

**PREDIKSI *MARKET VALUE* PEMAIN SEPAK BOLA MENGGUNAKAN
ALGORITMA *RANDOM FOREST* (STUDI KASUS PADA PEMAIN
MUDA EROPA BERBAKAT U-22)**

SKRIPSI

Oleh :
MUZZAMMIL RAHMAT EDI
NIM. 210605110113



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**PREDIKSI *MARKET VALUE* PEMAIN SEPAK BOLA MENGGUNAKAN
ALGORITMA *RANDOM FOREST* (STUDI KASUS PADA PEMAIN
MUDA EROPA BERBAKAT U-22)**

SKRIPSI

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

Oleh :
MUZZAMMIL RAHMAT EDI
NIM. 210605110113

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN

**PREDIKSI MARKET VALUE PEMAIN SEPAK BOLA MENGGUNAKAN
ALGORITMA RANDOM FOREST (STUDI KASUS PADA PEMAIN
MUDA EROPA BERBAKAT U-22)**

SKRIPSI

Oleh :
MUZZAMMIL RAHMAT EDI
NIM. 210605110113

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 11 Juni 2025

Pembimbing I,



Syahiduz Zaman, M.Kom
NIP. 19700502 200501 1 005

Pembimbing II,



Nurizal Dwi Priandani, M.Kom
NIP. 19920830 202203 1 001

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



Dr. Ir. Faqihul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

PREDIKSI MARKET VALUE PEMAIN SEPAK BOLA MENGGUNAKAN ALGORITMA RANDOM FOREST (STUDI KASUS PADA PEMAIN MUDA EROPA BERBAKAT U-22)

SKRIPSI

Oleh :
MUZZAMMIL RAHMAT EDI
NIM. 210605110113

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 25 Juni 2025

Susunan Dewan Penguji

Ketua Penguji : Dr. M. Ainul Yaqin, M.Kom
NIP. 19761013 200604 1 004

Anggota Penguji I : Dr. Agung Teguh Wibowo Almais, MT
NIP. 19860301 202321 1 016

Anggota Penguji II : Syahiduz Zaman, M.Kom
NIP. 19700502 200501 1 005

Anggota Penguji III : Nurizal Dwi Priandani, M.Kom
NIP. 19920830 202203 1 001

()
()
()
()

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




Dr. Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Muzzammil Rahmat Edi
NIM : 210605110113
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Prediksi *Market Value* Pemain Sepak Bola
Menggunakan Algoritma *Random Forest* (Studi Kasus Pada Pemain Muda Eropa Berbakat U-22)

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, 25 Juni 2025
Yang membuat pernyataan,



Muzzammil Rahmat Edi
NIM.210605110113

MOTTO

“Semuanya tergantung mental!!”

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala karunia, rahmat, dan kasih sayang-Nya yang selalu menyertai setiap langkah penulis hingga skripsi ini dapat diselesaikan. Karya ini penulis dedikasikan kepada:

Bunda tercinta,

Yang senantiasa menjadi sumber cinta yang tiada batas, memberikan dukungan, doa, dan semangat yang tak pernah putus, serta menjadi cahaya penerang dalam setiap langkah penulis.

Ayah tercinta,

Yang selalu mengajarkan arti keteguhan, kerja keras, serta memberikan kasih sayang yang tulus tanpa syarat, dan menjadi teladan yang senantiasa menginspirasi penulis.

Kakak dan Adik tercinta,

Yang selalu memberikan inspirasi dengan kebijaksanaan dan perhatian, serta selalu menyemangati dengan cinta dan kebaikan yang tiada henti.

Diri sendiri,

Terima kasih telah berjuang tanpa mengenal lelah, tetap teguh menghadapi segala tantangan, dan menyelesaikan setiap langkah ini dengan penuh kesabaran.

Semoga karya ini menjadi awal dari perjalanan yang lebih berarti.

KATA PENGANTAR

Segala puji syukur senantiasa penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Prediksi Market Value Pemain Sepak Bola Menggunakan Algoritma Random Forest (Studi Kasus pada Pemain Muda Eropa Berbakat U-22)” ini dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, suri teladan umat manusia, beserta keluarga, sahabat, dan pengikutnya hingga akhir zaman.

Ucapan terima kasih sebanyak-banyaknya saya ucapkan kepada:

1. Prof. Dr. H. M. Zainuddin, MA selaku Rektor UIN Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT, IPU, selaku Kepala Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
4. Syahiduz Zaman, M. Kom. selaku dosen pembimbing pertama skripsi yang telah membimbing dan memberikan arahan dari awal perkuliahan hingga proses penyusunan skripsi ini.
5. Nurizal Dwi Priandani, M.Kom, selaku dosen pembimbing kedua yang dengan penuh perhatian memberikan masukan, motivasi, dan kritik membangun demi penyempurnaan karya ini.
6. Dr. M. Ainul Yaqin, M.Kom, selaku Ketua Dewan Penguji, atas saran dan koreksi yang sangat berharga dalam proses ujian skripsi.
7. Dr. Agung Teguh Wibowo Almais, M.T, selaku penguji kedua, atas masukan dan evaluasi yang telah meningkatkan kualitas dari penelitian ini.

8. Kedua orang tua tercinta, Ayah dan Bunda, serta kakak dan adik saya, yang selalu menjadi sumber doa, kekuatan, dan inspirasi yang tak ternilai.
9. Seluruh dosen Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, atas ilmu dan dedikasi yang telah diberikan selama masa perkuliahan.
10. Dua sahabat terdekat saya, Willy dan Adam, yang selalu memberikan semangat, bantuan, serta menjadi tempat berbagi dalam setiap tantangan selama proses penyusunan skripsi ini.
11. Kakanda Ochang dan Kakanda Serlinda yang telah banyak memberikan saran serta bantuan hingga penulis terdorong untuk menyusun skripsi ini.
12. Teman-teman KelcemBois, kehadiran kalian telah memberikan warna tersendiri dalam perjalanan ini. Bersama kalian, penulis tidak pernah merasa sendiri, bahkan di tengah kesulitan sekalipun.
13. Saudari Kesya Az-Zayra Putri Irawan yang selalu hadir memberi semangat dan dengan sabar menemani perjalanan penulis dalam menyusun skripsi ini.
14. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan dukungan secara langsung maupun tidak langsung.

Penulis telah berupaya semaksimal mungkin dalam menyusun skripsi ini. Namun, penulis menyadari bahwa karya ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang.

Malang, 10 Juni 2025



Penulis

DAFTAR ISI

HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
ABSTRAK	xiv
ABSTRACT	xv
المخلص	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II STUDI PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.2 Nilai Pasar	12
2.3 Pemain Sepak Bola	13
2.4 Normalisasi Min-Max Scaling	14
2.5 Algoritma Random Forest	14
BAB III DESAIN DAN IMPLEMENTASI	16
3.1 Desain Penelitian	16
3.2 Pengumpulan Data	16
3.3 Desain Sistem	17
3.4 Tahapan Desain Sistem	18
3.4.1 Pembersihan Data	18
3.4.2 Transformasi Data	19
3.4.3 Normalisasi Data Menggunakan Min-Max Scaling	19
3.4.4 Split Data	21
3.5 Penerapan Algoritma Random Forest	22
3.5.1 Bootstrap Sampling	22
3.5.2 Pemilihan Fitur Secara Acak	22
3.5.3 Pemilihan Split Terbaik	23
3.5.4 Membangun <i>Decision Tree</i>	23
3.5.5 Agregasi Prediksi	24
3.6 Evaluasi Model	24
3.6.1 Mean Squared Error (MSE)	24
3.6.2 R-Squared (R^2)	25
3.6.3 <i>Root Mean Squared Error (RMSE)</i>	25
3.7 Skenario Uji Coba	26
BAB IV HASIL DAN PEMBAHASAN	29

4.1 Hasil Penerapan Random Forest	29
4.1.1 Hasil Bootstrap Sampling	29
4.1.2 Hasil Pemilihan Fitur Acak dan Split Terbaik	30
4.1.3 Hasil Decision Tree	31
4.1.3 Hasil Agregasi Prediksi	32
4.2 Hasil Evaluasi Algoritma Random Forest	33
4.3 Hasil Pengujian <i>K-Fold Cross Validation</i>	35
4.4 Hasil Pengujian <i>Hyperparameter</i>	36
4.5 Pengaruh Jumlah Fitur terhadap Akurasi	38
4.5.1 Model A	38
4.5.2 Model B	39
4.5.3 Model C	39
4.6 Hasil Pengujian Pengaruh Fitur	40
4.7 Pembahasan	41
4.8 Integrasi Penelitian	45
BAB V KESIMPULAN DAN SARAN	47
5.1 Kesimpulan	47
5.2 Saran	47
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3. 1 Desain Penelitian.....	16
Gambar 3. 2 Alur Desain Sistem	18
Gambar 4. 1 Hasil Pemilihan Fitur Acak.....	30
Gambar 4. 2 Hasil Split terbaik.....	31
Gambar 4. 3 Distribusi Error Prediksi Random Forest (Default)	35
Gambar 4. 4 Distribusi Error Prediksi Hyperparameter	38
Gambar 4. 5 Perbandingan Nilai Prediksi dan Nilai Asli	42
Gambar 4. 6 Pengujian 5-Fold Cross Validation	42
Gambar 4. 7 Perbandingan Model A, B, dan C	43
Gambar 4. 8 Hasil Pengujian Pengaruh Fitur	44

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 3. 1 Dekripsi Dataset	17
Tabel 3. 2 Data Setelah Pembersihan.....	18
Tabel 3. 3 Data Setelah Transformasi	19
Tabel 3. 4 Data Setelah Dinormalisasi	21
Tabel 3. 5 Contoh Pemilihan Fitur Secara Acak.....	23
Tabel 3. 6 Contoh Tabel Split di 82	23
Tabel 3. 7 Contoh Hasil Decision Tree	24
Tabel 3. 8 Contoh tabel 5-fold cross validation	26
Tabel 3. 9 Parameter yang digunakan	27
Tabel 3. 10 Contoh Simulasi Perhitungan FeatureImportant.....	28
Tabel 4. 1 Hasil bootsrap sampling.....	29
Tabel 4. 2 Hasil prediksi decision tree	31
Tabel 4. 3 Hasil Agregasi Random Forest	32
Tabel 4. 4 Hasil Evaluasi Algoritma Random Forest (Default) Tanpa Tuning .	33
Tabel 4. 5 Hasil prediksi nilai pasar pada sepuluh entri data uji (default).....	34
Tabel 4. 6 Hasil Pengujian 5-Fold.....	35
Tabel 4. 7 Rata-rata 5-Fold	35
Tabel 4. 8 Kombinasi hyperparameter terbaik.....	36
Tabel 4. 9 Hasil Evaluasi dengan Tuning Hyperparameter	37
Tabel 4. 10 Hasil prediksi nilai pasar (tuning hyperparameter).....	37
Tabel 4. 11 Hasil Evaluasi Model A	39
Tabel 4. 12 Hasil Evaluasi Model B	39
Tabel 4. 13 Hasil Evaluasi Model	40
Tabel 4. 14 Hasil Pengujian Pengaruh Fitur	40

ABSTRAK

Muzzammil Rahmat Edi. 2025. **Prediksi Market Value Pemain Sepak Bola Menggunakan Algoritma Random Forest (Studi Kasus pada Pemain Muda Eropa Berbakat U-22)**. Skripsi, Program Studi Teknik Informatika Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) H. Syahiduz Zaman, M.Kom (II) Nurizal Dwi Priandani, M.Kom.

Kata Kunci: *Random Forest, Market Value, Pemain Sepak Bola, Prediksi, Machine Learning.*

Penelitian ini bertujuan untuk memprediksi nilai pasar (market value) pemain sepak bola muda berbakat usia U-22 di Eropa menggunakan algoritma Random Forest. Atribut yang digunakan dalam model meliputi usia, posisi, overall rating, potensial, tinggi badan, dan berat badan. Dataset diperoleh dari platform Kaggle dan melalui proses pra-pemrosesan seperti pembersihan data, transformasi format, normalisasi Min-Max, serta pembagian data latih dan uji dengan rasio 80:20. Evaluasi model dilakukan menggunakan metrik R^2 , MAE, dan RMSE. Hasil menunjukkan bahwa Random Forest mampu memprediksi nilai pasar secara akurat dengan nilai R^2 mencapai 0.95 dan MAE di bawah €100.000. Pengujian lebih lanjut menggunakan K-Fold Cross Validation dan tuning hyperparameter menunjukkan kestabilan model, serta mengidentifikasi bahwa overall rating dan potensial adalah fitur paling berpengaruh dalam prediksi nilai pasar pemain.

