

Klasifikasi Sepatu Sepak Bola Berdasarkan Peminatan, *Brand*, Dan Posisi Pemain Dengan *Machine Learning*

Robby Kurniawan Syahfitra¹, Farid Akbar Siregar²

^{1,2} Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

Article Info	ABSTRAK
Article history: Received September 2, 2026 Revised September 2, 2026 Accepted September 2, 2026 Kata Kunci: Klasifikasi, Sepatu Sepak Bola, <i>Machine Learning</i>, <i>Random Forest</i>, SMOTE Keywords: <i>Classification</i>, <i>Football Boots</i>, <i>Machine Learning</i>, <i>Random Forest</i>, SMOTE	Pemilihan sepatu sepak bola yang tepat memengaruhi kenyamanan dan performa pemain, namun pemain amatir hingga semi-profesional sering mengalami kesulitan menentukan tipe sepatu yang sesuai karena kuatnya pengaruh merek dan kurangnya pertimbangan aspek teknis. Penelitian ini bertujuan mengklasifikasikan tipe sepatu sepak bola berdasarkan peminatan, brand, dan posisi pemain menggunakan algoritma Random Forest. Dataset awal berjumlah 3.971 data sepatu sepak bola. Setelah pembersihan data dan penanganan nilai kosong diperoleh 3.803 data, kemudian dibagi menjadi 3.042 data latih dan 761 data uji. Ketidakseimbangan kelas pada data latih ditangani menggunakan Synthetic Minority Over-sampling Technique (SMOTE) sehingga jumlah data latih menjadi 12.712. Hasil pengujian menunjukkan akurasi 99,61%, dengan 758 dari 761 data uji diklasifikasikan secara benar. Hasil feature importance menunjukkan BootsName sebagai fitur paling berpengaruh, diikuti Brand BootsName dan BootsPosition. Model menunjukkan performa klasifikasi sangat baik. Model kemudian diimplementasikan dalam prototipe sistem berbasis web SMART BOOTS untuk menghasilkan klasifikasi dan rekomendasi sepatu berdasarkan karakteristik yang dimasukkan pengguna. ABSTRACT <i>Choosing appropriate football boots affects player comfort and performance, yet amateur to semi-professional players often have difficulty selecting a suitable boot type because brand influence can dominate technical considerations. This study aims to classify football boot types based on player preference, brand, and playing position using the Random Forest algorithm. The initial dataset contained 3,971 football boot records. After data cleaning and missing-value handling, 3,803 records remained and were divided into 3,042 training records and 761 testing records. Class imbalance in the training data was handled using the Synthetic Minority Over-sampling Technique (SMOTE), increasing the training set to 12,712 records. Testing produced an accuracy of 99.61%, with 758 of 761 testing records classified correctly. Feature-importance analysis indicated BootsName as the most influential feature, followed by Brand BootsName and BootsPosition. The resulting model was implemented in the SMART BOOTS web-based prototype to provide football boot classification and recommendations according to the characteristics entered by users.</i> <i>This is an open access article under the CC BY license</i>



Corresponding Author:

Robby Kurniawan Syahfitra
Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara
Medan, Indonesia
Email: robbyk543@gmail.com

1. PENDAHULUAN

Sepak bola merupakan olahraga dengan basis pemain dan penggemar yang sangat luas. Pertumbuhan aktivitas sepak bola juga meningkatkan kebutuhan terhadap perlengkapan pendukung, termasuk sepatu sepak bola dengan beragam karakteristik teknis, merek, dan kategori penggunaan. Dalam kondisi tersebut, pemain terutama pada level amatir hingga semi-profesional menghadapi kesulitan ketika harus menentukan sepatu yang sesuai dengan karakter permainan. Keputusan pemilihan tidak selalu didasarkan pada kesesuaian teknis, tetapi sering dipengaruhi citra merek dan preferensi terhadap produk tertentu. Dalam kajian olahraga modern, pemanfaatan artificial intelligence dan machine learning telah berkembang untuk membantu analisis dan pengambilan keputusan berbasis data [1], sedangkan sistem rekomendasi olahraga semakin diarahkan pada personalisasi pilihan berdasarkan kebutuhan pengguna [2].

