TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

Perbandingan Metode Random Forest dengan Decision Tree pada Sistem Rekomendasi Olahraga Berdasarkan Karakteristik Kepribadian

Galih Tri Ardiansyah*, Titania Dwiandini

Fakultas Teknik dan Desain, Program Studi Teknik Informatika, Institut Teknologi & Bisnis Asia, Malang, Indonesia
Email: 1,*tugasnyagalih@email.com, 2titania.andini@asai.ac.id
Email Penulis Korespondensi: tugasnyagalih@email.com

Abstrak–Pemilihan olahraga yang sesuai dengan karakteristik individu merupakan faktor penting untuk meningkatkan motivasi, kenyamanan, dan konsistensi seseorang dalam melakukan aktivitas fisik. Setiap individu memiliki perbedaan karakteristik psikologis, preferensi aktivitas, tingkat intensitas olahraga, serta tujuan olahraga yang ingin dicapai. Salah satu pendekatan yang dapat digunakan untuk menghasilkan rekomendasi olahraga yang lebih personal adalah dengan memanfaatkan karakteristik kepribadian berdasarkan model *Big Five Personality (OCEAN).* Penelitian ini bertujuan untuk membandingkan performa algoritma *Decision Tree* dan *Random Forest* dalam melakukan klasifikasi kategori olahraga berdasarkan karakteristik kepribadian, intensitas olahraga, preferensi sosial, lokasi olahraga, dan tujuan olahraga. Kebaruan penelitian ini terletak pada penerapan serta evaluasi komparatif kedua algoritma tersebut dalam sistem rekomendasi olahraga berbasis *Big Five Personality* yang masih belum banyak dikembangkan. Dataset yang digunakan diperoleh dari pakar olahraga dengan jumlah 603 data dan terdiri dari atribut kepribadian serta faktor pendukung aktivitas olahraga. Dataset dibagi menggunakan metode *train-test split* dengan proporsi 67% data latih dan 33% data uji. Tahapan penelitian meliputi validasi data, transformasi atribut kategorikal menggunakan Ordinal *Encoding*, pembangunan model klasifikasi, serta evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan performa lebih baik dibandingkan *Decision Tree* dengan nilai *accuracy* sebesar 82,91%, *precision* 0,89, *recall* 0,83, dan *F1-score* 0,81. Sementara itu, *Decision Tree* memperoleh *accuracy* sebesar 77,89%, *precision* 0,65, *recall* 0,78, dan *F1-score* 0,70. Hasil tersebut menunjukkan bahwa pendekatan *ensemble* pada *Random* Forest mampu menangkap pola data yang lebih kompleks dan menghasilkan klasifikasi kategori olahraga yang lebih akurat. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem rekomendasi olahraga yang lebih adaptif dan personal berdasarkan karakteristik pengguna.

Kata Kunci: Big Five Personality; Rekomendasi Olahraga; Decision Tree; Random Forest; Machine Learning

Abstract–Selecting sports that match individual characteristics is an important factor in increasing motivation, comfort, and consistency in performing physical activities. Each individual has different psychological characteristics, activity preferences, exercise intensity levels, and desired exercise goals. One approach that can be used to provide more personalized sports recommendations is by utilizing personality characteristics based on the Big Five Personality (OCEAN) model. This study aims to compare the performance of the Decision Tree and Random Forest algorithms in classifying sports categories based on personality characteristics, exercise intensity, social preferences, exercise location, and exercise goals. The novelty of this research lies in the implementation and comparative evaluation of these two algorithms in a Big Five Personality-based sports recommendation system, which has not been widely developed. The dataset used was obtained from sports experts, consisting of 603 records containing personality attributes and supporting factors related to sports activities. The dataset was divided using the train-test split method with a proportion of 67% training data and 33% testing data. The research stages included data validation, categorical attribute transformation using Ordinal Encoding, classification model development, and evaluation using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results showed that the Random Forest algorithm achieved better performance than Decision Tree, with an accuracy of 82.91%, precision of 0.89, recall of 0.83, and F1-score of 0.81. Meanwhile, Decision Tree obtained an accuracy of 77.89%, precision of 0.65, recall of 0.78, and F1-score of 0.70. These results indicate that the ensemble approach in Random Forest is capable of capturing more complex data patterns and producing more accurate sports category classifications. This research is expected to serve as a foundation for developing a more adaptive and personalized sports recommendation system based on user characteristics.