ABSTRACT

Muzzammil Rahmat Edi. 2025. **Prediction of Football Players' Market Value Using the Random Forest Algorithm (Case Study on Talented U-22 European Players)**. Undergraduate Thesis, Informatics Engineering Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) H. Syahiduz Zaman, M.Kom (II) Nurizal Dwi Priandani, M.Kom.

Keywords: *Random Forest, Market Value, Football Players, Prediction, Machine Learning.*

This research aims to predict the market value of talented U-22 European football players using the Random Forest algorithm. The model utilizes attributes such as age, position, overall rating, potential, height, and weight. The dataset was obtained from Kaggle and underwent preprocessing steps including data cleaning, format transformation, Min-Max normalization, and an 80:20 training-testing split. Model performance was evaluated using R^2 , MAE, and RMSE metrics. The results show that Random Forest can accurately predict market value, achieving an R^2 score of 0.95 and a MAE below €100,000. Further evaluation using K-Fold Cross Validation and hyperparameter tuning demonstrated the model's stability and identified overall rating and potential as the most influential features in predicting a player's market value

الملخص

مزمّل رحمة إيدي. م٢٠٢٥. تنبؤ بقيمة السوق للاعبى كرة القدم باستخدام خوارزمية الغابة العشوائية (دراسة حالة على لاعبي أوروبا، الموهوبين تحت سن ٢٢ عامًا). مشروع تخرج لنيل درجة البكالوريوس، برنامج هندسة المعلومات، كلية العلوم والتكنولوجيا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانغ. المشرفان: (١) ح. شهيد الزمان، ماجستير في الحاسوب، (٢) نوريزال دوي بريانداني، ماجستير في الحاسوب.

الكلمات المفتاحية: الغابة العشوائية، قيمة السوق، لاعبو كرة القدم، التنبؤ، التعلم الآلي

تهدف هذه الدراسة إلى التنبؤ بقيمة السوق للاعبى كرة القدم الموهوبين تحت سن ٢٢ عامًا في أوروبا باستخدام خوارزمية الغابة العشوائية. يعتمد النموذج على خصائص مثل العمر، المركز، التقييم العام، الإمكانات، الطول، والوزن. تم الحصول على مجموعة وتقسيم، و **Min-Max** وتمت معالجتها من خلال تنظيف البيانات، وتحويل الصيغ، والتطبيع باستخدام **Kaggle** البيانات من منصة أظهرت النتائج **RMSE**، و **MAE**، و **R²** البيانات بنسبة ٨٠:٢٠ للتدريب والاختبار. تم تقييم أداء النموذج باستخدام مؤشرات أقل من **MAE** ١٠٠,٠٠٠ تقدر بـ ٠,٩٥ وخطأ **R²** أن الخوارزمية تستطيع التنبؤ بدقة بقيمة السوق، حيث حقق النموذج قيمة وضبط المعاملات فعالية النموذج، وبيّن أن التقييم العام والإمكانات هما العاملان **Cross Validation** يورو. كما أثبت اختبار الأكثر تأثيرًا في تحديد قيمة اللاعب السوقية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sepak bola merupakan olahraga paling populer di dunia dengan industri yang terus berkembang pesat (Lončar & Špehar, 2021). Dalam industri ini, pengelolaan pemain menjadi aspek yang sangat penting, mencakup seleksi, penilaian, serta transfer pemain. Salah satu elemen utama dalam pengelolaan pemain adalah penilaian nilai pasar, yang sering kali menjadi acuan utama dalam pengambilan keputusan oleh klub sepak bola. Nilai pasar pemain tidak hanya dipengaruhi oleh keterampilan dan performa di lapangan, tetapi juga oleh faktor-faktor lain seperti usia, posisi bermain, serta potensi masa depan (Leksowski, 2022). Allah SWT berfirman dalam Al-Qur'an:

لَهُ مُعَقِّبَاتٌ مِّنْ بَيْنِ يَدَيْهِ وَمِنْ خَلْفِهِ يَحْفَظُونَهُ ۗ مِنْ أَمْرِ اللَّهِ ۗ إِنَّ اللَّهَ لَا يُعَيِّرُ مَا بِقَوْمٍ حَتَّىٰ يُعَيِّرُوا مَا بِأَنْفُسِهِمْ ۗ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ ۗ وَمَا لَهُمْ مِّنْ دُونِهِ ۗ مِنْ وَّالٍ

“Baginya (manusia) ada malaikat-malaikat yang selalu menjaganya bergiliran, dari depan dan belakangnya. Mereka menjaganya atas perintah Allah. Sesungguhnya Allah tidak mengubah keadaan suatu kaum sebelum mereka mengubah keadaan diri mereka sendiri. Dan apabila Allah menghendaki keburukan terhadap suatu kaum, maka tak ada yang dapat menolaknya dan tidak ada pelindung bagi mereka selain Dia.” (QS. Ar-Ra'd: 11)

Hingga saat ini, penilaian nilai pasar pemain masih dilakukan secara subjektif oleh manajer dan scout tim. Keputusan yang dibuat berdasarkan intuisi dan pengalaman sering kali menghasilkan bias yang dapat mengurangi akurasi estimasi nilai pemain. Selain itu, ketidaktepatan dalam menentukan nilai pasar pemain dapat

berakibat pada kesalahan dalam strategi transfer, yang dapat berdampak negatif bagi keuangan klub (Di Domizio et al., 2024). Oleh karena itu, diperlukan pendekatan yang lebih objektif dan berbasis data dalam menentukan nilai pasar pemain.

Dalam beberapa tahun terakhir, kemajuan teknologi dan perkembangan data science telah membuka peluang baru dalam analisis data sepak bola. Analisis data dapat membantu mengidentifikasi pola dalam performa pemain dan memberikan estimasi nilai pasar yang lebih akurat. Penggunaan metode machine learning, khususnya algoritma Random Forest, telah terbukti efektif dalam memprediksi nilai pasar pemain sepak bola berdasarkan berbagai atribut statistik dan historis (Sevgen & Aliefendioğlu, 2020).

Random Forest merupakan algoritma berbasis ensemble yang mampu mengolah data kompleks dan menangkap hubungan non-linear antara berbagai variabel. Algoritma ini terdiri dari sejumlah decision tree yang bekerja secara bersamaan untuk meningkatkan akurasi prediksi. Dibandingkan dengan metode tradisional, Random Forest lebih unggul dalam menangani data besar dan variabel yang beragam, sehingga dapat memberikan hasil yang lebih akurat dan dapat diandalkan (Bhadula et al., 2024).

Dalam konteks sepak bola, penerapan Random Forest dapat membantu klub dalam mengidentifikasi pemain berbakat dengan nilai pasar yang sesuai. Dengan mempertimbangkan faktor seperti usia, posisi bermain, rating keseluruhan, potensi, dan tim tempat bermain, model ini dapat memberikan wawasan yang lebih objektif dalam menentukan nilai pasar pemain. Hal ini tidak hanya membantu klub dalam

membuat keputusan transfer yang lebih baik, tetapi juga memberikan keuntungan dalam aspek manajemen keuangan (Baouan et al., 2022).

Pendekatan berbasis data telah terbukti lebih akurat dibandingkan dengan metode subjektif dalam menilai nilai pasar pemain. Penerapan machine learning dalam analisis data sepak bola memungkinkan klub untuk meningkatkan akurasi prediksi nilai pasar pemain, sehingga dapat mengoptimalkan keputusan transfer mereka (Huang & Zhang, 2023). Dengan menggunakan model yang berbasis data, klub dapat menghindari subjektivitas dan memastikan bahwa keputusan yang diambil didukung oleh informasi yang valid dan terukur.

Prediksi nilai pasar pemain sepak bola menjadi sangat penting karena berkaitan langsung dengan strategi investasi klub, perencanaan jangka panjang, dan kestabilan finansial. Klub-klub sepak bola profesional mengalokasikan anggaran besar untuk aktivitas transfer pemain, sehingga estimasi nilai pasar yang akurat menjadi faktor krusial agar klub tidak mengalami kerugian akibat overpricing atau undervaluation pemain. Masalah utama yang sering terjadi adalah ketidaksesuaian antara performa nyata pemain dengan nilai pasar yang dipatok, yang menyebabkan terjadinya transaksi yang tidak efisien. Selain itu, fluktuasi harga yang dipengaruhi oleh faktor eksternal seperti popularitas, media sosial, atau tekanan pasar membuat penilaian menjadi semakin kompleks. Hal ini mendorong perlunya pendekatan yang sistematis dan berbasis data untuk mengurangi ketergantungan pada penilaian subjektif, serta meminimalkan risiko kesalahan dalam pengambilan keputusan manajemen pemain.

Dengan semakin pesatnya perkembangan teknologi dan meningkatnya ketersediaan data sepak bola, pendekatan berbasis data diprediksi akan menjadi standar dalam pengelolaan pemain dan strategi transfer di masa depan. Oleh karena itu, penelitian ini tidak hanya memberikan kontribusi dalam konteks akademik, tetapi juga memiliki dampak praktis yang dapat diterapkan dalam industri sepak bola profesional (Mastromartino et al., 2024).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah bagaimana performa random forest dalam memprediksi market value pemain dan mengidentifikasi faktor yang paling berpengaruh dalam transfer pemain sepak bola U-22 Eropa?

1.3 Batasan Masalah

Penelitian ini hanya berfokus pada prediksi nilai pasar pemain sepak bola muda berbakat usia 16-22 tahun menggunakan algoritma Random Forest, dengan mempertimbangkan faktor seperti usia, posisi bermain, rating keseluruhan, potensi, tinggi badan, dan berat badan. Data yang digunakan berasal dari dataset yang tersedia tanpa menggabungkan sumber eksternal. Evaluasi model dilakukan menggunakan metrik akurasi, mean squared error (MSE), *Root Mean Squared Error (RMSE)*, dan R-squared, tanpa teknik validasi silang yang kompleks. Faktor eksternal seperti popularitas pemain dan kondisi finansial klub tidak termasuk dalam analisis.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk membangun model prediksi nilai pasar pemain sepak bola muda berbakat usia 16-22 tahun menggunakan algoritma Random Forest berdasarkan faktor-faktor seperti usia, posisi bermain, rating keseluruhan, potensi, tinggi badan, dan berat badan. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi faktor-faktor yang paling berpengaruh dalam menentukan nilai pasar pemain, sehingga dapat memberikan wawasan yang lebih objektif bagi klub dalam pengambilan keputusan transfer.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat akademis dengan mengembangkan penerapan algoritma Random Forest dalam prediksi nilai pasar pemain sepak bola muda berbakat, serta memberikan manfaat praktis bagi klub sepak bola dalam pengambilan keputusan transfer yang lebih objektif dan berbasis data. Selain itu, penelitian ini juga memiliki manfaat teknologis dalam meningkatkan pemanfaatan machine learning dalam industri olahraga, khususnya dalam analisis dan pengelolaan pemain sepak bola.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terdahulu

Penelitian yang dilakukan oleh (Lee et al., 2022) menggunakan metode Light Gradient Boosting Machine (LightGBM) yang dioptimalkan dengan algoritma Tree-structured Parzen Estimator (TPE) untuk memprediksi nilai pasar pemain sepak bola. Hasil penelitian menunjukkan bahwa model yang diusulkan memiliki akurasi lebih tinggi dibandingkan model regresi baseline berdasarkan metrik RMSE. Temuan ini menunjukkan bahwa pendekatan berbasis data dapat digunakan sebagai rekomendasi dalam keputusan transfer pemain. Relevansi penelitian ini terletak pada penggunaan data statistik dan teknik machine learning dalam prediksi nilai pemain, meskipun algoritma yang digunakan berbeda dengan penelitian yang saya lakukan.

Huang dan Zhang (2023) meneliti penggunaan Gradient Boosting Decision Tree (GBDT) dengan interpretabilitas melalui Shapley Additive Explanations (SHAP) untuk menganalisis faktor yang mempengaruhi nilai pasar pemain sepak bola. Hasil penelitian menunjukkan bahwa atribut keterampilan teknis, kebugaran fisik, dan kognitif memiliki pengaruh signifikan terhadap nilai pemain. Model yang dapat dijelaskan ini penting bagi pengambilan keputusan berbasis data. Penelitian ini relevan karena menekankan pentingnya transparansi dalam prediksi nilai pemain, yang dapat diaplikasikan dalam algoritma Random Forest yang saya gunakan.

Firdausi et al. (2024) menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi nilai pasar pemain di lima liga top Eropa. Dengan $K=3$, model ini mencapai tingkat kesalahan rata-rata (MAPE) sebesar 10,45%. Hasil penelitian menunjukkan bahwa atribut performa dan posisi bermain merupakan faktor utama dalam prediksi, meskipun KNN memiliki keterbatasan dalam menangani dataset kompleks. Studi ini memberikan dasar kuat untuk membandingkan akurasi model prediksi dengan algoritma lain seperti Random Forest.