Kesesuaian antara karakteristik pemain dan perlengkapan menjadi penting karena posisi bermain memiliki tuntutan yang berbeda. Defender memerlukan stabilitas dan perlindungan, midfielder menuntut kenyamanan dan kontrol bola, forward berorientasi pada akselerasi dan akurasi, sedangkan goalkeeper memerlukan stabilitas lateral dan cengkeraman. Studi machine learning pada sepak bola juga menunjukkan bahwa posisi pemain dapat dipelajari dari karakteristik performa dan biomekanik [3]. Di sisi lain, data sepak bola telah banyak digunakan dalam pemodelan machine learning untuk mendukung keputusan, termasuk estimasi nilai pemain berdasarkan atribut yang tersedia [4].

Pada konteks produk olahraga, pemilihan sepatu berkaitan dengan preferensi terhadap karakteristik produk. Penggunaan model komputasional untuk produk sepatu juga telah diterapkan, antara lain melalui pemodelan neural network pada desain sepatu olahraga [5] dan sistem pengukuran objektif untuk mendukung kecocokan alas kaki atlet [6]. Namun, penelitian pada skripsi ini diarahkan pada masalah yang lebih spesifik, yaitu klasifikasi tipe sepatu sepak bola dengan menggabungkan atribut produk dan pemain dalam dataset terstruktur sehingga hasilnya dapat dimanfaatkan sebagai dasar rekomendasi.

Algoritma Random Forest dipilih sebagai metode utama karena merupakan ensemble dari sejumlah decision tree yang menggabungkan prediksi melalui voting. Random Forest dikembangkan sebagai metode klasifikasi dan regresi yang mampu menangkap hubungan nonlinier, mengurangi risiko overfitting melalui pengacakan sampel dan fitur, serta menyediakan ukuran kepentingan fitur [7]. Implementasinya telah banyak digunakan pada data terstruktur [8], sementara kajian teoritis menunjukkan fleksibilitas dan performanya pada beragam masalah klasifikasi [9]. Perbandingan banyak algoritma klasifikasi juga menempatkan keluarga Random Forest sebagai salah satu metode dengan performa kuat pada berbagai dataset [10].

Kesenjangan penelitian yang menjadi dasar penelitian ini adalah belum banyak pendekatan yang secara khusus mengintegrasikan peminatan pemain, brand, dan posisi pemain untuk mengklasifikasikan tipe sepatu sepak bola dalam satu model terstruktur. Riset yang dibahas dalam skripsi cenderung terpisah antara kajian pemasaran, rekomendasi produk, dan aplikasi machine learning pada data olahraga. Oleh karena itu, penelitian ini memanfaatkan Random Forest untuk memetakan pola atribut produk dan pemain terhadap kelas BootsType, sekaligus menggunakan feature importance untuk mengidentifikasi fitur yang paling berpengaruh. Pendekatan ini sejalan dengan karakter Random Forest yang dapat digunakan untuk mengevaluasi kontribusi prediktor dalam model [11], [12].

Penelitian ini bertujuan menerapkan algoritma machine learning untuk mengklasifikasikan sepatu sepak bola berdasarkan peminatan, brand, dan posisi pemain, mengevaluasi performa model pada data uji, serta mengimplementasikan hasil klasifikasi dalam prototipe sistem rekomendasi berbasis web SMART BOOTS.

2. METODE

2.1 Pendekatan Penelitian

Penelitian menggunakan pendekatan kuantitatif berbasis machine learning dengan Random Forest sebagai algoritma klasifikasi utama. Proses penelitian mencakup penyediaan dataset, preprocessing, transformasi data kategorikal, pembagian data latih dan data uji, penanganan ketidakseimbangan kelas, pelatihan model, evaluasi, serta implementasi model pada sistem berbasis web. Pemilihan Random Forest mengikuti mekanisme ensemble yang membangun banyak pohon keputusan dan mengagregasikan hasil prediksi, dengan parameter yang dapat disesuaikan untuk memperoleh performa klasifikasi yang stabil [7], [12].

