tepat. Namun, metrik ini dapat kurang representatif apabila jumlah kelas dalam dataset tidak seimbang (*imbalanced data*).

2. *Precision*

   Presisi menunjukkan tingkat ketepatan model dalam memprediksi kelas positif, atau seberapa banyak data yang benar-benar positif dari seluruh data yang diprediksi positif (Sujon *et al*., 2025).

$$Precision = \frac{TP}{TP + FP} \, x \; 100\% \tag{3.5}$$

   Nilai presisi yang tinggi menunjukkan bahwa model jarang melakukan kesalahan *false positive*, sehingga hasil prediksi positifnya dapat dipercaya.

3. *Recall*

   Recall mengukur kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya. Nilai *Recall* yang tinggi menandakan bahwa model mampu meminimalkan kesalahan *false negative* (Sujon *et al*., 2025).

$$Recall = \frac{TP}{TP + FN} x\ 100\% \quad (3.6)$$

Dalam konteks penelitian ini, *Recall* penting untuk mengetahui sejauh mana model dapat mengenali seluruh atlet yang berpotensi menang.

4. *F1-Score*

   F1-score merupakan rata-rata harmonik antara *Precision* dan *Recall* yang memberikan keseimbangan antara keduanya (Sujon *et al*., 2025).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} x\ 100\% \quad (3.7)$$

   Nilai *F1-score* berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan keseimbangan yang baik antara ketepatan dan kelengkapan prediksi.

Tahap evaluasi ini digunakan untuk menilai performa algoritma *Random Forest* dalam memprediksi kemenangan atlet Pencak Silat Tapak Suci berdasarkan data performa dan hasil pertandingan. Hasil perhitungan metrik evaluasi tersebut menjadi dasar dalam menilai efektivitas model yang telah dibangun.

3.7 *K-Fold Cross Validation*

K-Fold Cross Validation didefinisikan sebagai suatu metode verifikasi model yang diterapkan demi menakar tingkat stabilitas serta konsistensi capaian sistem saat dieksekusi menggunakan variasi partisi dataset yang berbeda. Melalui pendekatan ini, peneliti dapat menjamin bahwa efisiensi algoritma bersifat objektif dan tidak sekadar bertumpu pada satu skenario pembagian data latih maupun data uji (*data splitting*).

Teknik ini bekerja dengan cara membagi dataset menjadi beberapa bagian (*fold*) dengan ukuran yang relatif sama. Pada setiap iterasi, model dilatih menggunakan k−1 *fold* dan diuji menggunakan satu *fold* yang tersisa. Proses ini dilakukan secara berulang hingga seluruh *fold* pernah digunakan sebagai data uji.

Kelebihan metode ini adalah mampu memberikan estimasi performa model yang lebih representatif karena seluruh data memiliki kesempatan untuk digunakan sebagai data latih maupun data uji. Selain itu, metode ini juga dapat mengurangi kemungkinan terjadinya bias akibat pemilihan data latih dan data uji tertentu.

Pada penelitian ini digunakan 10-*Fold Cross Validation* (k = 10) dengan pengacakan data (*shuffle = True*). Teknik validasi ini diterapkan pada model *Random Forest* untuk mengukur kestabilan performa model pada setiap skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Hasil validasi digunakan untuk mendukung interpretasi hasil evaluasi model yang diperoleh dari metrik *Accuracy, Precision, Recall, dan F1-score*.

3.8 Skenario Pengujian

Skenario Formulasi skenario eksperimen ini dirancang demi memetakan kapabilitas arsitektur Random Forest dalam mengestimasi hasil kemenangan pesilat Tapak Suci pada ajang Kejuaraan Dunia 2025. Agenda pengujian tersebut diselenggarakan guna memberikan kepastian bahwa sistem yang dirancang menembus tingkat presisi yang optimal serta andal dalam melacak variabel penentu yang memengaruhi luaran laga secara konsisten.

Fase pengujian dieksekusi secara terpisah sesudah tahapan pembelajaran (training) tuntas dikerjakan, dengan mengandalkan subset data uji yang disisihkan

dari basis dataset utama lewat proporsi rasio 80:20. Himpunan data latih dialokasikan sepenuhnya untuk mengonstruksi fondasi model, sementara data uji dimanfaatkan sebagai instrumen validasi dalam menilai ketepatan model kala memproyeksikan data baru yang belum pernah dipelajari sebelumnya.

Tahap pengujian dilakukan setelah proses pelatihan *(training)* selesai, dengan menggunakan data uji yang telah dipisahkan dari dataset utama dengan rasio 80:20. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk menilai kemampuan model dalam memprediksi data baru yang belum pernah dilihat sebelumnya.

a. Pengujian *Hyperparameter*

Pengujian Pengujian *hyperparameter* dilakukan sebagai bagian dari proses *hyperparameter tuning* untuk menemukan konfigurasi model *Random Forest* yang menghasilkan performa terbaik.

Pengujian dilakukan dengan memvariasikan beberapa parameter utama pada algoritma *Random Forest*. Parameter yang diuji ditunjukkan pada Tabel 3.6 berikut.

Tabel 3. 6 Konfigurasi *Hyperparameter* yang Diuji

No	Parameter	Nilai yang Diuji	Keterangan
1.	*N_estimators*	[100, 200, 300]	Jumlah pohon keputusan dalam forest
2.	*Max_depth*	[10, 20, *None*]	Kedalaman maksimum pohon keputusan

Setiap Proses pengujian *hyperparameter* dilakukan menggunakan metode *Grid Search Cross Validation*, di mana setiap kombinasi parameter diuji menggunakan data yang sama agar hasil evaluasi dapat dibandingkan secara objektif.

Model dengan nilai akurasi tertinggi dipilih sebagai model terbaik yang akan digunakan pada tahap analisis fitur selanjutnya.

b. Analisis Feature Importance

Tahap berikutnya adalah analisis *feature importance, ,* yaitu proses untuk mengetahui sejauh mana setiap fitur individu berkontribusi terhadap hasil prediksi. Nilai *feature importance,* dihitung berdasarkan seberapa besar pengaruh suatu fitur dalam mengurangi ketidakmurnian *(impurity reduction)* pada setiap pohon keputusan di dalam model *Random Forest*. Analisis ini dilakukan terhadap model terbaik yang diperoleh dari tahap pengujian *hyperparameter*. Hasil perhitungan *feature importance,* digunakan untuk mengidentifikasi fitur yang paling menentukan kemenangan atlet, seperti frekuensi latihan, jumlah kemenangan, atau tingkat motivasi.

Informasi ini diharapkan dapat menjadi acuan bagi pelatih Tapak Suci dalam menyusun strategi latihan dan pembinaan atlet berbasis data, dengan fokus pada faktor-faktor yang terbukti paling memengaruhi peluang kemenangan.

Melalui serangkaian pengujian ini, diperoleh gambaran menyeluruh mengenai kinerja algoritma *Random Forest* dalam memprediksi kemenangan atlet Tapak Suci. Model dengan konfigurasi dan kombinasi fitur terbaik akan digunakan sebagai hasil akhir penelitian serta dasar dalam analisis performa atlet berbasis data.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Sub bab ini menyajikan hasil penelitian yang diperoleh dari seluruh tahapan penelitian yang telah dilaksanakan sebagaimana dijelaskan pada Bab III. Penyajian hasil penelitian difokuskan pada kondisi data penelitian setelah dilakukan pengolahan awal, hasil penerapan *preprocessing* data, proses pelatihan model menggunakan algoritma *Random Forest*, hasil penentuan parameter terbaik melalui *tuning hyperparameter*, serta evaluasi performa model dalam memprediksi kemenangan atlet pencak silat Tapak Suci. Selain itu, pada subbab ini juga disajikan hasil analisis *feature importance* serta implementasi model ke dalam sistem prediksi berbasis web.

4.1.1 Hasil Pengolahan Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui survei terhadap atlet pencak silat Tapak Suci. Berdasarkan hasil pengolahan data awal, jumlah data yang digunakan dalam penelitian ini sebanyak 226 data atlet. Setiap data mempresentasikan karakteristik atlet yang mencangkup aspek fisik pengalaman bertanding, intensitas latihan, serta aspek psikologis yang berkaitan dengan performa atlet dalam pertandingan.