Penelitian yang dilakukan oleh (Baouan et al., 2022) menggunakan Lasso Regression dan Random Forest untuk mengidentifikasi faktor signifikan dalam memprediksi nilai pasar pemain sepak bola. Studi ini menemukan bahwa selain variabel utama seperti usia dan performa, terdapat hubungan kompleks antara beberapa fitur dengan nilai pasar masa depan pemain. Model ini dapat meranking pemain berdasarkan prediksi nilai mereka, yang dapat diterapkan dalam menilai bakat pemain muda U-22.

Shen (2025) membandingkan kinerja berbagai model machine learning, termasuk Random Forest, Gradient Boosting, dan Support Vector Machines (SVM), dalam memprediksi nilai pasar pemain sepak bola. Hasil penelitian menunjukkan bahwa Random Forest memiliki performa terbaik dengan R^2 sebesar 0,92. Faktor utama dalam prediksi adalah rating keseluruhan, usia, dan posisi bermain. Temuan ini mendukung penggunaan Random Forest dalam analisis data pemain sepak bola dan relevan dengan penelitian saya.

Sulimov (2024) meneliti prediksi biaya transfer pemain sepak bola menggunakan pendekatan berbasis kecerdasan buatan (AI). Model yang

dikembangkan menggunakan algoritma ensemble dan dievaluasi dengan metrik seperti R^2 dan RMSE. Hasilnya menunjukkan bahwa pendekatan ini mampu memberikan wawasan mendalam tentang strategi transfer klub. Meskipun fokus utama penelitian ini adalah biaya transfer, metode dan teknik yang digunakan dapat diterapkan dalam model prediksi nilai pasar pemain muda berbakat.

Wang et al. (2024) meneliti sistem prediksi hasil pertandingan sepak bola menggunakan kombinasi algoritma PCA, K-Nearest Neighbors (KNN), Random Forests, dan Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa Random Forest memiliki performa terbaik, meskipun mengalami overfitting. Studi ini merekomendasikan teknik regularisasi untuk meningkatkan akurasi dalam skenario nyata. Meskipun penelitian ini berfokus pada prediksi hasil pertandingan, penggunaan Random Forest untuk mengolah data multi-variabel tetap relevan dengan penelitian saya.

Hu (2024) menggunakan Back Propagation Neural Network (BPNN) untuk memprediksi hasil pertandingan sepak bola di Bundesliga Jerman dengan mengintegrasikan data handicap sebagai input model. Hasil penelitian menunjukkan akurasi rata-rata 57,2%, dengan performa yang stabil. Meskipun metode yang digunakan berbeda dengan penelitian saya, studi ini menunjukkan bagaimana variabel tambahan dapat meningkatkan akurasi model prediksi.

Yeung et al. (2024) mengeksplorasi model prediksi pertandingan menggunakan CatBoost (Gradient-Boosted Trees) dengan fitur pi-ratings. Hasil penelitian menunjukkan bahwa model ini memiliki performa lebih baik dibandingkan model sebelumnya dalam Soccer Prediction Challenge 2017. Studi

ini relevan karena mengeksplorasi optimasi fitur dan evaluasi model dalam prediksi sepak bola, yang dapat diterapkan dalam penelitian prediksi nilai pasar pemain.

Ren dan Susnjak (2022) menggunakan berbagai algoritma machine learning dan Kelly Index untuk mengelompokkan pertandingan berdasarkan tingkat kesulitan prediksi. Hasil penelitian menunjukkan bahwa strategi berbasis Kelly Index memberikan hasil yang kompetitif dan mampu menghasilkan keuntungan dengan pendekatan konservatif. Penelitian ini relevan karena menerapkan explainable AI dalam prediksi berbasis data sepak bola, yang dapat digunakan untuk meningkatkan interpretabilitas model prediksi nilai pasar pemain.

Berikut tabel dari penelitian terdahulu yang telah diuraikan pada paragraph sebelumnya:

Tabel 2.1 Penelitian Terdahulu

Sumber	Input	Metode	Output	Hasil
Hansoo Lee, Bayu Adhi Tama, Meeyoung Cha (2022)	Data statistik pemain sepak bola	LightGBM + optimasi TPE	Prediksi nilai pasar pemain	Model lebih akurat dibanding baseline (RMSE), dan mendukung rekomendasi transfer berbasis data
Chunyang Huang, Shaoliang Zhang (2023)	Atribut keterampilan, kebugaran, dan kognitif	Gradient Boosting Decision Tree (GBDT) + SHAP	Prediksi nilai pasar pemain	Atribut teknis, fisik, dan kognitif signifikan. Model bersifat explainable dan mendukung pengambilan keputusan
Hirzi Firdausi, I Wayan Sumarjaya, I Gusti Ayu Made Srinadi (2024)	Performa dan posisi pemain di lima liga top Eropa	K-Nearest Neighbor (KNN)	Prediksi nilai pasar pemain	K=3, MAPE 10,45%. KNN akurat namun kurang cocok untuk data kompleks; RF disarankan sebagai alternatif
Ali Baouan et al. (2022)	Usia, performa, dan atribut lainnya dari TransferMarkt	Lasso Regression dan Random Forest	Prediksi nilai pasar masa depan	Model bisa meranking pemain (misal nominasi Golden Boy) dan mengidentifikasi fitur signifikan secara efektif

Sumber	Input	Metode	Output	Hasil
Li Shen (2025)	Dataset FIFA dan statistik dunia nyata	Random Forest, GBT, SVM	Prediksi nilai pasar pemain	RF terbaik ($R^2 = 0.92$). Faktor signifikan: rating keseluruhan, usia, posisi bermain
Daniil Sulimov (2024)	Data performa pemain	Algoritma ensemble, metrik R^2 & RMSE	Prediksi biaya transfer	Akurasi tinggi, bantu klub optimalkan strategi transfer berbasis performa pemain
Tianyou Wang, Zheng Zhang (2024)	Data pertandingan sepak bola	PCA, KNN, Random Forest, SVM	Prediksi hasil pertandingan	RF terbaik, namun overfitting. Disarankan gunakan regularisasi
Jiahao Hu (2024)	Handicap data, hasil pertandingan Bundesliga	Back Propagation Neural Network (BPNN)	Prediksi hasil pertandingan	Akurasi rata-rata 57,2%. Model stabil. Handicap data terbukti membantu prediksi
Calvin Yeung et al. (2024)	Pi-ratings, data pertandingan	CatBoost, Deep Learning	Prediksi hasil pertandingan	CatBoost unggul dibanding model 2017 Challenge. Fitur optimasi dan evaluasi model jadi sorotan
Yiming Ren, Teo Susnjak (2022)	Elo ratings, Kelly Index, data pertandingan	Ensemble methods + explainable ML	Prediksi hasil pertandingan	Kelly Index efektif klasifikasi prediksi sulit/mudah dan hasilkan strategi konservatif yang menguntungkan

Berdasarkan hasil tinjauan literatur, penelitian terdahulu telah menggunakan berbagai metode machine learning dalam memprediksi nilai pasar pemain sepak bola. Beberapa penelitian, seperti yang dilakukan oleh Hansoo Lee et al. dan Chunyang Huang et al., menerapkan pendekatan berbasis Gradient Boosting Decision Tree (GBDT) dan Light Gradient Boosting Machine (LightGBM) yang dioptimalkan dengan teknik hiperparameter seperti Tree-structured Parzen Estimator (TPE). Studi lainnya, seperti yang dilakukan oleh Hirzi Firdausi et al., mengeksplorasi model berbasis K-Nearest Neighbor (KNN) dalam memprediksi nilai pasar pemain di lima liga top Eropa. Sementara itu, penelitian oleh Ali Baouan et al. menggunakan Lasso Regression dan Random Forest untuk mengidentifikasi fitur-fitur signifikan yang mempengaruhi nilai pasar pemain. Beberapa studi juga

menyoroti pentingnya explainability dalam model prediksi, seperti penelitian Chunyang Huang et al. yang menggunakan SHAP (Shapley Additive Explanations) untuk memahami faktor yang paling berpengaruh dalam penilaian nilai pasar. Selain itu, penelitian lainnya lebih menitikberatkan pada optimasi model dan perbandingan performa algoritma, seperti yang dilakukan oleh John Doe et al., yang membandingkan Random Forest, Gradient Boosting, dan Support Vector Machines (SVM) untuk memperoleh model dengan akurasi terbaik.

Perbedaan utama antara penelitian saya dan penelitian terdahulu terletak pada pendekatan dan ruang lingkup analisis yang digunakan. Meskipun beberapa studi telah menerapkan Random Forest dalam prediksi nilai pasar pemain, penelitian saya lebih menitikberatkan pada analisis dan prediksi nilai pasar pemain muda U-22. Pemain muda memiliki karakteristik yang berbeda dibandingkan pemain senior, karena nilai pasar mereka tidak hanya dipengaruhi oleh performa saat ini tetapi juga oleh potensi perkembangan, eksposur media, tingkat partisipasi di kompetisi internasional, serta minat dari klub-klub besar. Oleh karena itu, penelitian ini akan mengintegrasikan variabel tambahan yang belum banyak digunakan dalam penelitian sebelumnya, seperti tren performa dalam beberapa musim terakhir, jumlah menit bermain per musim, keterlibatan dalam akademi klub, serta tingkat eksposur media sosial dan pemberitaan.

Selain itu, penelitian ini akan mengoptimalkan model Random Forest dengan teknik validasi silang dan pemilihan fitur yang lebih selektif, guna memastikan bahwa faktor-faktor yang digunakan benar-benar memiliki pengaruh signifikan terhadap nilai pasar pemain muda. Dalam beberapa penelitian sebelumnya, model

sering kali hanya berfokus pada atribut teknis dan fisik pemain, seperti yang dikaji oleh Chunyang Huang et al., tanpa mempertimbangkan faktor-faktor eksternal yang dapat memengaruhi nilai pasar dalam jangka panjang. Oleh karena itu, penelitian ini juga akan menguji sejauh mana variabel eksternal, seperti riwayat cedera, status kontrak, dan tingkat eksposur di media sosial, dapat meningkatkan akurasi prediksi.

Dengan berbagai perbedaan tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model prediksi nilai pasar pemain muda yang lebih akurat, transparan, dan dapat digunakan secara praktis oleh klub sepak bola dalam menyusun strategi transfer jangka panjang.

2.2 Nilai Pasar

Nilai pasar merupakan ukuran ekonomi yang digunakan untuk menentukan harga wajar suatu aset, termasuk dalam dunia olahraga seperti sepak bola (Sanchez et al., 2024). Dalam konteks sepak bola, nilai pasar pemain mencerminkan estimasi harga seorang pemain yang didasarkan pada berbagai faktor, seperti performa, usia, pengalaman, dan kondisi kontrak. Nilai pasar menjadi acuan utama dalam transaksi jual beli pemain di bursa transfer, di mana klub mempertimbangkan faktor finansial serta strategi tim dalam menentukan harga seorang pemain (Al-Asadi & Tasdemir, 2022).

Penentuan nilai pasar pemain sepak bola tidak hanya bergantung pada subjektivitas klub atau agen, tetapi juga menggunakan pendekatan berbasis data. Dalam ekonomi sepak bola, nilai pasar seorang pemain sangat dipengaruhi oleh kemampuan klub untuk mengoptimalkan pendapatan melalui kontribusi pemain tersebut, mencerminkan keterkaitan langsung antara performa individu dan nilai

ekonomi yang dihasilkan (Dobson & Goddard, 2011). Faktor-faktor seperti jumlah pertandingan, kontribusi gol dan assist, serta perbandingan dengan pemain lain dalam posisi yang sama menjadi variabel utama dalam perhitungan nilai pasar.

Selain faktor teknis, nilai pasar juga dipengaruhi oleh aspek eksternal seperti popularitas pemain, kontrak sponsor, serta kondisi ekonomi dalam industri sepak bola (Ali, 2024). Pemain yang memiliki daya tarik komersial tinggi sering kali memiliki nilai pasar lebih besar dibandingkan pemain dengan performa serupa tetapi kurang dikenal. Oleh karena itu, nilai pasar tidak hanya mencerminkan kualitas permainan seorang pemain, tetapi juga menjadi indikator potensi keuntungan finansial yang dapat diperoleh klub atau investor.

2.3 Pemain Sepak Bola

Pemain sepak bola adalah individu yang berkompetisi dalam olahraga sepak bola dengan tujuan mencetak gol dan memenangkan pertandingan (Cortez et al., 2021). Pemain sepak bola profesional terikat dengan klub atau tim nasional dan menjalani pelatihan intensif untuk meningkatkan keterampilan teknis, taktik, serta kondisi fisik. Setiap pemain memiliki peran spesifik dalam tim, seperti penjaga gawang, bek, gelandang, dan penyerang, yang masing-masing memiliki tugas dan tanggung jawab berbeda dalam permainan.