2.2 Dataset dan Preprocessing

Dataset aktual yang digunakan pada tahap hasil penelitian berasal dari Kaggle dan berjumlah 3.971 data sepatu sepak bola. Data mencakup atribut yang digunakan dalam proses klasifikasi, antara lain Brand, BootsName, BootsPosition, PlayerPosition, serta BootsType sebagai target klasifikasi. Kategori brand meliputi Adidas, Nike, Puma, dan Mizuno, sedangkan target BootsType terdiri dari Agility, Chaos, Control, Heritage, Modern C, Speed, dan Striking. Tahap preprocessing dimulai dengan membaca data menggunakan Python, menghapus baris yang tidak lengkap, menangani nilai kosong, dan memastikan atribut dapat diproses model. Dari 3.971 data awal, sebanyak 168 data dihapus sehingga tersisa 3.803 data yang layak digunakan.

Data kategorikal kemudian ditransformasikan menjadi representasi numerik menggunakan Label Encoding. Hasil encoding disimpan agar dapat digunakan kembali ketika model melakukan prediksi pada sistem. Dataset bersih disimpan dalam format CSV sebagai dasar pelatihan dan pengujian Random Forest.

2.3 Pembagian Data dan Penyeimbangan Kelas

Dataset bersih sebanyak 3.803 data dibagi dengan rasio 80:20 menjadi 3.042 data latih dan 761 data uji. Ketidakseimbangan jumlah data pada setiap kelas target ditangani menggunakan Synthetic Minority Over-sampling Technique (SMOTE). SMOTE membentuk sampel sintetis pada kelas minoritas sehingga distribusi data latih menjadi lebih seimbang [13]. Setelah proses tersebut, jumlah data latih meningkat dari 3.042 menjadi 12.712 data.

2.4 Pelatihan dan Evaluasi Model

Model Random Forest dilatih menggunakan data yang telah melalui proses penyeimbangan. Dalam proses klasifikasi, setiap pohon menghasilkan prediksi dan kelas akhir ditentukan melalui majority voting. Model kemudian disimpan dalam berkas sepatu_rf_model.pkl untuk digunakan pada aplikasi web. Evaluasi dilakukan pada 761 data uji menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Evaluasi terpisah antara data latih dan data uji diperlukan agar performa model mencerminkan kemampuan generalisasi, sebagaimana prinsip evaluasi prediktif yang menekankan sensitivitas prosedur validasi terhadap estimasi kesalahan [14]. Penelitian sejenis oleh Zikry dan Siregar juga menunjukkan bahwa tahapan preprocessing, ekstraksi fitur, pemilihan algoritma, dan evaluasi metrik menentukan performa klasifikasi machine learning [15].

2.5 Implementasi Sistem

Model yang telah dilatih diintegrasikan ke prototipe web SMART BOOTS. Alur sistem dimulai dari autentikasi pengguna, pengisian karakteristik sepatu dan pemain, validasi relasi data, pemrosesan oleh model Random Forest, kemudian penampilan BootsType dan tingkat keyakinan prediksi. Sistem juga menyediakan halaman informasi dan hasil klasifikasi sehingga pengguna memperoleh penjelasan singkat mengenai kategori sepatu yang diprediksi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing dan Pembagian Dataset

Hasil pengolahan menunjukkan bahwa dataset awal sebanyak 3.971 data berkurang menjadi 3.803 data setelah proses pembersihan. Pengurangan sebanyak 168 data berasal dari observasi yang tidak memenuhi kelengkapan atribut yang dibutuhkan. Dataset bersih kemudian dibagi menjadi 3.042 data latih dan 761 data uji. Ringkasan perubahan jumlah data disajikan pada Tabel 1.