Sebagai gambaran umum, sebagian data penelitian ditampilkan pada Tabel 4.1

Tabel 4. 1 Dataset Atlet Kejuaraan Dunia Tapak Suci 2025

No	Nama Atlet	A Muh Imam S	Arief Saputra	Julianastasia H
1	Jenis Kelamin	Laki-laki	Laki-laki	Perempuan

No	Nama Atlet	A Muh Imam S	Arief Saputra	Julianastasia H
2	Usia	27	28	28
3	Tinggi Badan	172 cm	166 cm	160 cm
4	Berat Badan	82 kg	64 kg	58 kg
5	Kontingen	Sulsel	Lampung	Jawa Timur
6	Kelas Tanding	Kelas C	Kelas G	Kelas C
7	Jumlah Pertandingan	Lebih dari 20 pertandingan	11-20 pertandingan	11-20 pertandingan
8	Jumlah Kemenangan	Lebih dari 7 kali kemenangan	4-7 kali menang	Lebih dari 7 kali menang
9	Jumlah Kekalahan	4-7 kali kalah	1-3 kali kalah	1-3 kali kalah
10	Total Medali	1-3 medali	4-7 medali	4-7 medali
11	Latihan dalam sehari	2 kali	2 kali	Lebih dari 2 kali
12	Latihan dalam seminggu	3-4 kali	5-7 kali	5-7 kali
13	Lama latihan dalam satu hari	Lebih dari 60 menit	Lebih dari 60 menit	Lebih dari 60 menit
14	Berapa lama aktif bertanding	Sebelum 2020	Sebelum 2020	2020-2025
15	Kondisi fisik saat ini	Sedang	Baik	Baik
16	Gaya bertanding	Campuran	Menyerang	Menyerang
17	Tingkat percaya diri	4	5	3
18	Seberapa besar motivasi	5	5	4
19	Kontrol emosi	2	3	4
20	Hasil pertandingan	Kalah	Menang	Menang

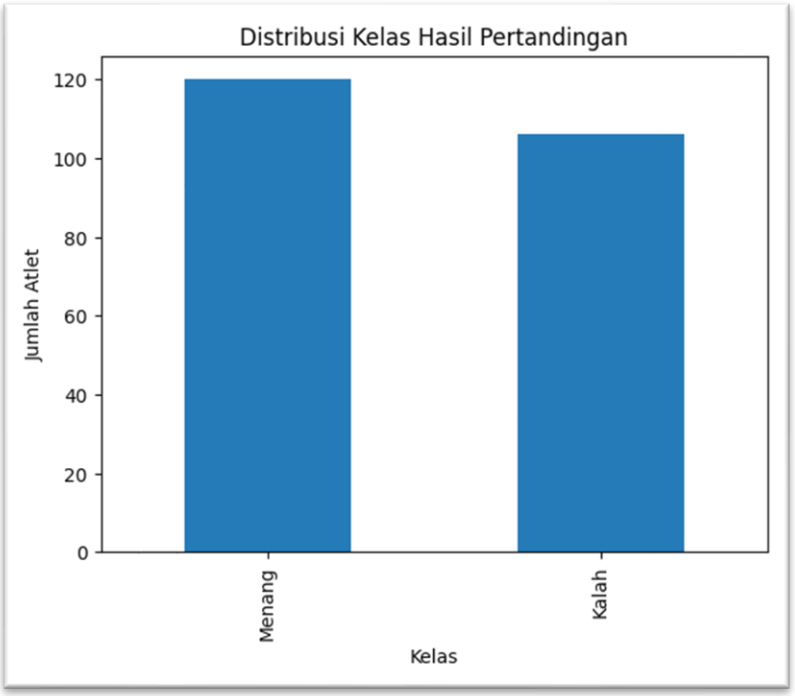
Alur penelitian secara umum dibagi menjadi tiga tahapan utama, yaitu tahap masukkan (*input*), tahap pemrosesan (*process*), dan tahap luaran (*output*). Pada tahap *input*, penelitian ini menggunakan *Athlete Performance Dataset* yang memuat rekam jejak klinis, latihan, dan psikologis atlet. Selanjutnya, data tersebut masuk ke tahap *process* yang diawali dengan prapemrosesan data (*preprocessing*), meliputi pembersihan data (*cleaning*), pengodean label (*encoding*), dan normalisasi nilai. Setelah data siap, dilakukan pemodelan menggunakan metode *Random Forest Classification* yang kemudian diuji mutunya pada sub-tahap evaluasi menggunakan metrik akurasi (*accuracy*), presisi (*precision*), *recall*, dan *f1-score*. Terakhir, pada tahap *output*, model yang telah optimal diekstrak untuk menghasilkan performa kedudukan fitur (*feature importance performance*) sebagai basis rekomendasi evaluasi atlet.

Sebagai gambaran umum mengenai instrumen data yang digunakan untuk melatih algoritma, seluruh variabel prediktor dikelompokkan secara terstruktur. Atribut prediktor tersebut mencakup 11 variabel utama yang merepresentasikan karakteristik fisik, aktivitas latihan, pengalaman kompetisi, dan profil psikologis atlet. Rincian data serta deskripsi operasional dari masing-masing fitur disajikan secara komprehensif pada Tabel 4.2 berikut di bawah ini.

Tabel 4. 2 Distribusi Data Penelitian

No	**Hasil Pertandingan**	**Jumlah Data**	**Persentase (%)**
1	Menang	120	53,10
2	Kalah	106	46,90
Total		226	100

Variabel target pada penelitian ini adalah hasil pertandingan, yang diklasifikasikan ke dalam dua kelas, yaitu menang dan kalah. Hasil pengolahan data menunjukkan bahwa dari keseluruhan data yang digunakan, terdapat 120 data atlet yang termasuk dalam kategori menang dan 106 data atlet yang termasuk dalam kategori kalah. Komposisi data tersebut menunjukkan bahwa distribusi kelas berada dalam kondisi relatif seimbang sehingga tidak menimbulkan permasalahan ketimpangan kelas pada proses klasifikasi.



Gambar 4. 1 Distribusi Kelas Hasil Pertandingan

Hasil pemeriksaan terhadap kelengkapan data menunjukkan bahwa seluruh atribut pada dataset tidak memiliki nilai kosong *(missing value)*. Dengan demikian, data tidak memerlukan proses penangan data hilang dan dapat langsung digunakan pada tahapan selanjutnya tanpa dilakukan imputasi tambahan.

Secara umum, dataset penelitian terdiri dari kombinasi atribut numerik dan atribut kategorikal yang menggambarkan kondisi atlet pencak silat secara komprehensif, baik dari sisi fisik, pengalaman bertanding, maupun kesiapan mental. Berdasarkan hasil pengolahan data awal tersebut, dataset dinyatakan layak dan siap digunakan untuk tahap *preprocessing* dan pemodelan menggunakan agoritma *Random Forest*.

4.1.2 Hasil Prepocessing Data

Tahap prepocessing data dilakukan untuk memastikan data penelitian berada dalam kondisi yang sesuai sebelum digunakan dalam proses pelatihan model. Tahapan prepocessing yang dilakukan meliputi data *cleaning, encoding,* dan normlasisasi data.

1. *Data Cleaning*

   Pada tahap awal dilakukan pemeriksaan terhadap kelengkapan data untuk mengidentifikasi adanya nilai kosong *(missing value)* pada dataset.

   Hasil pemeriksaan menunjukkan bahwa seluruh atribut dalam dataset tidak memiliki nilai kosong. Dengan demikian, data dinyatakan bersih dan tidak memerlukan proses penanganan *missing value.*

   Selain itu, dilakukan penyesuaian format data pada beberapa atribut yang masih mengandung satuan, seperti tinggu badan(cm) dan berat badan(kg), sehingga dapat dikonversi menjadi nilai numerik.

Tabel 4. 3 Hasil Data *Cleaning*

No	Fitur	Missing Count
1.	Nama lengkap atlet	0
2.	Jenis kelamin	0
3.	Usia saat ini	0
4.	Tinggi badan	0
5.	Berat badan	0
6.	Kontingen	0
7.	Kategori tanding	0
8.	Berapa jumlah pertandingan resmi yang pernah anda ikuti sejauh ini?	0
9.	Dari seluruh pertandingan yang anda ikuti, berapa jumlah kemenangan yang telah anda raih?	0
10.	Dari seluruh pertandingan yang anda ikuti, berapa jumlah kekalahan yang pernah anda alami?	0
11.	Total keseluruhan medali (emas, perak, perunggu) yang pernah anda dapatkan selama mengikuti kejuaraan?	0
12	Berapa kali anda berlatih dalam satu hari selama persiapan kejuaraan	0
13.	Berapa kali anda berlatih dalam seminggu?	0
14.	Berapa lama rata-rata durasi latihan anda dalam satu sesi?	0

No	Fitur	Missing Count
15.	Sejak tahun berapa anda mulai aktif bertanding di kejuaraan pencak silat?	0
16.	Bagaimana kondisi fisik anda saat mengikuti kejuaraan ini?	0
17.	Gaya bertanding seperti apa yang paling menggambarkan anda?	0
18.	Seberapa tinggi tingkat kepercayaan diri anda saat bertanding?	0
19.	Seberapa besar motivasi anda untuk menang dalam pertandingan ini?	0
20.	Sejauh mana anda mampu mengontrol emosi selama pertandingan berlangsung?	0
21.	Hasil pertandingan	0

2. *Encoding* Data

Tahap *encoding* dilakukan untuk mengubah data kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma *Random Forest*.

a. *Encoding* variabel target

Variabel target yaitu hasil pertandingan dikonversi menjadi bentuk numerik, dengan ketentuan sebagai berikut:

Tabel 4. 4 Hasil Encoding Variabel Target

No	Kategori	Nilai
1.	Menang	1
2.	Kalah	0

b. *Encoding* variabel ordinal

Variabel yang memiliki tingkatan, seperti kepercayaan diri, motivasi, dan kontrol emosi, dikonversi ke dalam skala numerik bertingkat.