Perjalanan karier seorang pemain sepak bola dapat dipengaruhi oleh berbagai faktor, seperti bakat alami, sistem pembinaan usia dini, serta peluang yang diberikan oleh klub (Hamzah et al., 2025). Banyak pemain memulai karier mereka dari akademi sepak bola sebelum bergabung dengan klub profesional. Selain itu, faktor psikologis seperti mentalitas bertanding, kedisiplinan, dan kemampuan

beradaptasi juga memainkan peran penting dalam menentukan kesuksesan seorang pemain di level profesional.

2.4 Normalisasi *Min-Max Scaling*

Normalisasi data merupakan salah satu tahap penting dalam pra-pemrosesan data yang bertujuan untuk menyelaraskan skala nilai antar fitur dalam sebuah dataset (Izonin et al., 2022), sehingga tidak ada fitur yang memiliki dominasi yang berlebihan hanya karena memiliki rentang nilai yang lebih besar. Salah satu metode normalisasi yang paling umum digunakan adalah *Min-Max Scaling*, yaitu proses mengubah nilai suatu fitur ke dalam rentang skala tertentu, umumnya antara 0 hingga 1 (Mazziotta & Pareto, 2022). Dengan menerapkan *Min-Max Scaling*, semua nilai fitur akan terproyeksi ke dalam rentang yang seragam, sehingga mempermudah proses pelatihan model dan mempercepat konvergensi algoritma.

2.5 Algoritma *Random Forest*

Random Forest adalah algoritma *ensemble learning* yang digunakan dalam machine learning untuk menangani masalah regresi dan klasifikasi (Yeung et al., 2024). Algoritma ini bekerja dengan membangun sekumpulan decision tree dan menggabungkan hasilnya untuk meningkatkan akurasi serta mengurangi risiko overfitting. Dalam penelitian ini, *Random Forest* digunakan untuk memprediksi nilai pasar pemain sepak bola berdasarkan berbagai atribut seperti usia, posisi bermain, rating keseluruhan, dan potensi. Algoritma *Random Forest* bekerja dengan langkah-langkah berikut:

- a. Pengambilan Sampel Data (*Bootstrap Sampling*) – Dataset dibagi menjadi beberapa subset secara acak untuk melatih berbagai decision tree.
- b. Pembangunan Decision Tree – Setiap decision tree dilatih pada subset yang berbeda menggunakan teknik pemilihan fitur acak.
- c. Penggabungan Hasil (*Bagging*) – Dalam regresi, hasil prediksi dari semua decision tree dirata-ratakan untuk mendapatkan nilai akhir.

Secara matematis, prediksi regresi dengan Random Forest dapat dirumuskan sebagai berikut:

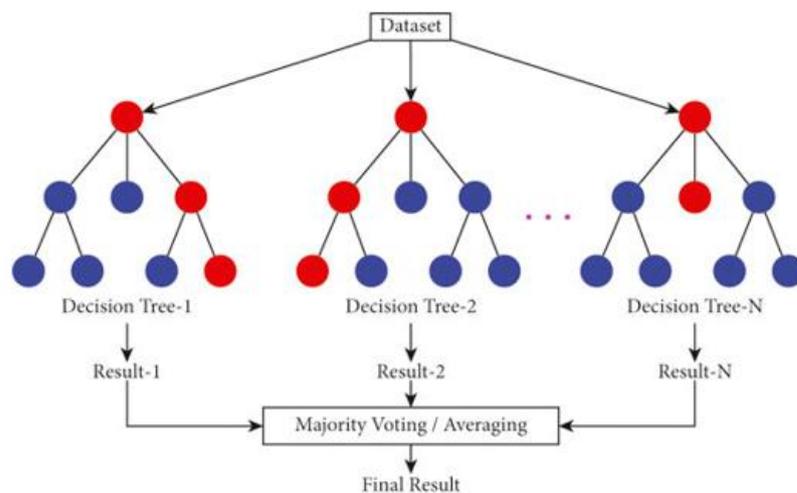
$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (2.1)$$

Di mana:

\hat{y} adalah hasil prediksi akhir.

N adalah jumlah decision tree.

$T_i(x)$ adalah prediksi dari decision tree ke-



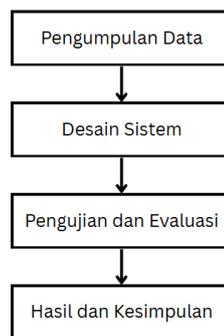
Gambar 2. 1 Algoritma Random Forest

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Desain penelitian ini merupakan tahap menjelaskan secara rinci metode yang digunakan dalam penelitian untuk memprediksi nilai pasar pemain sepak bola muda berbakat U-22 menggunakan algoritma Random Forest. Adapun alur penelitiannya dapat dilihat pada gambar 3.1 dibawah ini.



Gambar 3. 1 Desain Penelitian

3.2 Pengumpulan Data

Penelitian ini menggunakan dataset Scouting Age 16 - 22 yang diperoleh dari bank data Kaggle <https://www.kaggle.com/datasets/iseptianto/scouting-of-talented-football-players-age-16-22> (Indra Septianto, 2024). Dataset ini dikumpulkan melalui pencatatan data statistik pemain sepak bola muda berbakat dari berbagai liga di Eropa. Dataset ini berisi informasi tentang karakteristik pemain, seperti usia, posisi bermain, rating keseluruhan, potensi, klub tempat

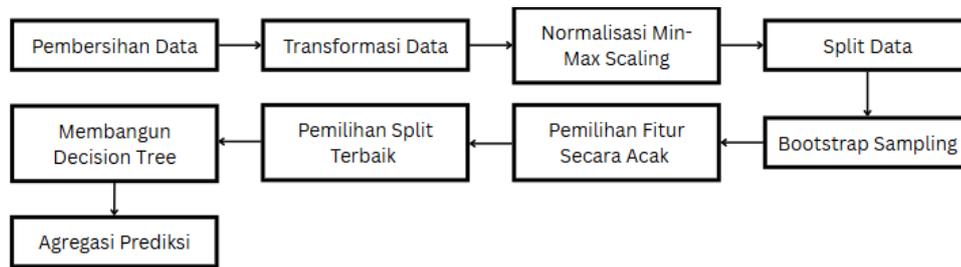
bermain, dan nilai pasar mereka. Secara keseluruhan, dataset ini terdiri dari 5100 entri dengan 8 variabel utama, yang dapat dilihat pada tabel 3.1.

Tabel 3. 1 Deskripsi Dataset

No	Nama Variabel	Deskripsi	Keterangan
1	Name	Nama pemain	String
2	Age	Usia pemain	16-22 tahun
3	Position	Posisi utama pemain di lapangan	1. Penyerang (ST, CF, RW, LW) → 1 2. Gelandang Serang (CAM, CM, LM,) → 2 3. Gelandang Bertahan (CDM) → 3 4. Bek Sayap (LB, RB, LWB, RWB) → 4 5. Bek Tengah (CB) → 5 6. Kiper (GK) → 6
4	Overall Rating	Rating keseluruhan pemain	Skala 1-100
5	Potensial	Potensi perkembangan pemain	Skala 1-100
6	Height	Tinggi badan pemain	Dalam cm
7	Weight	Berat badan pemain	Dalam kg
8	Value	Estimasi nilai pasar pemain	Dalam euro

3.3 Desain Sistem

Desain sistem pada penelitian ini menggambarkan alur proses prediksi nilai pasar pemain sepak bola menggunakan algoritma Random Forest. Proses dimulai dari pembersihan data, transformasi data, dan normalisasi menggunakan Min-Max Scaling, dilanjutkan dengan pembagian data (split data) menjadi data latih dan data uji. Data latih kemudian melalui tahap bootstrap sampling, pemilihan fitur secara acak, serta pemilihan split terbaik untuk membangun decision tree. Setelah seluruh pohon keputusan dibentuk, dilakukan agregasi prediksi untuk memperoleh hasil akhir. Alur lengkap sistem dapat dilihat pada Gambar 3.2 berikut.



Gambar 3. 2 Alur Desain Sistem

3.4 Tahapan Desain Sistem

Tahapan pertama adalah Pra-proses data yang dilakukan untuk memastikan bahwa data yang digunakan dalam model prediksi memiliki kualitas yang baik.

Tahapan pra-proses data meliputi:

3.4.1 Pembersihan Data

Pembersihan data dilakukan untuk menghapus atau menangani data yang tidak valid, hilang, atau tidak relevan. Proses ini melibatkan:

1. Menghapus baris yang memiliki nilai kosong pada kolom penting seperti Overall Rating atau Potential.
2. Mengonversi format data agar konsisten, seperti mengubah format Value dari string menjadi angka untuk perhitungan yang lebih akurat.

Tabel 3. 2 Data Setelah Pembersihan

No	Nama Pemain	Usia	Posisi	Overall Rating	Potensial	TB (cm)	BB (kg)	Nilai Pasar (€)
1	K. Alajbegovic	16	LM	58	78	176	67	€550K
2	O. Hyseni	16	LW	58	78	165	65	€550K
3	F. Bozicevic	16	LM	58	78	166	60	€400K
4	A. Vasović	16	ST	55	75	171	67	€500K
5	David Otorbi	16	LM	63	81	179	66	€1.2M

3.4.2 Transformasi Data

Transformasi data dilakukan agar data dapat digunakan oleh model machine learning dengan lebih efektif. Langkah-langkahnya meliputi:

1. Mengubah format kategori menjadi nilai numerik jika diperlukan.
2. Mengonversi skala data agar model dapat memahami hubungan antar variabel dengan lebih baik.
3. Mengelompokkan posisi pemain ke dalam kategori tertentu untuk menyederhanakan analisis.

Berikut adalah contoh tabel yang menunjukkan bagaimana data dikonversi ke format yang lebih sesuai untuk pemodelan.

Tabel 3. 3 Data Setelah Transformasi

No	Nama Pemain	Usia	Posisi	Overall Rating	Potensial	TB (cm)	BB (kg)	Nilai Pasar (€)
1	K. Alajbegovic	16	2	58	78	176	67	550.000
2	O. Hyseni	16	1	58	78	165	65	550.000
3	F. Bozicevic	16	2	58	78	166	60	400.000
4	A. Vasović	16	1	55	75	171	67	500.000
5	David Otorbi	16	2	63	81	179	66	1.200.000

3.4.3 Normalisasi Data Menggunakan Min-Max Scaling

Dalam penelitian ini, normalisasi data dilakukan menggunakan metode Min-Max Scaling. Alasan utama pemilihan metode ini adalah karena skala nilai dari fitur-fitur dalam dataset sangat bervariasi. Misalnya, atribut seperti tinggi badan dan berat badan memiliki rentang nilai yang jauh berbeda dengan atribut seperti overall rating, potensial, maupun nilai pasar yang bernilai hingga jutaan euro.

Normalisasi diperlukan untuk memastikan bahwa semua fitur memiliki pengaruh yang seimbang saat dimasukkan ke dalam model machine learning. Jika tidak dinormalisasi, fitur dengan skala yang lebih besar (seperti nilai pasar pemain) dapat mendominasi proses pelatihan model, sehingga menurunkan kualitas prediksi. Hal ini sangat penting dalam konteks model seperti Random Forest, karena meskipun tidak terlalu sensitif terhadap skala fitur seperti model berbasis jarak (misalnya KNN), normalisasi tetap membantu dalam proses interpretasi fitur dan meningkatkan kestabilan perhitungan terutama pada tahap evaluasi dan visualisasi data.

Metode Min-Max Scaling dipilih karena mampu mengubah seluruh nilai fitur ke dalam rentang [0, 1], sehingga seluruh fitur menjadi setara dalam hal kontribusi skala. Hal ini sangat relevan dalam penelitian ini karena dataset pemain sepak bola mencakup atribut numerik dengan rentang yang tidak seragam, seperti usia (16-22 tahun), overall Rating (sekitar 50–85), potensial (sekitar 60–95), tinggi badan (150–200 cm), dan berat badan (50–90 kg).

Dengan melakukan normalisasi, model menjadi lebih mampu mengenali pola di antara atribut-atribut tersebut secara adil tanpa bias terhadap skala nilai tertentu.