Tabel 1. Ringkasan Dataset dan Preprocessing

Tahap	Jumlah Data	Keterangan
Data awal	3.971	Dataset sepatu sepak bola
Data dihapus	168	Data tidak lengkap/missing value
Data setelah cleaning	3.803	Data siap diproses
Training	3.042	80% data bersih
Testing	761	20% data bersih
Training setelah SMOTE	12.712	Data latih seimbang

Penyeimbangan menggunakan SMOTE meningkatkan jumlah data latih menjadi 12.712. Tahap ini dilakukan karena distribusi target BootsType tidak seimbang. Penanganan kelas minoritas dibutuhkan agar model tidak hanya mempelajari kelas yang dominan. Dalam penelitian ini, data hasil SMOTE digunakan untuk pelatihan, sedangkan data uji tetap sebanyak 761 sehingga evaluasi dilakukan pada observasi yang tidak dibentuk melalui oversampling.

3.2 Hasil Pengujian Model

Pengujian model dilakukan dengan memasukkan kombinasi Brand, BootsName, BootsPosition, dan PlayerPosition. Model mampu menghasilkan kelas BootsType beserta confidence untuk setiap kombinasi input. Beberapa contoh pengujian yang tercantum dalam skripsi ditampilkan pada Tabel 2.

Tabel 2. Contoh Hasil Pengujian Model Random Forest

No.	Input Utama	Output
1	Puma; Ultra 1.1; Attack; Center Back	Speed; 80,63%
2	Mizuno; Morelia II; Defence; Defensive Midfielder	Heritage; 99,80%
3	Nike; Magista Opus; Midfield; Left Back	Control; 88,46%
4	Adidas; F50 Adizero Messi; Standard	Control; 92,00%
5	Nike; Hypervenom Phantom II; Defence; Support Strike	Modern C; 34,46%
6	Mizuno; Alpha; Attack; Center Back	Agility; 48,90%
7	Nike; Phantom VNM; Attack; Center Back	Striking; 99,62%
8	Adidas; X 16+ Purechaos; Attack; Center Back	Chaos; 99,89%
9	Nike; Tiempo Legend 9; Defence; Central Midfielder	Modern C; 99,68%
10	Adidas; Nameziz.1; Attack; Attacking Midfielder	Agility; 99,99%

3.3 Evaluasi Akurasi dan Metrik Klasifikasi

Pada 761 data uji, sebanyak 758 data diprediksi dengan benar sehingga accuracy mencapai 99,61%. Perhitungan tersebut diperoleh dari $758/761 \times 100\%$. Hasil classification report per kelas disajikan pada Tabel 3. Sebagian besar kelas memperoleh precision, recall, dan F1-score yang sangat tinggi. Kelas Control, Modern C, dan Striking memperoleh nilai 1,00 pada ketiga metrik, sedangkan kelas Speed memiliki precision 0,99 dan recall serta F1-score 1,00. Kelas Agility memperoleh precision 1,00, recall 0,94, dan F1-score 0,97; Heritage memperoleh precision 1,00, recall 0,91, dan F1-score 0,95.

Tabel 3. Hasil Evaluasi Model

Kelas Sepatu	Precision	Recall	F1-Score
Agility	1,00	0,94	0,97
Chaos	0,00	0,00	0,00
Control	1,00	1,00	1,00
Heritage	1,00	0,91	0,95
Modern C	1,00	1,00	1,00
Speed	0,99	1,00	1,00
Striking	1,00	1,00	1,00

Keterbatasan utama terlihat pada kelas Chaos yang memiliki precision, recall, dan F1-score sebesar 0,00. Confusion matrix menunjukkan bahwa sebagian besar prediksi berada pada diagonal utama: 228 data Control, 454 data Speed, 44 data Modern C, 10 data Heritage, 15 data Agility, dan 7 data Striking diklasifikasikan dengan benar. Terdapat tiga kesalahan klasifikasi, yaitu masing-masing satu data Agility, Chaos, dan Heritage yang diprediksi sebagai Speed. Dengan demikian, nilai akurasi keseluruhan yang sangat tinggi tetap perlu dibaca bersama distribusi performa per kelas karena kelas yang sangat sedikit belum memperoleh pola yang memadai.