Tabel 4. 5 Hasil Encoding Variabel Ordinal

No	Kategori	Nilai
1.	Sangat rendah	1
2.	Rendah	2
3.	Sedang	3
4.	Tinggi	4
5.	Sangat tinggi	5

c. *Encoding* variabel ordinal

Variabel jenis kelamin dikonversi ke dalam bentuk numerik menggunakan label *encoding*.

Tabel 4. 6 Hasil Encoding Variabel Kategorikal

No	Kategori	Nilai
1.	Laki - laki	1
2.	Perempuan	0

d. Hasil konversi tipe data

Setelah proses *encoding*, seluruh atribut berhasil dikonversi ke dalam bentuk numerik.

Tabel 4. 7 Hasil Encoding Variabel Kategorikal

No	Fitur	Tipe Data
1.	Jenis kelamin	Integer
2.	Usia saat ini	Integer
3.	Tinggi badan	Float
4.	Berat badan	Float
5.	Jumlah pertandingan	Float
6.	Jumlah kemenangan	Float
7.	Jumlah kekalahan	Float
8.	Total medali	Float
9.	Latihan per hadi	Float
10.	Latihan per minggu	Float
11.	Durasi latihan	Float
12.	Tahun mulai bertanding	Float
13.	Kepercayaan diri	Integer
14.	Motivasi	Integer
15.	Kontrol emosi	Integer

3. Normalisasi Data

Tahap selanjutnya adalah normalisasi data menggunakan metode *Min-Max Scaling*. Proses ini dilakukan untuk menyamakan rentang nilai seluruh fitur ke dalam skala 0 hingga 1.

Berikut ini contoh perubahan nilai sebelum dan sesudah normalisasi:

Tabel 4. 8 Hasil *Encoding* Variabel Kategorikal

Fitur	Sebelum	Sesudah
Usia	27	0.58
Tinggi badan	172	0.67
Berat badan	82	0.43
Jumlah pertandingan	15	0.44
Latihan per minggu	6	0.85

Merujuk pada keseluruhan rangkaian pra-pemrosesan (*preprocessing*) yang telah dieksekusi, luaran dataset final menyajikan total rekaman data dari 226 pesilat yang mengintegrasikan 15 atribut fitur sebagai variabel input pada arsitektur sistem. Transformasi data ke dalam format angka (numerik) telah tuntas dikerjakan secara menyeluruh guna menjamin kompatibilitas pemrosesan oleh algoritma *Random Forest*. Di samping itu, tahapan pengecekan mengonfirmasi absensi nilai kosong (*missing value*) pada dataset pasca-pemrosesan selesai dilakukan. Penyeragaman skala melalui pendekatan *Min-Max Scaling* juga telah diterapkan, yang berimplikasi pada standarisasi sebaran nilai fitur ke dalam batasan interval antara 0 sampai 1. Oleh karena itu, himpunan data ini dikategorikan lolos kualifikasi kriteria dan berada dalam status siap pakai untuk masuk ke fase pemisahan data (*data splitting*) serta tahapan pembelajaran model.

4.1.3 Hasil Pembagian Data (Split Data)

Riset ini menerapkan tiga skema pemisahan dataset, yang membagi porsi data latih dan data uji ke dalam rasio perbandingan 70:30, 80:20, serta 90:10. Implementasi dari ketiga skenario partisi tersebut ditujukan untuk menguji tingkat konsistensi maupun stabilitas capaian model *Random Forest* saat dihadapkan pada fluktuasi perbandingan volume data. Lewat strategi tersebut, peneliti dapat

membedah secara mendalam apakah pergeseran kuantitas data latih berimplikasi langsung terhadap raihan tingkat akurasi serta parameter evaluasi lainnya.

Prosedur pemisahan dataset itu sendiri dijalankan lewat metode *stratified sampling*, sebuah teknik partisi data yang didesain untuk mempertahankan rasio kelas target (kategori menang dan kalah) di setiap subset. Pendekatan tersebut diintegrasikan guna menjamin penyebaran label atau kelas tetap selaras dan seimbang, baik pada himpunan data latih maupun data uji.

Proses pembagian data dilakukan menggunakan metode *stratified sampling*, yaitu teknik pembagian data yang mempertahankan proporsi kelas target (menang dan kalah) pada setiap subset data. Pendekatan ini digunakan untuk memastikan distribusi kelas tetap seimbang pada data latih maupun data uji.

Hasil pembagian data pada masing-masing scenario ditampilkan pada tabel

Tabel 4. 9 Hasil pembagian data pada tiga skenario split

No	Skenario Split	Data Latih	Data Uji	Jumlah Total
1	70:30	158	68	226
2	80:20	180	46	226
3	90:10	203	23	226

Merujuk pada paparan Tabel 4.9, tampak bahwa peningkatan porsi data latih berbanding terbalik dengan kuantitas data uji yang tersedia. Pada skema 70:30, diperoleh sebaran berupa 158 rekaman data latih serta 68 data uji; skenario 80:20 mengondisikan 180 data latih dan 46 data uji; sedangkan pembagian 90:10 menyisakan 203 data latih dengan 23 data uji. Ketimpangan rasio ini berpeluang memengaruhi kedalaman sistem dalam menangkap karakteristik pola sekaligus mutu dari hasil pengujian yang didapatkan. Ketika volume data latih mendominasi, sistem memperoleh kesempatan yang lebih luas untuk mengenali sifat data secara

mendalam. Walau demikian, minimnya kuantitas data uji berisiko memicu bias, sehingga luaran penilaian menjadi kurang representatif kala menguji kompetensi model terhadap data baru. Di sisi lain, porsi data uji yang melimpah memang mampu menyajikan hasil pengujian yang lebih menyeluruh, namun terbatasnya pasokan data latih berpotensi menghambat optimalisasi fase pembelajaran sistem. Oleh sebab itu, ketiga rancangan pembagian ini diuji demi mengomparasikan kapabilitas model *Random Forest* sekaligus memetakan rasio dataset paling ideal yang mampu menyajikan titik keseimbangan terbaik antara efisiensi fase training dan fase evaluasi.

4.1.4 Hasil Pelatihan Model *Random Forest*

Proses ini berfokus pada *training* model prediktif menggunakan algoritma *Random Forest* dengan memanfaatkan data yang telah melewati fase pra-pemrosesan. Proses tersebut dirancang untuk menghasilkan sistem klasifikasi yang mampu memperkirakan hasil akhir pertandingan pesilat—baik itu menang maupun kalah berdasarkan karakteristik unik dari setiap atlet.

Tahap *training* dijalankan dengan memanfaatkan sekumpulan data latih yang bersumber dari pembagian basis data sebelumnya. Riset ini menerapkan tiga macam opsi rasio distribusi data, yakni 70:30, 80:20, serta 90:10. Kendati demikian, penalaan hiperparameter (*hyperparameter tuning*) beserta pembelajaran model di awal berfokus pada skema utama 80:20, yang dijadikan acuan untuk menetapkan setelan parameter paling optimal.

Algoritma *Random Forest* bekerja dengan membangun sejumlah pohon keputusan *(decision tree)* secara independen melalui teknik *bootstrap sampling*,

yaitu pengambilan sampel data latih secara acak dengan pengembalian. Selain itu, pada setiap percabangan pohon dilakukan pemilihan fitur secara acak *(random feature selection)* untuk menentukan pemisahan terbaik.

Setiap pohon keputusan menghasilkan prediksi kelas, kemudian hasil akhir model ditentukan berdasarkan mekanisme *majority voting*, yaitu kelas yang paling banyak diprediksi oleh seluruh pohon akan menjadi hasil prediksi akhir. Pendekatan ini bertujuan untuk meningkatkan stabilitas model serta mengurangi risiko *overfitting* dibandingkan penggunaan satu pohon keputusan tunggal.

Model *Random Forest* yang telah dilatih pada tahap ini selanjutnya digunakan pada proses *tuning hyperparameter* untuk memperoleh konfigurasi parameter terbaik, sebelum dilakukan evaluasi performa menggunakan data uji.