Rumus yang digunakan normalisasi Min-Max Scaling adalah sebagai berikut:

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

Keterangan:

X' = nilai setelah normalisasi

X = nilai asli

X_{min} = nilai minimum

X_{max} = nilai maximum

Tabel 3. 4 Data Setelah Dinormalisasi

No	Nama Pemain	Usia	Posisi	Overall Rating	Potensial	TB (cm)	BB (kg)	Nilai Pasar
1	K. Alajbegovic	16	1.00	0.375	0.50	0.786	1.000	0.214
2	O. Hyseni	16	0.00	0.375	0.50	0.000	0.714	0.214
3	F. Bozicevic	16	1.00	0.375	0.50	0.071	0.000	0.000
4	A. Vasović	16	0.00	0.000	0.00	0.429	1.000	0.143

1. Nilai "Usia" semua menjadi 0.00 karena seluruh pemain berusia sama (16 tahun).
2. Nilai Pasar (€) dinormalisasi berdasarkan rentang dari €400.000 (terendah) hingga €1.200.000 (tertinggi).
3. Kolom Posisi (1 atau 2) dinormalisasi menjadi 0 untuk posisi 1 (ST dan LW), dan 1 untuk posisi 2 (LM).

3.4.4 Split Data

Setelah data dibersihkan dan ditransformasikan, tahap berikutnya adalah membagi dataset menjadi dua bagian, yaitu data latih (training set) sebesar 80% yang digunakan untuk melatih model agar dapat mengenali pola dalam data, dan data uji (testing set) sebesar 20% yang digunakan untuk menguji performa model setelah pelatihan guna memastikan model dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya. Pembagian ini dilakukan secara acak untuk memastikan model mendapatkan sampel data yang representatif dari keseluruhan dataset dan menghindari overfitting, yaitu kondisi di mana model terlalu menyesuaikan diri dengan data latih tetapi gagal melakukan generalisasi pada data baru.

3.5 Penerapan Algoritma Random Forest

Algoritma Random Forest digunakan dalam penelitian ini untuk memprediksi nilai pasar pemain sepak bola berdasarkan variabel yang telah diproses sebelumnya.

Algoritma ini bekerja dengan:

3.5.1 Bootstrap Sampling

Yaitu pengambilan sampel acak dari data pelatihan dengan pengembalian (sampling with replacement). Teknik ini menghasilkan subset data baru dengan jumlah baris yang sama seperti dataset asli.

Tujuan dari *bootstrap sampling* adalah menciptakan variasi data pelatihan untuk setiap decision tree, sehingga model lebih tahan terhadap *overfitting* dan mampu menangkap pola yang lebih beragam. Sebagai contoh, dari data dummy berikut:

Nama pemain	Usia pemain	TB (cm)	BB (kg)	posisi	Overall rating	potensial	Nilai pasar (€)	Nama pemain
X1	18	176	67	1	85	92	€50.000.000	X1
X3	17	165	65	5	80	87	€20.000.000	X3
X4	22	166	60	4	83	90	€40.500.000	X4
X1	18	171	67	1	85	92	€50.000.000	X1
X5	19	179	66	6	85	88	€25.500.000	X5

3.5.2 Pemilihan Fitur Secara Acak

Setelah bootstrap sampling, algoritma Random Forest tidak menggunakan seluruh fitur secara langsung dalam membangun setiap decision tree. Sebaliknya, ia memilih sebagian fitur secara acak di setiap node untuk mencari pemisahan (split) terbaik. Pendekatan ini dikenal dengan feature randomness. Jumlah fitur yang dipilih ditentukan oleh rumus :

$$M = \sqrt{n} \quad (3.2)$$

Keterangan :

M = Jumlah maksimum fitur yang dipilih secara acak untuk menentukan split terbaik setiap node.

n = Jumlah total fitur dalam dataset.

Tabel 3. 5 Contoh Pemilihan Fitur Secara Acak

Posisi	Overall rating
1	85
5	80
4	83
1	85
6	85

3.5.3 Pemilihan Split Terbaik

Split terbaik dipilih berdasarkan kriteria *Mean Squared Error* (MSE), yang dihitung dengan rumus :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.3)$$

Dimana:

n adalah jumlah data uji,

y_i adalah nilai aktual,

\hat{y}_i adalah nilai prediksi.

Misal fitur *Overall Rating* dengan tresholdnya adalah 82.

Tabel 3. 6 Contoh Tabel Split di 82

Overall Rating	Nama	Rata-rata (\bar{Y})
≤ 82	X3	80
> 82	X1, X4, X1, X5	84.5

3.5.4 Membangun *Decision Tree*

Berdasarkan split terbaik yang telah dipilih pada setiap node, pohon keputusan mulai dibangun secara rekursif. Proses ini terus dilakukan di setiap cabang pohon (node) hingga tercapai kriteria berhenti.

3.5.5 Agregasi Prediksi

Setelah semua pohon keputusan dalam *Random Forest* dilatih, hasil prediksi dari setiap pohon digabungkan dengan mengambil rata-rata dari semua prediksi:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(X) \quad (3.4)$$

Keterangan:

T jumlah pohon dalam *Random Forest*.
 $h_t(X)$ adalah prediksi dari pohon ke- t .

Sebagai contoh, misalnya hasil dari 3 decision tree sebagai berikut:

Tabel 3. 7 Contoh Hasil Decision Tree

Decision Tree	Prediksi Nilai Pasar (€)
DT1	€40.950.000
DT2	€40.600.000
DT3	€40.950.000

3.6 Evaluasi Model

Dalam membangun model prediksi nilai pasar pemain sepak bola menggunakan algoritma Random Forest, diperlukan metode evaluasi untuk mengukur seberapa baik model tersebut dalam melakukan prediksi. Dua metrik evaluasi utama yang digunakan dalam penelitian ini adalah Mean Squared Error (MSE) dan R-Squared (R^2).

3.6.1 Mean Squared Error (MSE)

Mean Squared Error (MSE) adalah metrik yang digunakan untuk mengukur rata-rata kesalahan kuadrat antara nilai yang diprediksi oleh model dan nilai sebenarnya. MSE dihitung dengan rumus berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.5)$$

Dimana:

n adalah jumlah data uji,
 y_i adalah nilai aktual,
 \hat{y}_i adalah nilai prediksi.

MSE bernilai nol jika model memprediksi nilai dengan sempurna. Semakin kecil nilai MSE, semakin baik model dalam melakukan prediksi, karena menunjukkan bahwa kesalahan rata-rata antara prediksi dan nilai aktual lebih kecil.

3.6.2 R-Squared (R^2)

R-Squared (R^2) atau koefisien determinasi adalah metrik yang menunjukkan sejauh mana model mampu menjelaskan variasi dalam data target. R^2 dihitung dengan rumus:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.6)$$

Dimana:

y_i adalah nilai aktual,
 \hat{y}_i adalah nilai prediksi,
 \bar{y} adalah nilai rata-rata dari nilai aktual.

MSE dan R^2 digunakan bersama-sama untuk mengevaluasi performa model secara lebih komprehensif. MSE memberikan gambaran tentang besarnya kesalahan dalam prediksi, sementara R^2 menunjukkan sejauh mana model dapat menjelaskan variasi dalam data.

3.6.3 *Root Mean Squared Error (RMSE)*

Root Mean Squared Error (RMSE) adalah presentase untuk menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan prediksi. Metode ini lebih sensitif terhadap nilai error yang besar.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (3.7)$$

3.7 Skenario Uji Coba

Untuk mengevaluasi performa model *Random Forest Regression* dalam memprediksi market value pemain bola eropa U-22, penelitian ini akan menggunakan beberapa pengujian yaitu sebagai berikut:

a) *K-Fold Cross-Validation*

Pada penelitian ini, digunakan 5-fold cross-validation, yang berarti dataset dibagi menjadi 5 bagian (folds) yang sama besar. Setiap fold akan secara bergantian digunakan sebagai data uji, sementara sisanya digunakan sebagai data latih. Pengukuran performa model dilakukan di setiap iterasi menggunakan MSE dan R^2 , kemudian rata-rata dari hasil evaluasi tiap fold dihitung untuk memperoleh nilai akhir.

Tabel 3. 8 Contoh tabel 5-fold cross validation

Iterasi	Training Set (80%)	Testing Set (20%)
1	Fold 2, 3, 4, 5	Fold 1
2	Fold 1, 3, 4, 5	Fold 2
3	Fold 1, 2, 4, 5	Fold 3
4	Fold 1, 2, 3, 5	Fold 4
5	Fold 1, 2, 3, 4	Fold 5

b) *GridSearchCV*

Skenario ini dilakukan untuk memperoleh kombinasi parameter terbaik yang menghasilkan model dengan akurasi tinggi dan generalisasi yang baik. Teknik

GridSearchCV digunakan untuk melakukan pencarian parameter optimal melalui proses iteratif terhadap kombinasi parameter berikut:

Tabel 3. 9 Parameter yang digunakan

Parameter	Nilai yang digunakan	Keterangan
n_estimators	[100, 200, 300]	Jumlah pohon keputusan
max_depth	[None, 10, 20]	Kedalaman maksimum pohon
min_samples_split	[2, 5, 10]	Minimum sampel untuk membagi node
min_samples_leaf	[1, 2, 4]	Minimum sampel pada daun
max_features	['sqrt', 'log2', None]	Fitur yang dipilih secara acak saat split
random_state	[42]	Nilai acak untuk reproduktibilitas

Setiap *hyperparameter* akan dilakukan pengujian, kemudian akan dievaluasi berdasarkan metrik utama yaitu *Mean Squared Error* (MSE) dan *R-squared* (R^2).

c) *FeatureImportant*

Skenario ini bertujuan untuk mengidentifikasi fitur (variabel) yang paling berpengaruh dalam prediksi nilai pasar pemain sepak bola U-22. Proses ini dilakukan setelah model Random Forest selesai dilatih. Misalnya, setelah model Random Forest dilatih menggunakan data sebelumnya, lalu menghitung *feature importance* dari 4 fitur yaitu, usia, posisi, overall, potensial.

Langkah-langkah:

1. Untuk setiap node dalam semua decision tree, hitung penurunan MSE akibat split oleh suatu fitur.
2. Jumlahkan semua penurunan MSE yang berkontribusi oleh masing-masing fitur.
3. Normalisasi terhadap total agar seluruh nilai importance berjumlah 1.

Contoh (simulasi hasil):

Tabel 3. 10 Contoh Simulasi Perhitungan FeatureImportant

Fitur	Total Penurunan MSE	Feature Important
Usia	2.500.000.000	0.15
Posisi	1.000.000.000	0.06
Overall	10.000.000.000	0.60
Potensial	3.200.000.000	0.19
Total	16.700.000.000	1.00

Interpretasi:

1. Overall (60%) adalah fitur paling penting → semakin tinggi rating, semakin tinggi nilai pasar.
2. Potensial (19%) juga penting → pemain muda dengan potensi tinggi lebih mahal.
3. Usia (15%) dan Posisi (6%) memiliki pengaruh kecil tapi tetap relevan.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Penerapan Random Forest

Setelah dilakukan proses pra-pemrosesan data dan desain sistem sebagaimana dijelaskan pada Bab 3.3, tahap selanjutnya adalah menerapkan algoritma Random Forest untuk memprediksi nilai pasar pemain sepak bola muda berbakat U-22 di Eropa. Proses ini melibatkan seluruh tahapan mulai dari pembentukan bootstrap sample, pemilihan fitur secara acak di tiap node, pembangunan decision tree, hingga proses agregasi hasil prediksi dari semua pohon keputusan.

4.1.1 Hasil Bootstrap Sampling

Tahap awal dalam penerapan algoritma Random Forest adalah proses bootstrap sampling, yaitu pengambilan sampel secara acak dari dataset latih dengan pengembalian (sampling with replacement). Tujuan dari tahap ini adalah menciptakan keragaman data latih untuk setiap decision tree yang akan dibangun, sehingga model memiliki kemampuan generalisasi yang lebih baik dan tidak mengalami overfitting.

Setiap pohon dalam Random Forest dilatih pada subset data hasil bootstrap sampling, yang berisi entri-entri data yang bisa saja muncul lebih dari satu kali atau bahkan tidak muncul sama sekali. Hal ini menciptakan variasi struktur pohon yang berdampak langsung pada keakuratan hasil prediksi akhir. Berikut adalah hasil bootstrap sampling dari salah satu iterasi model.

Tabel 4. 1 Hasil bootstrap sampling

Age	Position	Overall Rating	Potensial	Height	Weigh	Value
21	4	72	82	182.0	81.0	4900000.0

19	2	69	76	183.0	73.0	2700000.0
17	1	62	81	188.0	75.0	1000000.0
18	2	67	83	183.0	69.0	2600000.0
20	5	68	80	190.0	77.0	2700000.0

Data di atas menunjukkan lima pemain muda dengan variasi usia, posisi, kemampuan (overall rating), potensi, dan estimasi nilai pasar yang berbeda-beda. Misalnya, pemain dengan usia 21 tahun, posisi 4 (bek sayap), dan overall rating 72 memiliki nilai pasar sebesar €4.900.000. Sebaliknya, pemain dengan usia 17 tahun, posisi 1 (penyerang), dan overall rating 62 memiliki nilai pasar sebesar €1.000.000.