3.4 Feature Importance dan Interpretasi Model

Analisis hasil penelitian menunjukkan bahwa BootsName merupakan fitur yang paling berpengaruh dalam proses klasifikasi, kemudian diikuti Brand_BootsName dan BootsPosition. Temuan ini menunjukkan bahwa model tidak hanya bergantung pada satu atribut, tetapi menangkap pola kombinasi karakteristik produk dan posisi. Penggunaan feature importance merupakan salah satu kelebihan Random Forest untuk membantu mengidentifikasi prediktor yang berkontribusi pada keputusan model [16]. Namun, ukuran kepentingan fitur juga perlu diinterpretasikan secara hati-hati karena dapat dipengaruhi struktur dan korelasi antarvariabel [17].

3.5 Implementasi SMART BOOTS

Model Random Forest diimplementasikan pada SMART BOOTS. Halaman login digunakan sebagai gerbang autentikasi, sedangkan dashboard menyediakan formulir Brand, Boots Name, Player Position, dan Boots Position. Setelah input divalidasi, pengguna dapat menjalankan analisis dan memperoleh BootsType beserta tingkat keyakinan. Salah satu contoh hasil pada skripsi adalah klasifikasi Adidas 11 Pro sebagai BootsType Speed dengan confidence 82,05%. Sistem juga memberikan analisis singkat mengenai karakteristik tipe sepatu yang diprediksi sehingga hasil klasifikasi dapat digunakan sebagai dasar informasi pemilihan produk.

3.6 Pembahasan

Hasil 99,61% menunjukkan bahwa Random Forest mampu mempelajari pola dataset terstruktur yang digunakan dalam penelitian. Secara metodologis, hasil tersebut konsisten dengan karakter Random Forest sebagai ensemble yang mengurangi varians melalui penggabungan banyak pohon [7], serta dengan kajian yang menunjukkan kemampuan Random Forest pada berbagai masalah klasifikasi [9], [10]. Penggunaan SMOTE sebelum pelatihan juga sesuai dengan tujuan penelitian untuk memperbaiki representasi kelas minoritas [13]. Meskipun demikian, hasil kelas Chaos memperlihatkan bahwa oversampling tidak otomatis menjamin performa yang sama untuk setiap kelas ketika jumlah contoh asli sangat sedikit.

Nilai akurasi yang sangat tinggi juga perlu dipahami bersama aspek validasi dan generalisasi. Random Forest memiliki mekanisme evaluasi internal, tetapi estimasi performa dapat terpengaruh distribusi data, tuning parameter, dan pemisahan sampel [12], [18]. Karena itu, hasil 758 prediksi benar dari 761 data uji merupakan performa pada dataset penelitian ini dan tidak serta-merta menyatakan tingkat akurasi yang sama pada data baru di luar cakupan dataset. Saran penelitian dalam skripsi untuk memperluas dataset, menambahkan atribut seperti jenis lapangan, harga, atau material sepatu, dan membandingkan Random Forest dengan XGBoost, SVM, atau Neural Network menjadi langkah penting untuk menguji kemampuan generalisasi model.

Dalam konteks olahraga, penggunaan machine learning untuk data sepak bola telah berkembang pada analisis posisi dan keputusan berbasis data [1], [3], [4]. Implementasi SMART BOOTS memperluas penerapan tersebut ke klasifikasi produk dan rekomendasi sepatu, sejalan dengan arah sistem rekomendasi olahraga yang memanfaatkan informasi pengguna dan karakteristik item [2]. Penelitian lokal terkait klasifikasi machine learning juga menunjukkan pentingnya pemilihan algoritma dan preprocessing untuk memperoleh performa yang baik [15]. Dengan demikian, kontribusi utama penelitian ini terletak pada integrasi dataset sepatu, Random Forest, penanganan ketidakseimbangan kelas, feature importance, dan implementasi prototipe web dalam satu alur klasifikasi.