4.1.5 Hasil Pelatihan Model *Random Forest*

Optimasi hiperparameter dieksekusi demi mendapatkan pengaturan arsitektur *Random Forest* yang mampu menyajikan kinerja paling optimal. Langkah penalaan ini krusial lantaran efektivitas algoritma tersebut sangat bergantung pada kombinasi nilai parameter yang diterapkan, terutama kedalaman maksimal pohon (*max_depth*) serta kuantitas pohon keputusan (*n_estimators*) yang digunakan.

Dalam studi ini, skema pemisahan data dengan rasio 80:20 dipilih sebagai acuan utama untuk proses penyetelan tersebut. Penentuan proporsi ini didasarkan pada pertimbangan bahwa pembagian tersebut mampu menyeimbangkan ketersediaan data training yang berlimpah guna melatih algoritma, sekaligus

menyediakan porsi data testing yang proporsional untuk penilaian performa di tahap awal.

Pengujian dilakukan terhadap tiga variasi nilai *n_estimators*, yaitu 100, 200, dan 300, serta tiga variasi nilai *max_depth*, yaitu 10, 20, dan None. Dengan demikian, jumlah kombinasi *hyperparameter* yang diuji adalah:

$$3\ x\ 3\ =\ 9\ kombinasi$$

Seluruh kombinasi tersebut diuji menggunakan metode *Grid Search Cross Validation* dengan skema *10-Fold Cross Validation*. Pada setiap kombinasi parameter, data latih dibagi menjadi sepuluh bagian, kemudian model dilatih dan divalidasi secara bergantian sehingga diperoleh nilai rata-rata akurasi *(mean test score)* untuk masing-masing kombinasi. Hasil pengujian kombinasi *hyperparameter* ditampilkan pada Tabel 4.10.

Tabel 4. 10 Hasil *Tuning hyperparameter Random Forest*

No	***n_estimator***	***max_depth***	**Akurasi (CV)**
1	100	10	88,33%
2	300	10	88,33%
3	300	20	88,33%
4	300	None	88,33%
5	100	20	87,78%
6	200	20	87,78%
7	100	None	87,78%
8	200	10	87,78%
9	200	None	87,78%

Berdasarkan Tabel 4.10, terdapat empat kombinasi parameter yang menghasilkan nilai akurasi rata-rata tertinggi sebesar 0,8833 (88,33%), yaitu (*n_estimators* = 100, *max_depth* = 10), (*n_estimators* = 300, *max_depth* = 10), (*n_estimators* = 300, *max_depth* = 20), dan (*n_estimators* = 300, *max_depth* = None).

Di antara kombinasi tersebut, konfigurasi dengan *n_estimators* = 100 dan *max_depth* = 10 dipilih sebagai model terbaik karena memiliki tingkat kompleksitas yang lebih rendah dibandingkan konfigurasi dengan jumlah pohon yang lebih besar, sehingga lebih efisien secara komputasi dan berpotensi lebih stabil dalam proses generalisasi model.

4.1.6 Hasil Pengolahan Data Penelitian

Pengujian model dilakukan demi mengukur tingkat akurasi algoritma *Random Forest* saat memprediksi hasil tanding atlet pencak silat. Proses validasi ini memanfaatkan sekumpulan data uji (*testing data*) yang sepenuhnya terpisah dari tahap pelatihan. Langkah ini krusial agar hasil penilaian dapat merepresentasikan secara objektif bagaimana kemampuan model dalam menggeneralisasi data-data baru yang belum pernah dikenali sebelumnya.

Berbeda dari riset terdahulu yang umumnya terpaku pada satu skema pembagian data saja, penelitian kali ini menerapkan evaluasi berbasis tiga skenario pemisahan data yang berbeda. Ketiga skema tersebut meliputi perbandingan data latih dan data uji sebesar 70:30, 80:20, serta 90:10. Penerapan variasi ini ditujukan untuk meninjau tingkat stabilitas maupun konsistensi performa yang ditunjukkan model sewaktu proporsi datanya diubah-ubah. Adapun untuk mengukur efektivitasnya secara menyeluruh, penilaian akhir bersandar pada empat parameter evaluasi utama, yakni *accuracy*, *precision*, *recall*, beserta *F1-score*.

Tabel 4. 11 Hasil Evaluasi Model *Random Forest*

No	Skenario	Split	*Accuracy*	*Precision*	*Recall*	F-1 Score
1	Skenario 1	70:30	95,59%	95,93%	95,59%	95,57%
2	Skenario 2	80:20	95,65%	95,99%	96,56%	95,64%
3	Skenario 3	90:10	95,65%	96,01%	95,65%	95,65%

Berdasarkan Tabel 4.11, dapat diketahui bahwa seluruh skenario pembagian data menghasilkan performa yang sangat baik dan relatif stabil. Nilai *accuracy* pada ketiga skenario berada di atas 95%, menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data uji dengan benar.

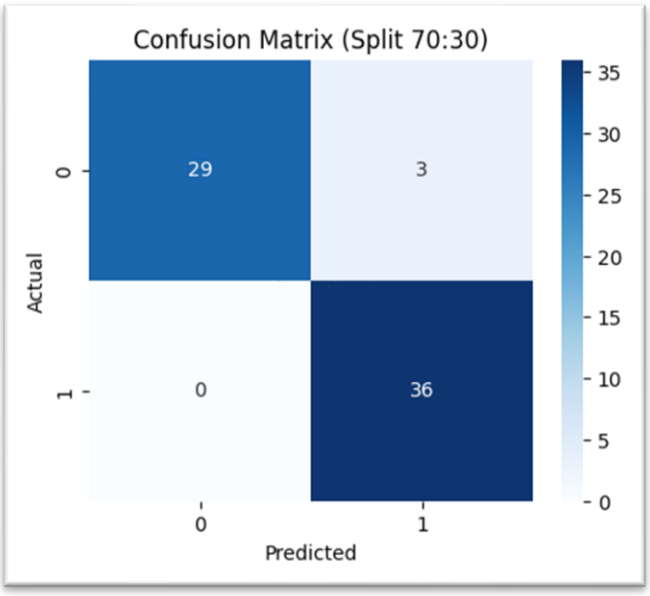
Tingkat akurasi paling tinggi diperoleh pada eksperimen dengan proporsi 80:20 serta 90:10, yakni mencapai 95,65%. Sementara itu, skema 70:30 menunjukkan hasil yang sedikit di bawahnya dengan angka 95,59%. Tipisnya selisih angka ini mengindikasikan bahwa model prediktif yang dibangun mempunyai stabilitas performa yang sangat kokoh meskipun diuji menggunakan variasi pembagian data yang berbeda-beda.

Ditinjau dari parameter *precision*, rasio data 90:10 mengungguli skenario lainnya dengan perolehan nilai maksimal sebesar 0,9601. Kondisi tersebut menandakan bahwa rasio ketepatan model dalam memprediksi kelas positif secara benar pada skema ini sedikit lebih unggul.

Untuk metrik *recall*, ketiga variasi eksperimen menyajikan skor yang cenderung seragam di kisaran 0,956. Hal ini membuktikan efektivitas model dalam mengidentifikasi mayoritas data pada kategori positif dengan sangat baik di semua skema distribusi data. Selaras dengan pencapaian tersebut, nilai *F1-score* yang mencerminkan keseimbangan performa antara *precision* dan *recall* turut memperlihatkan hasil yang ajeg di setiap pengujian.

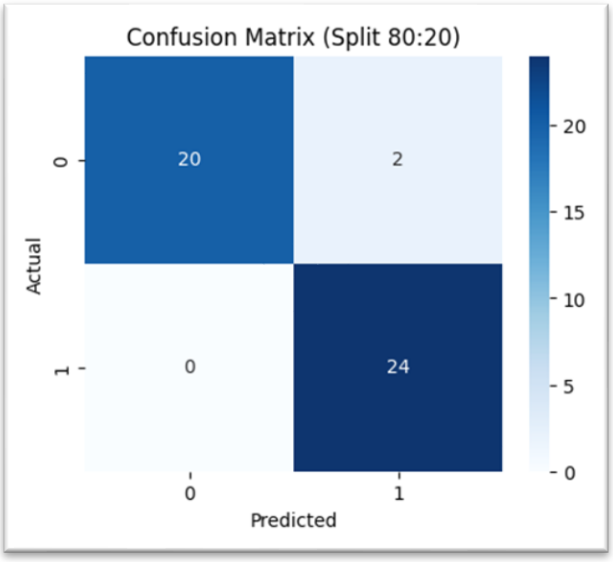
Guna menyajikan visualisasi yang lebih mendalam mengenai pola kesalahan dalam proses klasifikasi, pengujian pada tiap skenario dilengkapi dengan analisis *Confusion Matrix*.