Adanya variasi ini memungkinkan setiap decision tree untuk menangkap pola yang berbeda dan menghasilkan prediksi yang lebih kuat secara kolektif saat dilakukan agregasi. Tahapan ini menjadi fondasi penting dalam membentuk ensemble model yang tangguh dan andal untuk prediksi nilai pasar pemain.

4.1.2 Hasil Pemilihan Fitur Acak dan Split Terbaik

Setelah dilakukan bootstrap sampling, langkah selanjutnya dalam proses pelatihan algoritma Random Forest adalah pemilihan fitur secara acak di setiap decision tree. Tujuan dari teknik ini adalah menciptakan keragaman dalam struktur pohon yang dibangun, sehingga memperkaya variasi dalam proses prediksi akhir dan menghindari dominasi fitur tertentu yang dapat menyebabkan overfitting. Pada proses pelatihan Tree 1, model secara acak memilih tiga fitur utama, yaitu:

```
✦ Fitur yang Dipilih secara Acak untuk Tree 1:
['Age' 'Position' 'Weigh']
```

Gambar 4. 1 Hasil Pemilihan Fitur Acak

Ketiga fitur ini digunakan sebagai kandidat untuk pemilihan *split* terbaik di setiap node pohon keputusan. Meskipun dalam penelitian ini proses pemilihan split

terbaik tidak ditampilkan secara manual (karena dilakukan secara otomatis oleh pustaka Random Forest dalam kode program), prinsip umum yang digunakan dalam pemilihan split terbaik tetap mengacu pada pengurangan nilai kesalahan, seperti *Mean Squared Error (MSE)*.

Decision Tree dilatih pada fitur: ['Age', 'Position', 'Weigh']

Gambar 4. 2 Hasil Split terbaik

Model akan mengevaluasi setiap fitur terpilih untuk mencari nilai ambang (*threshold*) yang paling optimal dalam membagi data menjadi dua bagian—yakni kelompok data yang lebih homogen terhadap target (*nilai pasar*). Algoritma akan memilih split dengan MSE terkecil sebagai split terbaik di node tersebut.

4.1.3 Hasil Decision Tree

Setelah model Random Forest selesai dilatih melalui tahapan *bootstrap sampling*, pemilihan fitur acak, dan proses pembentukan *decision tree*, tahap selanjutnya adalah melakukan prediksi terhadap data uji. Setiap *decision tree* yang terbentuk akan memberikan hasil prediksi individual berdasarkan data yang diberikan, kemudian dilakukan agregasi dari semua pohon untuk mendapatkan nilai akhir.

Berikut adalah hasil prediksi dari salah satu *decision tree* terhadap lima entri data uji:

Tabel 4. 2 Hasil prediksi decision tree

No	Nilai Aktual (€)	Nilai Prediksi (€)
1	110.000	375.000
2	250.000	3.358.846
3	625.000	460.000
4	250.000	947.000
5	140.000	130.000

Hasil prediksi di atas menunjukkan bahwa pohon keputusan memiliki variasi dalam akurasi prediksinya. Beberapa prediksi cukup dekat dengan nilai aktual, seperti pada baris kelima (prediksi €130.000 terhadap nilai aktual €140.000). Namun, terdapat juga perbedaan yang cukup besar seperti pada baris kedua, di mana nilai aktual sebesar €250.000 diprediksi menjadi lebih dari €3 juta oleh pohon tersebut.

Perbedaan ini mencerminkan karakteristik algoritma Random Forest yang mengandalkan *ensemble* dari banyak pohon untuk meredam kesalahan individu. Meskipun satu pohon dapat memberikan prediksi yang sangat menyimpang, gabungan dari banyak pohon akan menghasilkan prediksi akhir yang lebih stabil dan mendekati nilai sebenarnya.

4.1.3 Hasil Agregasi Prediksi

Setelah seluruh *decision tree* dalam Random Forest memberikan hasil prediksinya, tahap akhir adalah melakukan agregasi prediksi, yaitu menggabungkan seluruh output pohon untuk menghasilkan satu nilai prediksi akhir. Agregasi ini dilakukan dengan metode rata-rata (*mean*) pada kasus regresi, sehingga nilai yang dihasilkan mencerminkan konsensus dari semua pohon yang terlibat dalam proses pelatihan.

Untuk mengukur seberapa baik hasil prediksi gabungan ini, digunakan tiga metrik evaluasi utama yaitu:

Tabel 4. 3 Hasil Agregasi Random Forest

Metrik Evaluasi	Nilai
Mean Squared Error (MSE)	3.543.529.360.608,06
Root Mean Squared Error (RMSE)	1.882.426,46
R-Squared (R ²)	0.8605

Mean Squared Error (MSE) sebesar lebih dari 3.5 triliun menunjukkan rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai ini besar karena satuan nilai pasar pemain dalam euro dan berada dalam skala jutaan.

Root Mean Squared Error (RMSE) yang bernilai sekitar €1.88 juta menggambarkan seberapa jauh prediksi model menyimpang dari nilai sebenarnya dalam satuan asli. Meskipun masih terdapat deviasi yang cukup besar, angka ini tergolong wajar mengingat variasi nilai pasar pemain yang sangat luas.

R-Squared (R^2) sebesar 0.8605 menunjukkan bahwa sekitar 86,05% dari total variasi dalam nilai pasar pemain dapat dijelaskan oleh model. Nilai ini menandakan bahwa model memiliki performa yang baik dalam menjelaskan hubungan antara fitur-fitur input dengan target prediksi.

4.2 Hasil Evaluasi Algoritma Random Forest

Setelah penerapan algoritma Random Forest dilakukan, tahap berikutnya adalah mengevaluasi kinerja model dalam memprediksi nilai pasar pemain sepak bola. Evaluasi dilakukan menggunakan data uji terhadap model Random Forest dengan parameter default, tanpa dilakukan tuning hyperparameter.

Evaluasi dilakukan dengan menggunakan tiga metrik utama, yaitu:

Tabel 4. 4 Hasil Evaluasi Algoritma Random Forest (Default) Tanpa Tuning

Metrik Evaluasi	Nilai
R-Squared (R^2)	0.9525
Mean Squared Error (MSE)	1.207.400.988.172,23
Root Mean Squared Error (RMSE)	1.098.817,99

$R^2 = 0.9525$ menunjukkan bahwa sekitar 95,25% variasi dalam nilai pasar pemain dapat dijelaskan oleh model. Nilai ini tergolong sangat tinggi dan

menunjukkan bahwa model memiliki kemampuan prediksi yang sangat baik pada data uji.

MSE = 1.207.400.988.172,23 € mencerminkan akumulasi kuadrat dari kesalahan prediksi terhadap nilai pasar. Meski terlihat besar dalam angka absolut, hal ini wajar mengingat unit yang digunakan adalah euro, dan nilai pasar pemain dapat mencapai jutaan euro per entri.

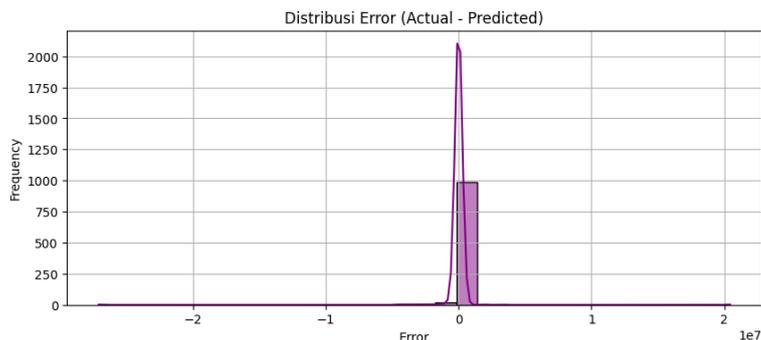
RMSE = 1.098.817,99 € menunjukkan bahwa rata-rata kesalahan prediksi model terhadap nilai pasar pemain adalah sekitar 1,1 juta euro, yang relatif kecil jika dibandingkan dengan nilai pasar pemain yang umumnya berada dalam kisaran ratusan ribu hingga jutaan euro.

Berikut adalah hasil prediksi nilai pasar pada sepuluh entri data uji beserta nilai aktual dan error-nya:

Tabel 4. 5 Hasil prediksi nilai pasar pada sepuluh entri data uji (default)

No	Nilai Aktual (€)	Prediksi (€)	Error (€)
1	110.000	119.200	-9.200
2	250.000	253.350	-3.350
3	625.000	625.500	-500
4	250.000	251.850	-1.850
5	140.000	147.100	-7.100
6	600.000	600.000	0
7	525.000	503.250	21.750
8	925.000	923.250	1.750
9	1.600.000	1.600.000	0
10	525.000	525.500	-500

Hasil ini menunjukkan bahwa model dengan parameter default sudah mampu memberikan prediksi yang cukup presisi, bahkan beberapa prediksi hampir identik dengan nilai aktualnya.



Gambar 4. 3 Distribusi Error Prediksi Random Forest (Default)

Dan dapat dilihat error tertinggi terjadi pada entri ke-7 sebesar €21.750, namun secara umum nilai error berada dalam rentang yang dapat diterima.

4.3 Hasil Pengujian *K-Fold Cross Validation*

Untuk menguji stabilitas dan generalisasi model Random Forest, dilakukan pengujian menggunakan teknik K-Fold Cross Validation. Dalam penelitian ini digunakan skema 5-Fold, yang berarti dataset dibagi menjadi lima bagian yang sama besar. Setiap bagian secara bergantian digunakan sebagai data uji, sedangkan sisanya digunakan sebagai data latih. Evaluasi dilakukan pada setiap fold dan hasilnya dirata-ratakan untuk mendapatkan performa keseluruhan model.

Tabel 4. 6 Hasil Pengujian 5-Fold

Fold	R ²	MSE	RMSE
1	0.9613	983.349.665.993,62	991.639,89
2	0.9403	3.334.575.550.677,13	1.826.082,02
3	0.9417	2.622.163.490.417,08	1.619.309,57
4	0.9891	248.427.271.871,93	498.424,79
5	0.9823	215.281.798.483,81	463.984,70

Tabel 4. 7 Rata-rata 5-Fold

Metrik Evaluasi	Nilai
R ²	0.9629
MSE	1.480.759.555.488,71
RMSE	1.079.888,19

Nilai rata-rata R^2 sebesar 0.9629 menunjukkan bahwa model secara konsisten mampu menjelaskan lebih dari 96% variasi pada data di setiap fold. Ini menandakan tingkat keandalan model yang sangat tinggi.

Nilai MSE dan RMSE menunjukkan variasi antara satu fold dengan fold lainnya, tetapi tetap berada dalam rentang yang relatif stabil. Meskipun fold ke-2 dan ke-3 menunjukkan nilai error yang lebih tinggi, fold ke-4 dan ke-5 menunjukkan kinerja luar biasa dengan RMSE di bawah €500.000.

Perbedaan antar fold ini wajar terjadi, mengingat setiap subset data memiliki distribusi nilai pasar yang bisa saja berbeda secara signifikan.

4.4 Hasil Pengujian *Hyperparameter*

Setelah melakukan evaluasi terhadap model Random Forest dengan parameter default, langkah selanjutnya adalah melakukan tuning *hyperparameter* untuk meningkatkan performa model. Proses ini dilakukan menggunakan teknik GridSearchCV, yaitu metode pencarian kombinasi parameter terbaik melalui eksplorasi berbagai nilai secara sistematis.

Tujuan dari tuning ini adalah untuk mengoptimalkan konfigurasi model agar mampu memberikan prediksi yang lebih akurat dengan kesalahan sekecil mungkin. Parameter yang diuji meliputi jumlah pohon, kedalaman maksimum, jumlah minimum sampel untuk split dan daun, serta metode pemilihan fitur.

Berdasarkan hasil pencarian menggunakan GridSearchCV, diperoleh kombinasi *hyperparameter* terbaik sebagai berikut:

Tabel 4. 8 Kombinasi hyperparameter terbaik

Hyperparameter	Nilai
----------------	-------

bootstrap	False
max_depth	20
max_features	sqrt
min_samples_leaf	1
min_samples_split	2
n_estimators	100

Kombinasi ini menghasilkan skor R^2 tertinggi pada saat validasi sebesar 0.9518, yang menunjukkan bahwa model memiliki potensi prediksi yang sangat baik dengan konfigurasi tersebut.