4. KESIMPULAN

Penelitian berhasil membangun model klasifikasi sepatu sepak bola menggunakan Random Forest berdasarkan data yang memuat karakteristik brand, nama sepatu, posisi sepatu, dan posisi pemain. Dari 3.971 data awal, proses preprocessing menghasilkan 3.803 data yang selanjutnya dibagi menjadi 3.042 data latih dan 761 data uji. Penyeimbangan menggunakan SMOTE meningkatkan data latih menjadi 12.712. Pengujian menghasilkan akurasi 99,61% dengan 758 dari 761 data uji diklasifikasikan secara benar. Sebagian besar kelas memiliki precision, recall, dan F1-score yang sangat tinggi, tetapi kelas Chaos belum dapat diklasifikasikan dengan baik karena jumlah datanya sangat terbatas. Feature importance menunjukkan BootsName sebagai fitur paling berpengaruh, diikuti Brand_BootsName dan BootsPosition. Model kemudian diterapkan pada prototipe SMART BOOTS untuk menampilkan kelas BootsType dan tingkat keyakinan prediksi. Pengembangan berikutnya diarahkan pada perluasan dan diversifikasi dataset, penambahan atribut seperti jenis lapangan, harga, dan material sepatu, serta perbandingan Random Forest dengan algoritma klasifikasi lain untuk menguji generalisasi dan memperoleh model terbaik.

REFERENSI

- [1] N. Chmait and H. Westerbeek, "Artificial intelligence and machine learning in sport research: An introduction for non-data scientists," *Front. Sports Act. Living*, vol. 3, Art. no. 682287, 2021, doi: 10.3389/fspor.2021.682287.

- [2] A. Felfernig, M. Wundara, T. N. T. Tran, V.-M. Le, S. Lubos, and S. Polat-Erdeniz, "Sports recommender systems: Overview and research directions," *J. Intell. Inf. Syst.*, vol. 62, pp. 1125–1164, 2024, doi: 10.1007/s10844-024-00857-w.
- [3] F. H. Yagin, U. C. H. Hasan, F. M. Clemente, O. Eken, G. Badicu, and M. Gulu, "Using machine learning to determine the positions of professional soccer players in terms of biomechanical variables," *Proc. Inst. Mech. Eng. P J. Sports Eng. Technol.*, vol. 239, no. 4, pp. 726–733, 2025, doi: 10.1177/17543371231199814.
- [4] I. Behravan and S. M. Razavi, "A novel machine learning method for estimating football players' value in the transfer market," *Soft Comput.*, vol. 25, no. 3, pp. 2499–2511, 2021, doi: 10.1007/s00500-020-05319-3.
- [5] Y.-E. Yeh, "Prediction of optimized color design for sports shoes using an artificial neural network and genetic algorithm," *Appl. Sci.*, vol. 10, no. 5, Art. no. 1560, 2020, doi: 10.3390/app10051560.
- [6] J. W. Wannop, D. J. Stefanyshyn, R. B. Anderson, M. J. Coughlin, and R. Kent, "Development of a footwear sizing system in the National Football League," *Sports Health*, vol. 11, no. 1, pp. 40–46, 2019, doi: 10.1177/1941738118789402.
- [7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [8] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [9] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, 2016, doi: 10.1007/s11749-016-0481-7.
- [10] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, pp. 3133–3181, 2014.
- [11] A.-L. Boulesteix, S. Janitza, J. Kruppa, and I. R. König, "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics," *WIREs Data Min. Knowl. Discov.*, vol. 2, no. 6, pp. 493–507, 2012, doi: 10.1002/widm.1072.
- [12] P. Probst, M. N. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 3, Art. no. e1301, 2019, doi: 10.1002/widm.1301.
- [13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [14] J. D. Rodríguez, A. Pérez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010, doi: 10.1109/TPAMI.2009.187.
- [15] A. Zikry and F. A. Siregar, "Performance analysis of logistic regression and SVM (support vector machine) algorithms on e-football mobile game review sentiment," *J. Sci. Technol. Innov.*, vol. 1, no. 3, pp. 212–224, 2026, doi: 10.65310/073zdt91.
- [16] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, 2010, doi: 10.1016/j.patrec.2010.03.014.
- [17] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinformatics*, vol. 9, Art. no. 307, 2008, doi: 10.1186/1471-2105-9-307.
- [18] S. Janitza and R. Hornung, "On the overestimation of random forest's out-of-bag error," *PLoS ONE*, vol. 13, no. 8, Art. no. e0201904, 2018, doi: 10.1371/journal.pone.0201904.