Pada skema distribusi data 70:30, algoritma mencatatkan persentase akurasi di angka 95,59%. Merujuk pada hasil pemetaan *Confusion Matrix*, mayoritas sampel data mampu dikelompokkan ke dalam kategori yang tepat oleh sistem. Tercatat ada 29 data dari label negatif (kalah) yang berhasil diperkirakan secara akurat, serta 36 data pada kategori positif (menang) yang juga sukses terprediksi secara tepat. Adapun kekeliruan prediksi hanya dijumpai sebanyak 3 data pada label negatif, sedangkan untuk label positif sama sekali tidak ditemukan adanya kesalahan klasifikasi.



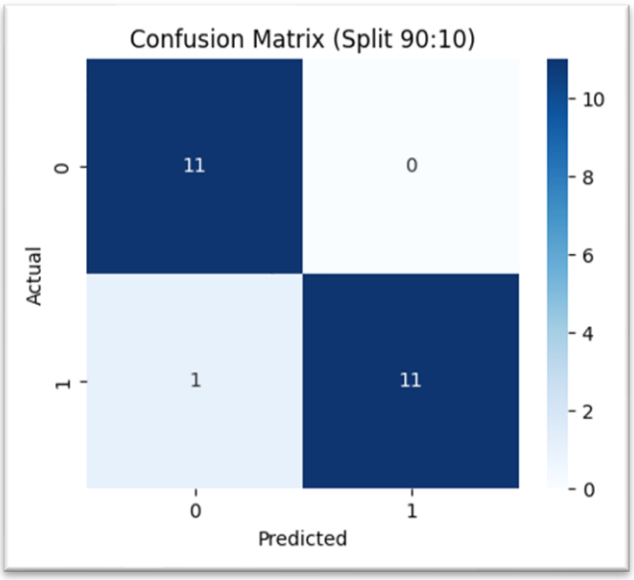
Gambar 4. 2 *Confusion matrix* Hasil Evaluasi (Split 70:30)

Pada skenario 80:20, model menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar 95,65%. *Confusion matrix* menunjukkan bahwa 20 data kelas negatif dan 24 data kelas positif berhasil diklasifikasikan dengan benar. Kesalahan prediksi hanya terjadi pada 2 data kelas negatif, sedangkan seluruh data kelas positif berhasil diprediksi dengan benar.



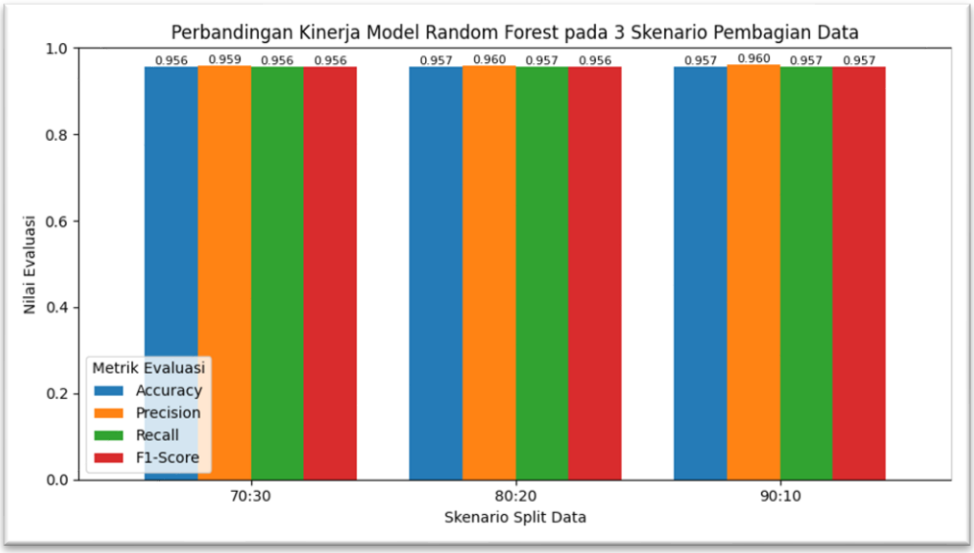
Gambar 4. 3 *Confusion matrix* Hasil Evaluasi (Split 80:20)

Pada skenario 90:10, model menghasilkan akurasi sebesar 95,65%. *Confusion matrix* menunjukkan bahwa 11 data kelas negatif dan 11 data kelas positif berhasil diklasifikasikan dengan benar. Terdapat 1 kesalahan klasifikasi pada kelas positif. Karena jumlah data uji lebih sedikit, variasi kesalahan menjadi lebih sensitif terhadap perubahan kecil pada prediksi.



Gambar 4. 4 *Confusion matrix* Hasil Evaluasi (Split 90:10)

Untuk memperjelas perbandingan performa model pada ketiga skenario pembagian data, ditampilkan visualisasi dalam bentuk grafik batang.



Gambar 4. 5 Perbandingan Kinerja Tiga Model Skenario Pengujian

Merujuk pada Gambar 4.5, perolehan skor untuk parameter *accuracy*, *precision*, *recall*, beserta *F1-score* di seluruh skenario pengujian hanya memperlihatkan selisih yang sangat tipis. Fenomena ini membuktikan bahwa arsitektur *Random Forest* yang dirancang mempunyai tingkat stabilitas yang amat kokoh terhadap perubahan rasio antara data training dan data testing.

Secara umum, capaian evaluasi ini menegaskan bahwa algoritma *Random Forest* mampu menyajikan performa yang mumpuni serta ajeg di tiga skema pembagian data tersebut. Berkat minimnya deviasi nilai pada metrik acuan (*accuracy*, *precision*, *recall*, dan *F1-score*), model ini dinilai mempunyai kapabilitas generalisasi yang sangat andal. Bersandarkan pada temuan tersebut, setelan parameter paling optimal yang didapatkan dari tahapan *tuning hyperparameter* terbukti cukup *robust* sehingga sangat representatif untuk

diaplikasikan pada fase analisis lanjutan maupun implementasi sistem prediksi ke depan.

4.1.7 Hasil *K-Fold Cross Validation*

Selain evaluasi menggunakan metode *hold-out* test, penelitian ini juga melakukan evaluasi menggunakan metode *10-Fold Cross Validation* (k = 10) pada seluruh skenario pembagian data. Penggunaan nilai k yang sama pada setiap skenario bertujuan untuk menjaga konsistensi prosedur evaluasi sehingga hasil yang diperoleh dapat dibandingkan secara lebih objektif.

Dalam implementasi teknik ini, data training pada tiap-tiap skenario dipecah menjadi sepuluh segmen (*fold*) dengan porsi yang relatif seimbang. Proses pembelajaran algoritma mengombinasikan sembilan *fold* sebagai basis latihan, sementara satu *fold* sisanya dialokasikan sebagai bahan pengujian yang diputar secara bergantian hingga seluruh bagian sempat berperan sebagai data uji. Mekanisme tersebut memicu lahirnya sepuluh skor akurasi pada tiap skema, yang nantinya dikalkulasikan guna memperoleh nilai rerata (*mean*) beserta standar deviasinya.

Penerapan metode *10-Fold Cross Validation* ini ditujukan demi mendapatkan estimasi kinerja model yang jauh lebih konsisten. Di samping itu, pendekatan ini efektif dalam meminimalisasi potensi bias akibat ketergantungan pada satu kali skema pemisahan data saja.

Hasil pengujian *10-Fold Cross Validation* pada masing-masing skenario ditampilkan pada Tabel 4.5.

Tabel 4. 12 Hasil Pengujian 10 *Fold* Cross Validation

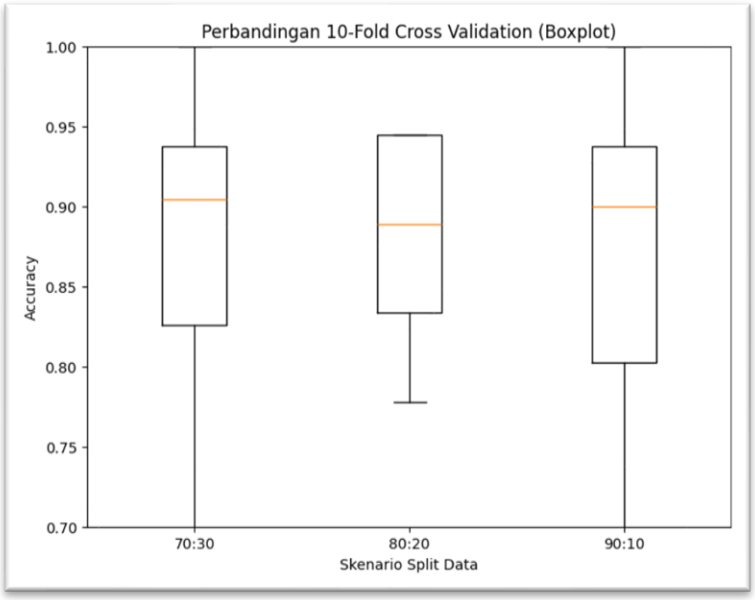
Split	k	Mean CV *Accuracy*	Std Dev
70:30	10	0,8675	0,1022
80:20	10	0,8833	0,0580
90:10	10	0,8676	0,1035

Berdasarkan tabel tersebut, dapat diketahui bahwa:

1. Skenario 80:20 memiliki nilai *mean cross-validation accuracy* tertinggi sebesar 0,8833.
2. Skenario 90:10 memiliki nilai *mean cross validation accuracy* sebesar 0,8676.
3. Skenario 80:20 memiliki nilai *mean cross validation accuracy* sebesar 0,8675.