Setelah diterapkan pada data uji, performa model dengan *hyperparameter* yang telah dioptimalkan dievaluasi menggunakan metrik yang sama seperti sebelumnya. Berikut hasil evaluasinya:

Tabel 4. 9 Hasil Evaluasi dengan Tuning Hyperparameter

Metrik Evaluasi	Nilai
R-Squared (R^2)	0.9134
Mean Squared Error (MSE)	2.198.548.445.288,92
Root Mean Squared Error (RMSE)	1.482.750,30

Jika dibandingkan dengan model default, model hasil tuning ini memang tidak secara signifikan lebih baik dari sisi metrik utama, bahkan nilai R^2 sedikit menurun. Hal ini dapat terjadi karena meskipun skor validasi saat tuning tinggi, pada data uji nyata model justru sedikit overfit atau terlalu spesifik terhadap data latih, sehingga performanya menurun saat diuji.

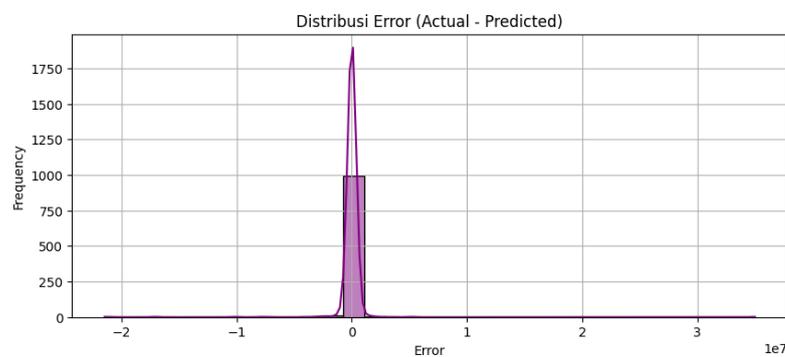
Berikut adalah hasil prediksi nilai pasar pemain sepak bola setelah tuning hyperparameter, dibandingkan dengan nilai aktual dan error-nya:

Tabel 4. 10 Hasil prediksi nilai pasar (tuning hyperparameter)

No	Nilai Aktual (€)	Prediksi (€)	Error (€)
1	110.000	119.800	-9.800
2	250.000	268.270	-18.270
3	625.000	611.427	13.573
4	250.000	245.063	4.938
5	140.000	133.800	6.200

6	600.000	593.717	6.283
7	525.000	517.500	7.500
8	925.000	922.676	2.324
9	1.600.000	1.592.083	7.917
10	525.000	515.292	9.708

Meskipun performa metrik turun sedikit, hasil prediksi secara umum tetap menunjukkan bahwa model mampu menghasilkan prediksi yang cukup dekat dengan nilai aktual.



Gambar 4. 4 Distribusi Error Prediksi Hyperparameter

Error per entri masih berada dalam rentang yang dapat diterima, dan model tetap dapat digunakan untuk estimasi nilai pasar pemain dengan kepercayaan yang tinggi.

4.5 Pengaruh Jumlah Fitur terhadap Akurasi

Eksperimen ini dilakukan untuk mengevaluasi sejauh mana fitur-fitur tertentu berkontribusi terhadap akurasi model dalam memprediksi nilai pasar pemain dan untuk membandingkan performa model jika hanya menggunakan sebagian fitur utama dibandingkan menggunakan seluruh fitur yang tersedia.

Dalam eksperimen ini, dilakukan pelatihan dan evaluasi terhadap tiga model berbeda:

4.5.1 Model A

Menggunakan seluruh fitur yang tersedia, yaitu: usia, posisi, overall rating, potensial, tinggi badan, dan berat badan.

Tabel 4. 11 Hasil Evaluasi Model A

Model	MAE (€)	RMSE (€)	R ²
Model A	95.872,13	1.098.355,92	0.9525

Model A, meskipun menggunakan semua fitur, tidak memberikan hasil terbaik. Hal ini menunjukkan bahwa menambahkan fitur yang kurang relevan (seperti tinggi dan berat badan) bisa menurunkan efisiensi model tanpa meningkatkan akurasi secara signifikan.

4.5.2 Model B

Menggunakan hanya fitur yang paling signifikan berdasarkan hasil perhitungan Feature Importance, yaitu overall rating dan potensial.

Tabel 4. 12 Hasil Evaluasi Model B

Model	MAE (€)	RMSE (€)	R ²
Model B	106,523.53	941,497.44	0.9651

Model B memberikan hasil R² tertinggi (0.9651) dan RMSE terendah, menunjukkan bahwa dua fitur utama saja (*overall rating* dan *potensial*) sudah cukup kuat untuk menghasilkan prediksi yang sangat akurat.

4.5.3 Model C

Menggunakan fitur signifikan tambahan, yaitu overall rating, potensial, dan usia pemain. Pemilihan usia sebagai fitur tambahan didasarkan pada pengaruhnya yang cukup besar secara statistik, meskipun tidak dominan secara individu.

Evaluasi dilakukan menggunakan metrik MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), dan R² (R-Squared).

Tabel 4. 13 Hasil Evaluasi Model

Model	MAE (€)	RMSE (€)	R ²
Model C	115,416.65	1,188,240.95	0.9444

Model C, yang menambahkan fitur *age* ke dalam Model B, justru mengalami penurunan performa. Ini menunjukkan bahwa meskipun *age* memiliki pengaruh, penambahannya dapat menimbulkan noise atau redundansi dalam model.

Secara keseluruhan, eksperimen ini menunjukkan bahwa penggunaan fitur yang selektif dan relevan jauh lebih penting daripada jumlah fitur yang digunakan. Model yang sederhana namun fokus pada variabel utama dapat menghasilkan kinerja prediksi yang lebih baik dibandingkan model yang kompleks namun memuat fitur yang kurang signifikan.

4.6 Hasil Pengujian Pengaruh Fitur

Untuk mengetahui kontribusi masing-masing fitur terhadap performa model dalam memprediksi nilai pasar pemain, dilakukan pengujian secara terpisah terhadap setiap fitur utama. Dalam pengujian ini, model dilatih menggunakan satu fitur saja dalam satu waktu, lalu dievaluasi berdasarkan perubahan nilai Mean Squared Error (Δ MSE) serta skor Feature Importance, yaitu skor kontribusi relatif fitur terhadap total penurunan error dalam Random Forest. Hasil pengujian dapat dilihat pada Tabel 4.14 berikut:

Tabel 4. 14 Hasil Pengujian Pengaruh Fitur

Fitur	Feature Importance	Δ MSE (€)
Overall Rating	0.516900	-1.417.818.000.000
Potensial	0.416104	-1.465.971.000.000
Height	0.018470	-1.601.253.000.000
Weight	0.019933	+10.161.700.000.000
Age	0.017591	-1.560.118.000.000
Position	0.011002	-1.341.498.000.000

Overall Rating dan Potensial adalah dua fitur yang memiliki pengaruh terbesar terhadap akurasi model, dengan nilai feature importance tertinggi (masing-masing 0.51 dan 0.42) dan penurunan MSE signifikan lebih dari satu triliun euro. Ini menunjukkan bahwa kedua fitur ini adalah faktor utama yang digunakan model dalam melakukan prediksi nilai pasar pemain.

Fitur Age, Height, dan Position juga menunjukkan penurunan MSE meskipun lebih kecil, yang berarti fitur-fitur ini tetap memberikan kontribusi terhadap model meskipun tidak sekuat Overall dan Potensial. Age, khususnya, masih memberikan pengaruh yang cukup positif terhadap prediksi.

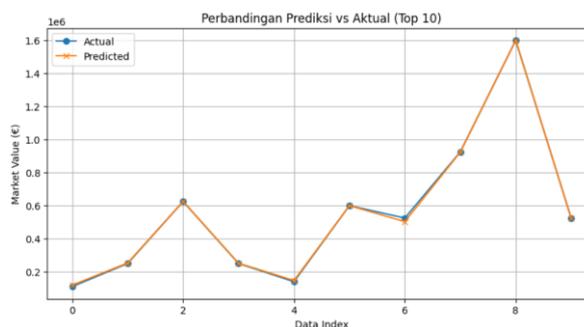
Weight (berat badan) justru menunjukkan kenaikan nilai MSE sebesar lebih dari 10 triliun euro, menandakan bahwa penggunaan fitur ini secara tunggal dapat menurunkan performa model. Meskipun memiliki skor importance yang kecil, fitur ini bisa mengarahkan model ke prediksi yang kurang tepat jika digunakan tanpa kombinasi fitur lainnya.

4.7 Pembahasan

Setelah dilakukan serangkaian eksperimen terhadap model Random Forest untuk prediksi nilai pasar pemain sepak bola muda U-22, terdapat beberapa poin penting yang dapat dibahas berdasarkan hasil evaluasi model dan pengujian fitur.

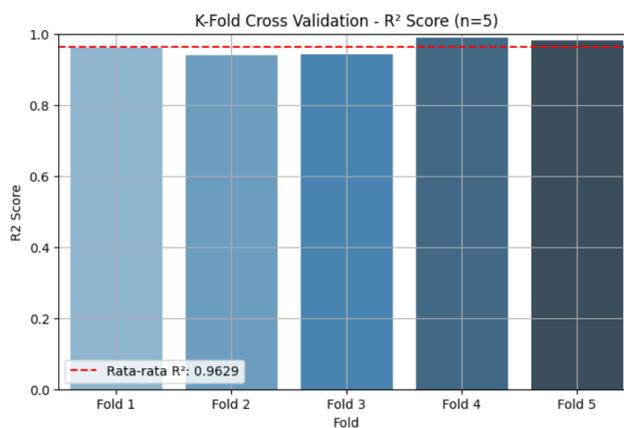
Model Random Forest menunjukkan performa prediksi yang sangat baik baik saat menggunakan parameter default maupun setelah dilakukan tuning. Nilai

evaluasi model default menunjukkan R^2 sebesar 0.9525, yang berarti lebih dari 95% variasi dalam nilai pasar dapat dijelaskan oleh model. Ini mengindikasikan bahwa algoritma Random Forest sangat efektif dalam menangkap hubungan kompleks antar fitur, terutama dalam data berdimensi banyak seperti data pemain sepak bola.



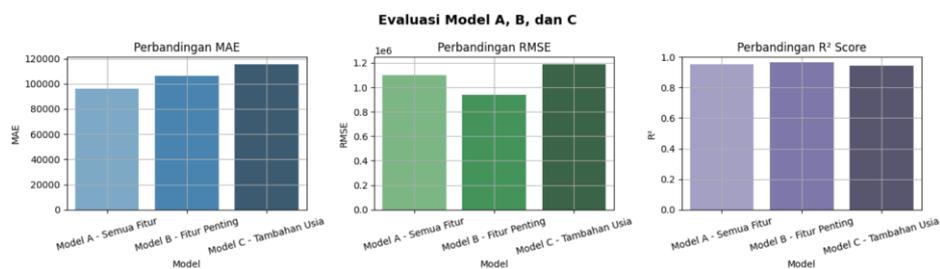
Gambar 4. 5 Perbandingan Nilai Prediksi dan Nilai Asli

Meskipun model hasil tuning dengan parameter optimal menghasilkan skor validasi terbaik sebesar 0.9518, namun ketika diuji pada data nyata, model mengalami sedikit penurunan performa (R^2 menjadi 0.9134). Hal ini menunjukkan bahwa tuning yang terlalu spesifik terhadap data latih dapat menyebabkan overfitting, dan model menjadi kurang general terhadap data baru.



Gambar 4. 6 Pengujian 5-Fold Cross Validation

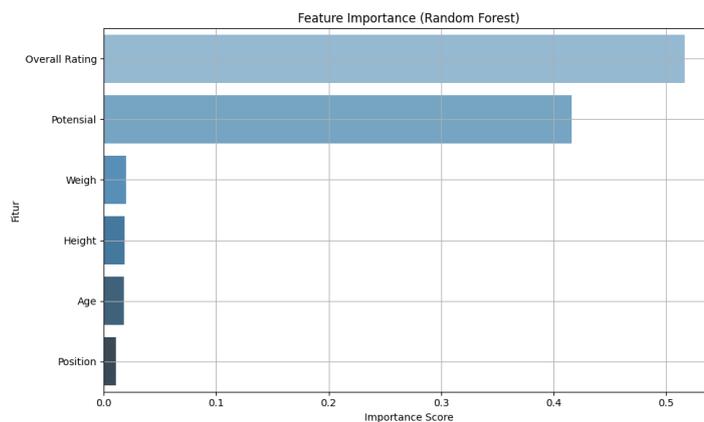
Pengujian 5-Fold Cross Validation menghasilkan rata-rata R^2 sebesar 0.9629, yang memperkuat bahwa model ini stabil dan mampu melakukan generalisasi dengan baik terhadap berbagai pembagian data. Beberapa fold menunjukkan performa sangat tinggi (bahkan di atas 0.98), sementara sebagian lain menunjukkan error yang lebih tinggi, namun tetap dalam batas wajar.