Keywords: Big Five Personality; Sports Recommendation; Decision Tree; Random Forest; Machine Learning

1. PENDAHULUAN

Olahraga Olahraga merupakan aktivitas yang melibatkan gerakan fisik dan aspek mental yang bertujuan untuk menjaga serta meningkatkan kualitas kesehatan tubuh (Nurmala Hindun, Wilyanti Agustin, 2022). Aktivitas olahraga tidak hanya berperan dalam meningkatkan kebugaran fisik, tetapi juga memberikan manfaat terhadap kondisi psikologis, seperti meningkatkan konsentrasi, mengurangi stres, dan mendukung produktivitas seseorang (Hardinsyah et al., 2024; Zulkarnaen & Novita, 2025) Selain itu, aktivitas fisik yang dilakukan secara rutin dapat membantu seseorang mempertahankan kondisi kesehatan serta meningkatkan kualitas hidup. Oleh karena itu, pemilihan jenis olahraga yang sesuai dengan kondisi dan karakteristik individu menjadi salah satu faktor penting agar seseorang dapat menjalankan aktivitas olahraga secara konsisten.

Pemilihan olahraga yang sesuai dengan karakteristik individu penting untuk meningkatkan motivasi, kenyamanan, dan konsistensi dalam berolahraga. Setiap individu memiliki perbedaan dalam menentukan jenis aktivitas fisik yang diminati berdasarkan faktor psikologis, tujuan olahraga, tingkat aktivitas, serta lingkungan sosial. Salah satu pendekatan yang dapat digunakan untuk memahami karakteristik individu adalah teori *Big Five Personality Traits (OCEAN)*, yang terdiri dari *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, *dan Neuroticism*. (Azhari, 2024; Farihah & Subekti, 2025)

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

Karakteristik kepribadian memiliki hubungan terhadap kecenderungan seseorang dalam memilih suatu aktivitas. Individu dengan tingkat *extraversion* yang tinggi misalnya cenderung lebih menyukai aktivitas yang melibatkan interaksi sosial, sedangkan individu dengan tingkat *openness* yang tinggi cenderung lebih terbuka terhadap aktivitas baru dan menantang. Namun, penelitian mengenai *Big Five Personality* selama ini lebih banyak berfokus pada pengukuran karakteristik psikologis atau klasifikasi kepribadian dan belum banyak diterapkan secara khusus untuk membangun sistem rekomendasi olahraga dengan pendekatan *machine learning*. Oleh karena itu, diperlukan pendekatan komputasi yang mampu mengolah karakteristik individu menjadi suatu rekomendasi olahraga yang lebih personal.

Perkembangan teknologi kecerdasan buatan *(Artificial Intelligence/AI)* dan *machine learning* memberikan peluang untuk mengembangkan sistem rekomendasi yang lebih personal. Pendekatan *machine learning* telah banyak digunakan dalam berbagai bidang, termasuk kesehatan dan kebugaran, untuk melakukan prediksi maupun memberikan rekomendasi berdasarkan pola karakteristik pengguna. Penelitian yang dilakukan oleh (Brons et al., 2024) menunjukkan bahwa metode machine learning mampu mendukung personalisasi strategi intervensi kesehatan digital dalam meningkatkan aktivitas fisik pengguna.