Selain itu, skenario 80:20 memiliki nilai standar deviasi terkecil, yaitu 0,0580. Hal ini menunjukkan bahwa performa model pada skenario tersebut lebih stabil dan konsisten dibandingkan skenario lainnya. Sebaliknya, skenario 70:30 dan 90:10 memiliki nilai standar deviasi yang lebih besar (di atas 0,10), yang mengindikasikan adanya variasi performa antar *fold* yang lebih tinggi. Dengan demikian, skenario 80:20 tidak hanya menghasilkan nilai mean cross-validation accuracy tertinggi, tetapi juga menunjukkan tingkat kestabilan model yang lebih baik dalam proses validasi.

Untuk memperjelas distribusi nilai akurasi pada masing-masing skenario, ditampilkan visualisasi dalam bentuk *boxplot* pada Gambar 4.6.



Gambar 4. 6 Perbandingan Mean *Accuracy K-Fold Cross Validation*

Berdasarkan Gambar 4.6, terlihat bahwa median akurasi pada skenario 80:20 berada sedikit lebih tinggi dibandingkan skenario lainnya serta memiliki rentang distribusi yang lebih sempit. Hal ini menunjukkan bahwa skenario 80:20 tidak hanya memberikan rata-rata akurasi tertinggi, tetapi juga tingkat kestabilan performa yang lebih baik.

Secara keseluruhan, hasil *10-Fold Cross Validation* menunjukkan bahwa model *Random Forest* memiliki performa yang cukup konsisten pada ketiga variasi pembagian data. Temuan ini memperkuat hasil evaluasi *hold-out* sebelumnya bahwa model yang dibangun memiliki kemampuan generalisasi yang baik dan tidak bergantung secara signifikan pada satu konfigurasi pembagian data tertentu.

4.2 Pembahasan

4.2.1 Analisis Pengaruh *Hyperparameter*

Pada tahap *tuning hyperparameter*, dilakukan pengujian terhadap dua parameter utama dalam algoritma *Random Forest*, yaitu jumlah pohon keputusan (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*). Kedua parameter ini berpengaruh langsung terhadap kompleksitas model serta kemampuan model dalam mempelajari pola data.

Pada tahap *tuning hyperparameter*, dilakukan pengujian terhadap dua parameter utama dalam algoritma *Random Forest*, yaitu jumlah pohon keputusan (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*). Kedua parameter ini berpengaruh langsung terhadap kompleksitas model serta kemampuan model dalam mempelajari pola data.

Parameter *n_estimators* menentukan jumlah pohon keputusan yang dibangun dalam proses pembelajaran. Secara teoritis, semakin banyak jumlah pohon yang digunakan, maka model akan semakin stabil karena hasil prediksi diperoleh melalui mekanisme agregasi (*majority voting*). Namun, peningkatan jumlah pohon juga meningkatkan waktu komputasi dan kompleksitas model. Pada penelitian ini digunakan tiga variasi nilai *n_estimators*, yaitu 100, 200, dan 300. Nilai 100 dipilih sebagai batas minimal jumlah pohon yang umumnya sudah cukup stabil, sedangkan 200 dan 300 digunakan untuk menguji apakah peningkatan jumlah pohon dapat meningkatkan performa model secara signifikan.

Parameter kedua, yaitu *max_depth*, menentukan batas kedalaman setiap pohon keputusan. Semakin besar nilai *max_depth*, maka model akan semakin

kompleks karena mampu membentuk percabangan yang lebih dalam. Namun, kedalaman yang terlalu besar berpotensi menyebabkan *overfitting*. Dalam penelitian ini digunakan tiga variasi nilai *max_depth*, yaitu 10, 20, dan *None*. Nilai 10 digunakan untuk membatasi kompleksitas model, nilai 20 untuk melihat pengaruh kedalaman yang lebih besar, dan nilai *None* untuk mengizinkan pohon berkembang tanpa batas kedalaman.

Tabel 4. 13 Hasil *Tuning hyperparameter Random Forest*

No	*n_estimator*	*max_depth*	Akurasi (CV)
1	100	10	0.883333
2	300	10	0.883333
3	300	20	0.883333
4	300	None	0.883333
5	100	20	0.877778
6	200	20	0.877778
7	100	None	0.877778
8	200	10	0.877778
9	200	None	0.877778

Berdasarkan Tabel 4.6, kolom *n_estimators* menunjukkan jumlah pohon yang digunakan dalam model, sedangkan kolom *max_depth* menunjukkan kedalaman maksimum pohon keputusan. Kolom *Akurasi (CV)* merupakan nilai rata-rata akurasi hasil *10-Fold Cross Validation* untuk setiap kombinasi parameter.

Berdasarkan hasil pengujian, terdapat empat kombinasi parameter yang menghasilkan nilai akurasi tertinggi sebesar 0,883333 (88,33%), yaitu (100, 10), (300, 10), (300, 20), dan (300, None). Meskipun menghasilkan performa yang sama, kombinasi (100, 10) dipilih sebagai konfigurasi terbaik karena mampu memberikan akurasi optimal dengan kompleksitas model yang lebih rendah.

Hasil ini menunjukkan bahwa peningkatan jumlah pohon dari 100 menjadi 300 tidak memberikan peningkatan akurasi yang signifikan pada dataset penelitian

ini. Demikian pula, peningkatan kedalaman pohon dari 10 menjadi 20 atau tanpa batas (*None*) tidak menghasilkan perbedaan performa yang berarti.

Meskipun beberapa kombinasi menghasilkan akurasi yang sama, konfigurasi dengan *n_estimators* = 100 dan *max_depth* = 10 dipilih sebagai parameter terbaik. Pemilihan ini didasarkan pada pertimbangan efisiensi komputasi dan kompleksitas model. Dengan jumlah pohon yang lebih sedikit dan kedalaman yang dibatasi, model tetap mampu menghasilkan performa optimal tanpa meningkatkan risiko *overfitting* secara berlebihan.

Dengan demikian, hasil analisis menunjukkan bahwa peningkatan kompleksitas model tidak selalu berbanding lurus dengan peningkatan performa. Pemilihan *hyperparameter* yang tepat menjadi faktor penting dalam menghasilkan model *Random Forest* yang stabil, efisien, dan memiliki kemampuan generalisasi yang baik.

4.2.2 Analisis Pengaruh Split Data

Pengujian terhadap tiga skenario pembagian data dilakukan untuk menganalisis pengaruh proporsi data latih (training data) dan data uji (testing data) terhadap performa model *Random Forest*. Secara teoritis, semakin besar proporsi data latih, maka model memiliki peluang lebih besar untuk mempelajari pola data secara optimal. Namun, proporsi data uji yang terlalu kecil dapat membuat hasil evaluasi menjadi kurang representatif dan lebih sensitif terhadap kesalahan prediksi.

Komposisi pembagian data pada masing-masing skenario ditampilkan pada Tabel 4.7 berikut.

Tabel 4. 14 Hasil pembagian data pada tiga skenario split

No	Skenario Split	Data Latih	Data Uji	Jumlah Total
1	70:30	158	68	226
2	80:20	180	46	226
3	90:10	203	23	226

Hasil evaluasi model pada ketiga skenario menunjukkan bahwa nilai *accuracy* relatif konsisten, yaitu:

Tabel 4. 15 Perbandingan Nilai *Accuracy* pada Tiga Skenario Split

No	Skenario Split	*Accuracy*
1	70:30	0,9559
2	80:20	0,9565
3	90:10	0,9565

Deviasi skor *accuracy* antarskenario terpantau amat tipis, yakni di bawah 0,1%. Kondisi ini mengindikasikan bahwa perubahan komposisi kuantitas data training maupun data testing tidak membawa dampak yang berarti bagi tingkat presisi prediksi sistem.

Ditinjau dari aspek teoritis, proporsi 90:10 menawarkan keunggulan pada fase pengenalan pola lantaran algoritma berkesempatan belajar dari himpunan data yang jauh lebih berlimpah. Kendati demikian, minimnya kuantitas sampel uji (hanya 29 data) mengakibatkan metrik penilaian menjadi lebih rentan berfluktuasi akibat satu kekeliruan prediksi saja. Sebaliknya pada skema 70:30, walaupun disokong oleh jumlah data latih yang lebih terbatas (sebanyak 158 data), sistem terbukti tetap sanggup mencatatkan persentase akurasi yang memuaskan. Hal tersebut menegaskan bahwa karakteristik basis data yang diolah sudah sangat representatif dalam melahirkan model dengan kapabilitas generalisasi yang andal.