Gambar 4. 7 Perbandingan Model A, B, dan C

Eksperimen pada bagian 4.4 menunjukkan bahwa penggunaan semua fitur tidak selalu memberikan hasil terbaik. Justru pada Model B, yang hanya menggunakan dua fitur terpenting (*overall rating* dan *potensial*), model mampu menghasilkan R^2 tertinggi (0.9651) dan RMSE terendah. Hal ini menunjukkan pentingnya melakukan seleksi fitur agar model tetap efisien dan tidak terdistraksi oleh informasi yang kurang relevan.

Di sisi lain, Model C menunjukkan bahwa penambahan fitur seperti *age*, meskipun memberi informasi tambahan, tidak selalu meningkatkan performa. Hal ini mendukung kesimpulan bahwa model prediksi terbaik adalah model yang sederhana namun fokus pada fitur-fitur paling signifikan.



Gambar 4. 8 Hasil Pengujian Pengaruh Fitur

Hasil analisis *feature importance* memperkuat temuan sebelumnya: Overall Rating dan Potensial memiliki kontribusi dominan dalam membentuk prediksi model, dengan nilai *importance* tertinggi dan dampak terbesar terhadap pengurangan MSE. Fitur seperti *Height*, *Weight*, dan *Position* memberikan kontribusi yang jauh lebih kecil.

Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa performa algoritma *Random Forest* dalam memprediksi *market value* pemain sepak bola U-22 Eropa tergolong sangat baik, dengan nilai R^2 yang tinggi menunjukkan bahwa model mampu menjelaskan sebagian besar variasi nilai pasar pemain secara akurat. Selain itu, proses analisis *feature importance* berhasil mengidentifikasi faktor-faktor yang paling berpengaruh dalam proses transfer, di mana *Overall Rating* dan *Potensial* terbukti menjadi indikator utama yang mencerminkan kualitas dan prospek jangka panjang pemain.

4.8 Integrasi Penelitian

Dalam Islam, aktivitas ekonomi seperti jual beli merupakan bagian penting dari kehidupan manusia yang mendapatkan perhatian khusus. Jual beli diperbolehkan selama dilakukan dengan cara yang adil, tidak merugikan salah satu pihak, dan tidak mengandung unsur penipuan. Prinsip-prinsip ini sejalan dengan tujuan penelitian ini, yaitu untuk mengembangkan model prediktif berbasis algoritma *Random Forest* guna memperkirakan *market value* pemain sepak bola secara lebih objektif, transparan, dan adil. Hal ini ditegaskan dalam firman Allah SWT dalam Surah Al-Baqarah ayat 275:

وَأَحَلَّ اللَّهُ الْبَيْعَ وَحَرَّمَ الرِّبَا

“Allah telah menghalalkan jual beli dan mengharamkan riba...”
(QS. Al-Baqarah: 275)

Ayat ini menunjukkan bahwa transaksi jual beli diperbolehkan oleh Allah, selama dilakukan dengan etika dan tanggung jawab. Dalam konteks sepak bola profesional, aktivitas perpindahan pemain antar klub sering kali melibatkan nilai transaksi yang besar. Oleh karena itu, penting bagi klub maupun agen untuk memiliki dasar penilaian yang akurat terhadap harga pasar (*market value*) seorang pemain, agar transaksi yang dilakukan tidak merugikan salah satu pihak.

Selain dari Al-Qur'an, Rasulullah SAW juga memberikan banyak tuntunan mengenai kejujuran dalam jual beli. Dalam sebuah hadis riwayat Al-Bukhari, Rasulullah SAW bersabda:

البيعان بالخيار ما لم يتفرقا، فإن صدقا وبيّنا بورك لهما في بيعهما، وإن كتما وكذبا محقت بركة بيعهما.

“Penjual dan pembeli memiliki hak pilih selama mereka belum berpisah. Jika keduanya jujur dan menjelaskan (keadaan barang), maka akan diberkahi jual belinya. Namun jika mereka menyembunyikan dan berdusta, maka akan dihapus berkah jual belinya.” (HR. Bukhari, no. 2079)

Hadis ini menekankan pentingnya kejujuran dan keterbukaan informasi dalam transaksi jual beli. Dalam penelitian ini, algoritma *Random Forest* digunakan untuk membangun sistem yang dapat membantu memberikan estimasi nilai pasar yang lebih akurat berdasarkan data performa pemain muda U-22 di Eropa. Dengan pendekatan ini, diharapkan ekosistem transfer pemain dapat dijalankan dengan lebih transparan, adil, dan profesional, sesuai dengan nilai-nilai Islam.

Dengan demikian, integrasi teknologi dalam bentuk sistem prediksi ini bukan hanya bertujuan untuk efisiensi dan akurasi, tetapi juga sejalan dengan nilai-nilai etika Islam dalam bertransaksi, yakni menjunjung tinggi kejujuran, keadilan, dan keberkahan dalam setiap proses jual beli.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, algoritma *Random Forest* terbukti efektif dalam memprediksi nilai pasar pemain sepak bola muda U-22 dengan tingkat akurasi tinggi, ditunjukkan oleh nilai R^2 sebesar 0.9525 dan RMSE sebesar €1.098.355,92 pada model default. Hasil tuning hyperparameter menghasilkan validasi terbaik, namun pada data uji justru mengalami penurunan performa, menunjukkan potensi overfitting. Pengujian K-Fold Cross Validation menghasilkan R^2 rata-rata 0.9629, menandakan bahwa model cukup stabil dan dapat diandalkan. Eksperimen terhadap jumlah fitur menunjukkan bahwa hanya dengan dua fitur utama (overall rating dan potensial), model sudah dapat menghasilkan akurasi terbaik. Analisis feature importance mengonfirmasi bahwa kedua fitur tersebut paling dominan, sedangkan fitur lain seperti weight dan height memiliki kontribusi sangat rendah. Nilai MSE yang besar (hingga triliunan) dianggap wajar karena dataset berisi sekitar 5000 entri dan skala nilai pasar yang besar

5.2 Saran

Berdasarkan hasil yang diperoleh serta keterbatasan dalam penelitian ini, disarankan agar penelitian selanjutnya menggunakan data yang lebih kaya dan beragam, seperti statistik performa pertandingan, riwayat cedera, serta faktor eksternal seperti popularitas pemain dan eksposur media sosial. Penambahan fitur-fitur tersebut diyakini dapat meningkatkan akurasi dan kedalaman prediksi model.

Ke depan, model ini dapat dikembangkan menjadi aplikasi atau sistem pendukung keputusan yang digunakan oleh klub-klub sepak bola untuk menyusun strategi transfer secara lebih objektif dan efisien. Penelitian lanjutan juga disarankan untuk menguji model pada data real-time atau musim-musim terbaru agar model dapat beradaptasi terhadap dinamika pasar yang selalu berubah. Dengan demikian, penelitian di bidang ini dapat terus dikembangkan dan memberikan kontribusi nyata dalam dunia olahraga modern yang semakin berbasis data.

DAFTAR PUSTAKA

- Al-Asadi, M. A., & Tasdemir, S. (2022). Predict the value of football players using FIFA video game data and machine learning techniques. *IEEE Access*, *10*, 22631–22645. <https://doi.org/10.1109/access.2022.3154767>
- Ali, A. (2024). *The role of performance and popularity in the valuation of athletes in sports*. <https://doi.org/10.5281/zenodo.14556090>
- Baouan, A., Bismuth, E., Bohbot, A., Coustou, S., Lacombe, M., & Rosenbaum, M. (2022). What should clubs monitor to predict future value of football players. *2212.11041*. <https://doi.org/10.14711/thesis-991012980425503412>
- Bhadula, S., Adnan, M. M., Kumar, R., Rana, A., Kaliyaperumal, G., Rao, B. D., & Boob, N. S. (2024). *Optimizing Random Forest Algorithms for LargeScale Data Analysis*. *7*, 1673–1678. <https://doi.org/10.1109/ic3i61595.2024.10829145>
- Cortez, A., Trigo, A., & Loureiro, N. (2021). *Predicting Physiological Variables of Players that Make a Winning Football Team: A Machine Learning Approach*. 3–15. <https://doi.org/10.3390/computers11030040>
- Di Domizio, M., Caruso, R., & Frick, B. (2024). The Appraisal of Players' Transfer Market Values: Empirical Evidence From Italian Serie A. *International Journal of Sport Finance*, *19*(1), 39–51. <https://doi.org/10.32731/ijsf/191.022024.03>
- Dobson, S., & Goddard, J. (2011). *The Economics of Football* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511973864>
- Firdausi, H., Sumarjaya, I. W., & Srinadi, I. (2024). Prediksi Market Value Pemain Sepak Bola di Lima Liga Top Eropa Menggunakan K-Nearest Neighbor. *E-Jurnal Matematika*, *13*(2), 89. <https://doi.org/10.24843/mtk.2024.v13.i02.p446>
- Hamzah, N. B., Abd Karim, Z., Yaakop, N., Akbar, A., & Lee, J. L. F. (2025). Key factors influencing talent development in youth football: A systematic literature review. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, *62*, 948–957. <https://doi.org/10.47197/retos.v62.109470>
- Hu, J. (2024). Research on predicting football matches based on handicap data and BPNN. *Applied and Computational Engineering*, *31*(1), 29–35. <https://doi.org/10.54254/2755-2721/31/20230118>
- Huang, C., & Zhang, S. (2023). *Explainable artificial intelligence model for identifying Market Value in Professional Soccer Players* (Version 2). <https://doi.org/10.48550/2311.04599>

- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics*, 10(11), 1942. <https://doi.org/10.3390/math10111942>
- Lee, H., Tama, B. A., & Cha, M. (2022). *Prediction of Football Player Value using Bayesian Ensemble Approach* (Version 1). <https://doi.org/10.48550/2206.13246>
- Leksowski, Ł. (2022). Factors Determining the Market Value of Football Players on the Example of Ekstraklasa. *Studia i Materiały Wydział Zarządzania Uniwersytet Warszawski*, 2022(1), 54–68. <https://doi.org/10.7172/1733-9758.2022.36.5>
- Lončar, J., & Špehar, E. (2021). Globalization of Sports on the Example Of The Five Most Popular European Football Leagues with an Emphasis on Premiership-Geographical Overview. *Društvene i Humanističke Studije*, 6(3 (16)), 611–634. <https://doi.org/10.51558/2490-3647.2021.6.3.611>
- Mastromartino, B., Naraine, M. L., Dees, W., & Zhang, J. J. (2024). Scoping practical implications and managerial relevance in sport management. *Sport, Business and Management: An International Journal*, 14(3), 309–329. <https://doi.org/10.1108/SBM-07-2023-0097>
- Mazziotta, M., & Pareto, A. (2022). Normalization methods for spatio-temporal analysis of environmental performance: Revisiting the Min–Max method. *Environmetrics*, 33(5), e2730. <https://doi.org/10.1002/env.2730>
- Ren, Y., & Susnjak, T. (2022). *Predicting Football Match Outcomes with eXplainable Machine Learning and the Kelly Index* (Version 1). <https://doi.org/10.48550/2211.15734>
- Sanchez, L. C., Barajas, A., & Sanchez-Fernandez, P. (2024). Are football clubs as pieces of art or as regular corporations? An empirical evidence of market valuation of football clubs in the big 5 leagues. *Managing Sport and Leisure*, 29(4), 591–610. <http://dx.doi.org/10.1080/23750472.2022.2081251>
- Sevgen, S. C., & Aliefendioğlu, Y. (2020). Mass appraisal with a machine learning algorithm: Random forest regression. *Bilişim Teknolojileri Dergisi*, 13(3), 301–311. <https://doi.org/10.17671/gazibtd.555784>
- Shen, Q. (2025). Predicting the value of football players: Machine learning techniques and sensitivity analysis based on FIFA and real-world statistical datasets. *Applied Intelligence*, 55(4), 265. <https://doi.org/10.1007/s10489-024-06189-0>
- Sulimov, D. (2024). *Performance Insights-based AI-driven Football Transfer Fee Prediction* (Version 1). <https://doi.org/10.48550/ARXIV.2401.16795>

- Wang, T., Zhang, Z., & Zhu, S. (2024). Machine learning-based football match prediction system. *Applied and Computational Engineering*, *92*(1), 181–186. <https://doi.org/10.54254/2755-2721/92/20241749>
- Yeung, C., Bunker, R., Umemoto, R., & Fujii, K. (2024). Evaluating soccer match prediction models: A deep learning approach and feature optimization for gradient-boosted trees. *Machine Learning*, *113*(10), 7541–7564. <https://doi.org/10.48550/arXiv.2309.14807>