Dalam penelitian di Indonesia, penerapan *machine learning* juga telah banyak digunakan untuk menyelesaikan berbagai permasalahan klasifikasi dan prediksi. Penelitian (Azmi & Voutama, 2024) menunjukkan bahwa algoritma klasifikasi seperti *Decision Tree* dan *Random Forest* mampu digunakan untuk melakukan prediksi dengan tingkat akurasi yang baik melalui tahapan *preprocessing* dan evaluasi menggunakan *confusion matrix*. Selain itu, penelitian (Utami et al., 2024) menunjukkan bahwa *Random Forest* memiliki kemampuan yang baik dalam melakukan klasifikasi pada data dengan pola yang kompleks karena menggunakan pendekatan *ensemble learning*. Hal tersebut menunjukkan bahwa machine learning dapat menjadi pendekatan yang sesuai dalam membangun sistem rekomendasi berbasis karakteristik pengguna.

Dalam penelitian ini digunakan algoritma *Decision Tree* dan *Random Forest* sebagai metode klasifikasi. *Decision Tree* dipilih karena memiliki struktur model yang mudah dipahami dan mampu menggambarkan hubungan antara variabel input dengan hasil klasifikasi. Algoritma ini bekerja dengan membentuk aturan keputusan berdasarkan proses pemisahan data menggunakan atribut tertentu sehingga menghasilkan model berbentuk pohon keputusan yang mudah diinterpretasikan. Kemampuan interpretasi tersebut menjadi salah satu keunggulan *Decision Tree* dalam proses pengambilan keputusan berbasis data. Namun, *Decision Tree* memiliki kelemahan berupa risiko *overfitting* ketika digunakan pada data dengan pola yang kompleks karena struktur pohon yang terlalu dalam dapat menyebabkan model kurang mampu melakukan generalisasi terhadap data baru (Adriansyah et al., 2022; Bagas et al., 2023).

Sebagai metode pembanding, *Random Forest* digunakan karena memiliki kemampuan untuk meningkatkan performa klasifikasi melalui penggabungan beberapa pohon keputusan. Random Forest termasuk metode *ensemble learning* yang bekerja dengan membangun beberapa *Decision Tree* dan menggabungkan hasil prediksi dari setiap pohon untuk memperoleh keputusan akhir. Pendekatan tersebut membuat *Random Forest* lebih stabil serta mampu mengurangi risiko *overfitting* dibandingkan penggunaan satu pohon keputusan saja. Penelitian (Yaman et al., 2024) menunjukkan bahwa *Random Forest* memiliki performa lebih baik dibandingkan *Decision Tree* dalam proses klasifikasi nutrisi makanan cepat saji. Selain itu, penelitian Adriansyah et al. (2022) juga menunjukkan bahwa *Random Forest* mampu memberikan hasil klasifikasi yang baik pada data dengan beberapa atribut masukan.

Berdasarkan penelitian sebelumnya, penerapan *machine learning* untuk klasifikasi telah banyak dilakukan pada berbagai bidang, tetapi penelitian yang mengimplementasikan perbandingan *Decision Tree* dan *Random Forest* dalam sistem rekomendasi olahraga berdasarkan karakteristik kepribadian masih relatif terbatas. Sebagian besar penelitian sebelumnya lebih berfokus pada klasifikasi kesehatan, pendidikan, atau prediksi berbasis data numerik. Oleh karena itu, penelitian ini dilakukan untuk mengetahui algoritma yang lebih optimal dalam menghasilkan rekomendasi olahraga berdasarkan karakteristik kepribadian pengguna.

Berdasarkan permasalahan tersebut, rumusan masalah dalam penelitian ini adalah bagaimana performa algoritma *Decision Tree* dan *Random Forest* dalam mengklasifikasikan kategori olahraga berdasarkan karakteristik kepribadian serta algoritma mana yang menghasilkan kinerja terbaik. Tujuan penelitian ini adalah menganalisis dan membandingkan performa kedua algoritma berdasarkan nilai *accuracy, precision, recall, dan F1-score.*

Kontribusi penelitian ini adalah memberikan evaluasi komparatif antara algoritma *Decision Tree* sebagai model klasifikasi tunggal dan *Random Forest* sebagai metode *ensemble* dalam sistem rekomendasi olahraga berbasis *Big Five Personality*. Hasil penelitian diharapkan dapat menjadi dasar dalam pengembangan sistem rekomendasi olahraga yang lebih personal sehingga pengguna dapat memperoleh pilihan aktivitas olahraga yang sesuai dengan karakteristik dirinya.