Selain *accuracy*, nilai *precision*, *recall*, dan *F1-score* pada ketiga skenario juga menunjukkan konsistensi yang tinggi. Hal ini mengindikasikan bahwa model

tidak hanya memiliki tingkat ketepatan prediksi yang baik secara keseluruhan, tetapi juga mampu menjaga keseimbangan dalam mengklasifikasikan kedua kelas (menang dan kalah).

Dengan demikian, dapat disimpulkan bahwa model *Random Forest* yang dibangun bersifat stabil dan robust terhadap variasi proporsi pembagian data. Performa model yang relatif konsisten pada ketiga skenario memperkuat kesimpulan bahwa model memiliki kemampuan generalisasi yang baik dan tidak bergantung pada satu konfigurasi split tertentu.

4.2.3 Analisis *Feature Importance* dan Implikasinya

Analisis *feature importance* dilakukan untuk mengetahui sejauh mana masing-masing variabel berkontribusi terhadap proses klasifikasi yang dilakukan oleh model *Random Forest*. Berbeda dengan subbab hasil yang hanya menyajikan nilai numerik, pada bagian ini dilakukan interpretasi terhadap faktor-faktor dominan yang memengaruhi prediksi kemenangan atlet.

Berdasarkan Tabel 4.9, seluruh fitur yang digunakan dalam penelitian ini berjumlah 15 atribut yang dikelompokkan ke dalam empat kategori utama, yaitu faktor fisik, teknik dan pengalaman, psikologis, serta latihan.

Faktor fisik terdiri dari usia, tinggi badan, berat badan, dan jenis kelamin. Faktor teknik dan pengalaman meliputi jumlah pertandingan, jumlah kemenangan, jumlah kekalahan, total medali, dan tahun mulai aktif bertanding. Faktor psikologis mencakup tingkat motivasi, kepercayaan diri, dan kemampuan mengontrol emosi. Sementara itu, faktor latihan meliputi frekuensi latihan per hari, frekuensi latihan per minggu, dan durasi latihan. Pengelompokan ini digunakan untuk mempermudah

analisis kontribusi masing-masing aspek terhadap hasil prediksi yang dihasilkan oleh model *Random Forest*.

Tabel 4.9 berikut menampilkan peringkat *feature importance* berdasarkan model terbaik yang telah melalui proses *tuning hyperparameter* dan evaluasi.

Tabel 4. 16 Analisis *Feature importance*s

No	Fitur	Nilai Importance
1	Berapa kali atlet berlatih dalam satu hari	0,1517
2	Kemampuan mengontrol emosi saat bertanding	0,1501
3	Total keseluruhan medali (emas, perak perunggu)	0,1137
4	Frekuensi Latihan dalam satu minggu	0,1022
5	Tingkat motivasi untuk menang	0,0961
6	Tinggi badan	0,0765
7	Berat badan	0,0674
8	Usia atlet saat ini	0,0598
9	Tingkat kepercayaan diri	0,0548
10	Jumlah pertandingan yang pernah diikuti	0,0294
11	Jumlah pertandingan resmi yang diikuti	0,0272
12	Pengalaman bertanding (total tahun aktif)	0,0253
13	Tahun mulai aktif bertanding	0,0216
14	Jenis kelamin	0,0153
15	Durasi rata-rata latihan	0,0087

Berdasarkan Tabel 4.9, terlihat bahwa fitur “berapa kali atlet berlatih dalam satu hari” memiliki nilai *importance* tertinggi sebesar 0,1517. Hal ini menunjukkan bahwa intensitas latihan harian merupakan faktor paling dominan dalam menentukan peluang kemenangan atlet.

Selain itu, kemampuan mengontrol emosi saat bertanding juga memiliki nilai *importance* yang hampir setara. Temuan ini menunjukkan bahwa aspek psikologis memiliki pengaruh yang signifikan dalam performa pertandingan. Atlet dengan kontrol emosi yang baik cenderung mampu mengambil keputusan lebih tepat dalam situasi tekanan tinggi.

Fitur pengalaman bertanding yang direpresentasikan melalui total medali dan jumlah pertandingan juga menunjukkan kontribusi yang cukup besar. Hal ini

mengindikasikan bahwa pengalaman kompetitif memberikan dampak terhadap kesiapan strategi dan adaptasi atlet selama pertandingan.

Jika ditinjau berdasarkan kategori fitur, terlihat bahwa fitur-fitur yang termasuk dalam faktor latihan dan psikologis cenderung memiliki nilai *importance* yang lebih tinggi dibandingkan dengan faktor fisik. Hal ini menunjukkan bahwa konsistensi latihan dan kesiapan mental memiliki peran yang lebih dominan dalam menentukan hasil pertandingan dibandingkan karakteristik fisik semata.

Sebaliknya, fitur seperti jenis kelamin dan durasi rata-rata latihan memiliki nilai *importance* yang relatif lebih rendah. Hal ini menunjukkan bahwa dalam konteks dataset penelitian ini, faktor tersebut tidak memberikan kontribusi signifikan terhadap hasil prediksi dibandingkan fitur lainnya.

Secara konseptual, hasil ini memperkuat pemahaman bahwa kemenangan dalam pencak silat tidak hanya dipengaruhi oleh kondisi fisik, tetapi juga oleh stabilitas mental dan intensitas latihan yang dilakukan secara konsisten. Dengan demikian, model *Random Forest* tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai alat analisis untuk mengidentifikasi faktor-faktor kunci yang perlu diperhatikan dalam proses pembinaan atlet.

Hasil analisis *feature importance* pada penelitian ini menunjukkan bahwa keberhasilan atlet tidak hanya dipengaruhi oleh faktor fisik semata, tetapi juga oleh aspek latihan, motivasi, dan kesiapan mental. Fitur-fitur seperti frekuensi latihan harian, motivasi, dan kemampuan mengontrol emosi memiliki kontribusi yang lebih besar dibandingkan beberapa atribut fisik lainnya. Temuan ini menunjukkan bahwa keberhasilan atlet dipengaruhi oleh kemampuan dalam mempersiapkan diri,

mengevaluasi kekurangan, serta melakukan perbaikan secara berkelanjutan melalui latihan yang konsisten dan pengendalian diri yang baik. Prinsip tersebut sejalan dengan ajaran Islam yang menekankan pentingnya muhasabah (evaluasi diri) dan persiapan yang matang untuk menghadapi masa depan. Allah SWT berfirman dalam Q.S. *Al-Hasyr* ayat 18:

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوا اتَّقُوا اللّٰهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍۚ وَاتَّقُوا اللّٰهَۗ اِنَّ اللّٰهَ خَبِيْرٌۢ ۢبِمَا تَعْمَلُوْنَ

"Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan." (QS. Al-Hasyr:18)

Menurut Tafsir Ibnu Katsir (1998), ayat ini mengandung perintah agar setiap manusia melakukan muhasabah (evaluasi diri) dan menyiapkan bekal amal untuk menghadapi hari esok. (Al-Sheikh, 2005).

Dalam konteks penelitian ini, proses evaluasi diri dapat diibaratkan sebagai upaya atlet dalam menilai kemampuan, memperbaiki kekurangan, serta meningkatkan kualitas latihan dan kesiapan mental sebelum menghadapi pertandingan. Hasil ini menunjukkan bahwa faktor-faktor yang berkaitan dengan proses pembinaan dan pengembangan diri, seperti intensitas latihan, motivasi, dan kemampuan mengontrol emosi, memiliki pengaruh yang besar terhadap peluang kemenangan atlet. Temuan tersebut sejalan dengan makna Q.S. *Al-Hasyr* ayat 18 yang mengajarkan pentingnya melakukan evaluasi terhadap diri sendiri dan mempersiapkan tindakan terbaik untuk menghadapi masa depan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model prediksi kemenangan atlet Pencak Silat pada Kejuaraan Dunia Tapak Suci 2025 berhasil dibangun menggunakan algoritma Random Forest berdasarkan data performa atlet. Hasil evaluasi menunjukkan bahwa seluruh skenario pembagian data (70:30, 80:20, dan 90:10) menghasilkan performa yang tinggi. Skenario pembagian data 80:20 menggunakan 10-*Fold Cross Validation* menghasilkan performa terbaik dengan mean *accuracy* sebesar 88,33% dan standar deviasi sebesar 0,0580, sehingga dipilih sebagai skenario yang paling optimal dan konsisten. Sedangkan pada skenario 80:20 tanpa *K-Fold Cross Validation*, model memperoleh nilai *accuracy* sebesar 95,65%, *precision* 95,99%, *recall* 96,56%, dan *F1-score* 95,64%, yang menunjukkan kemampuan generalisasi model yang baik dalam memprediksi hasil pertandingan. Selain itu, berdasarkan analisis *feature importance,* diketahui bahwa intensitas latihan harian, kemampuan mengontrol emosi saat bertanding, total perolehan medali, frekuensi latihan mingguan, dan tingkat motivasi untuk menang merupakan variabel yang paling berpengaruh terhadap hasil prediksi kemenangan atlet. Dengan demikian, tujuan penelitian untuk membangun model prediksi kemenangan atlet, mengevaluasi performa model, dan mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kemenangan atlet telah berhasil tercapai.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan adalah sebagai berikut:

1. Bagi Pelatih dan Pengurus Tapak Suci

   Hasil penelitian ini dapat dimanfaatkan sebagai bahan pertimbangan dalam proses pembinaan atlet, khususnya dalam meningkatkan konsistensi latihan dan kesiapan mental yang terbukti memiliki pengaruh besar terhadap kemenangan.

2. Bagi Pengembang Sistem

   Penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar dan melibatkan lebih banyak kategori pertandingan agar model memiliki kemampuan generalisasi yang lebih luas. Selain itu, dapat dilakukan perbandingan dengan algoritma klasifikasi lain untuk mengetahui metode yang paling optimal.

DAFTAR PUSTAKA

Ajgaonkar, Y., Bhoyar, K., Patil, A., & Shah, J. (2021). *Prediction of Winning Team using Machine.pdf* (p. 6).

Al-Sheikh, A. bin M. bin A. bin I. (2005). *Tafsir Ibnu Katsir 8.1.pdf*. Pustaka Imam asy-Syafi'i.

Alisti, S., Fitri, M., Paramitha, S. T., Apriady, H., & Ockta, Y. (2025). *Examining the Correlation between Mental.pdf* (p. 10). International Journal of Human Movement and Sports Sciences, HRPUB.

Allorerung, P. P., Erna, A., & Bagussahrir, M. (2024). *Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit*. *9*(3), 178–191.

Ashraf, J., Bhatti, S., & Talpur, S. (2021). Predicting the Best Team Players of Pakistan Super League using Machine Learning Algorithms. In *International Journal of Computer Applications* (Vol. 183, Issue 16, pp. 6–13). https://doi.org/10.5120/ijca2021921486

Bunker, R., & Susnjak, T. (2022). The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review. *Journal of Artificial Intelligence Research*, *73*, 1285–1322. https://doi.org/10.1613/jair.1.13509

Chen, J. (2024). Factor and Correlation Analysis for Predicting Marathon Race Performance Using Machine Learning Algorithms. In *Journal of Electrical Systems* (Vol. 20, Issue 6s, pp. 1948–1958). https://doi.org/10.52783/jes.3110

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Data Preprocessing and Classification for Predicting Student. In *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)* (Vol. 11, Issue 1, pp. 215–224).

Hadiana, O., Subarjah, H., Ma'mun, A., Mulyana, M., Budi, D. R., & Rahadian, A. (2022). Pencak Silat Tapak Suci: Overview in a Historical Perspective of Muhammadiyah Autonomic Organizations in Indonesia. In *JUARA : Jurnal Olahraga* (Vol. 7, Issue 2, pp. 408–416). https://doi.org/10.33222/juara.v7i2.1915

Jianjun, Q., Isleem, H. F., Almoghayer, W. J. K., & Khishe, M. (2025). Predictive athlete performance modeling with machine learning and biometric data integration. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-01438-9

John, A., & Damilola, A. (2021). *Design and Implementation of a Predictiv.pdf* (p. 17).

Jubeile Mark Baladjay, Nisce Riva, Ladine Ashley Santos, Dan Michael Cortez, Criselle Centeno, & Ariel Antwaun Rolando Sison. (2023). Performance evaluation of random forest algorithm for automating classification of mathematics question items. In *World Journal of Advanced Research and Reviews* (Vol. 18, Issue 2, pp. 034–043). https://doi.org/10.30574/wjarr.2023.18.2.0762

Kampezidou, S. I., Ray, A. T., Bhat, A. P., Pinon Fischer, O. J., & Mavris, D. N. (2024). *Styliani I. Kampezidou Fundamental Components and Principles of Supervised Machine Learning Workflows with Numerical and Categor.pdf* (p. 33).

KC, S., Kumar, B., Tolani, D., & Patel, H. (2021). *C. SrikantaiahK Prediction of IPL Match Outcome Using Machine Learning Techniques [2021].pdf* (p. 8).

Lampis, T., Ioannis, N., Vasilios, V., & Stravrianna, D. (2023). *predictions of european basketball match results with machine learning algorithms.pdf* (p. 20).

Li, S., Huo, S., & Ke, W. (2021). Intelligent Decision-Making System for Martial Arts Competition Using Deep Learning. *Mobile Information Systems*, *2021*. https://doi.org/10.1155/2021/9920751

Mienye, I. D., & Jere, N. (2024). *A Survey of Decision Trees: Concepts, Algorithms, and Applications.pdf* (p. 15).

Miller, C., Portlock, T., Nyaga, D. M., & O'Sullivan, J. . (2024). *A review of model evaluation metrics for machine learning in genetics and genomics Catriona.pdf* (p. 13).

Nandekdkar, A., Gadve, C., Gowda, C., Deep, S., & Mulla, R. (2024). *Ameya Nandedkar - Athlete Performance Prediction Using Random Forest [2024].pdf* (p. 10).

Nguyen, H. N. (2022). *Nguyen Hoang Nguyen The application of machine learning and deep learning in sport predicting NBA players' performance and popul.pdf* (p. 21).

Nguyen, J. M., Jezequel, P., Gillois, P., Silva, L., Azzouz, F. Ben, Lambert-Lacroix, S., Campone, M., Gaultier, A., & Moreau-Gaudry, A. (2022). *Random forest of perfect trees concept, performance, applications and perspectives.pdf* (p. 11).

Patah, I. A., Jumareng, H., Setiawan, E., Aryani, M., & Gani, R. A. (2021). *The Importance of Physical Fitness for P.pdf* (p. 12). Journal Sport Area, Universitas Islam Riau.

Pongthao, J., Na-udom, A., & Rungrattanaubol, J. (2025). *Machine Learning Classification with Logistic Regression Feature Selection Approach on Health Datasets.pdf* (p. 9).

Razak, B. (2024). *PERATURAN PERTANDINGAN PENCAK SILAT TERBARU 2024*. Pengurus Besar Ikatan Pencak Silat Indonesia (PB IPSI).

Salman, H. A., Kalakech, A., & Steiti, A. (2024). *View of Random Forest Algorithm Overview.pdf* (p. 11).

Sanjaykumar, S., Natarajan, S., Lakshmi, P. Y., & Boby, F. A. (2024a). Predicting Team Success in the Indian Premier League Cricket 2024 Season Using Random Forest Analysis. *Physical Education Theory and Methodology* , *24*(2), 304–309. https://doi.org/10.17309/tmfv.2024.2.16

Sanjaykumar, S., Natarajan, S., Lakshmi, P. Y., & Boby, F. A. (2024b). *Swamynathan Sanjaykumar Predicting Team Success in the Indian Premier League Cricket 2024 Season Using Random Forest Analysis.pdf* (p. 6).

Sasmita, K., Barlian, E., & Padli. (2023). *Khairul Sasmita - Pencak Silat Wajah Budaya Bangsa Indonesia [2023].pdf* (p. 12). Journal on Education (JoE), Universitas Negeri Padang.

Sindar, S. A., Saputri, B., & Aulia, N. (2025). *Pemodelan Classification and Regression Tree (CART) Pada Klasifikasi Gaya Hidup Sehat Menggunakan Pendekatan User-Based Classification*. *4*, 1028–1036.

Subekti, N., Syaifullah, R., Fatoni, M., Syaukani, A. A., Warthadi, A. N., & Arni-Rayhan, A. R. (2021). Pencak silat combat match: Time motion analysis in elite athletes championship. *Journal of Human Sport and Exercise*, *16*(Proc4), S1597–S1608. https://doi.org/10.14198/jhse.2021.16.Proc4.05

Sujon, K. M., Hassan, R., Choi, K., & Samad, M. A. (2025). *Accuracy_precision_recall_f1-score_or_MCC_empirica.pdf* (p. 46).

Yu, Y., Wang, L., Huang, H., & Yang, W. (2020). *An Improved Random Forest Algorithm*. https://doi.org/10.1088/1742-6596/1646/1/012070

Yulianto, E., Mulyana, M., Yudiana, Y., & Hendarsin, F. (2022). The Effect of Interval Training on Anaerobic Capacity Improvement related to Pencak Silat Athlete Performance. In *Jurnal Pendidikan Jasmani dan Olahraga* (Vol. 7, Issue 1, pp. 96–103). https://doi.org/10.17509/jpjo.v7i1.42620

Zhou, Q. (2022). *sports achievement prediction and influencing factors.pdf* (p. 8).