2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian

Penelitian ini menggunakan metode dengan menerapkan dua algoritma klasifikasi, yaitu *Decision Tree* dan *Random Forest*, untuk menentukan kategori olahraga yang paling sesuai berdasarkan karakteristik kepribadian pengguna. Data yang digunakan terdiri dari atribut kepribadian serta faktor pendukung aktivitas olahraga yang berperan sebagai variabel

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

prediktor dalam proses klasifikasi. Pendekatan ini bertujuan untuk membangun model yang mampu memberikan rekomendasi olahraga secara lebih personal sesuai dengan profil pengguna.

2.2 Sumber Data

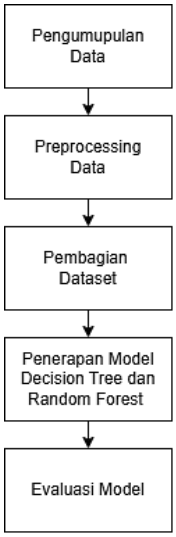
Data yang digunakan dalam penelitian ini merupakan data yang diperoleh melalui proses pengumpulan data dari pakar di bidang olahraga. Pengumpulan data dilakukan dengan menghimpun informasi mengenai hubungan antara karakteristik individu dengan kategori olahraga yang sesuai.Dataset yang diperoleh berjumlah 603 data yang berisi informasi karakteristik pengguna dan kategori olahraga. Setiap data memiliki atribut yang menggambarkan aspek kepribadian, kebiasaan olahraga, preferensi aktivitas, serta tujuan olahraga. Data tersebut kemudian digunakan sebagai dasar dalam proses pelatihan dan pengujian model klasifikasi.

Sebelum digunakan dalam proses pemodelan, dataset dilakukan proses validasi untuk memastikan kualitas dan kelayakan data. Tahap validasi meliputi pemeriksaan jumlah data, struktur atribut, tipe data, konsistensi kategori, serta pengecekan nilai kosong. Berdasarkan hasil validasi, seluruh data memiliki format yang sesuai dan tidak ditemukan nilai kosong sehingga dataset dapat digunakan untuk proses klasifikasi.

2.3 Variabel Data Penelitian

Variabel penelitian terdiri dari variabel independen sebagai atribut masukan *(input feature)* dan variabel dependen sebagai kelas target *(target class).* Variabel independen digunakan untuk merepresentasikan karakteristik pengguna yang menjadi dasar dalam proses klasifikasi, meliputi OCEAN sebagai indikator kepribadian berdasarkan model , intensitas olahraga, preferensi sosial, lokasi olahraga, dan tujuan olahraga. Sementara itu, variabel dependen yang digunakan adalah kategori olahraga yang berfungsi sebagai label klasifikasi atau hasil rekomendasi yang dihasilkan oleh model. Kategori olahraga tersebut terdiri atas delapan kelas, yaitu *Cardio, Endurance, Flexibility, Functional Training, HIIT, Mind-Body, Outdoor Adventure*, dan *Strength Training*. Seluruh variabel input digunakan untuk memprediksi kategori olahraga yang paling sesuai dengan karakteristik dan preferensi pengguna.

2.4 Tahap Penelitian



Gambar 1. Kerangka Penelitian

Tahapan penelitian pada gambar 1 menggambarkan proses yang dilakukan mulai dari pengumpulan data hingga evaluasi model. Tahapan penelitian terdiri dari beberapa proses sebagai berikut:

1. Pengumpulan data
   Tahap awal penelitian dilakukan dengan mengumpulkan dataset dari pakar olahraga. Data yang diperoleh berisi karakteristik kepribadian pengguna berdasarkan pendekatan OCEAN serta faktor pendukung dalam menentukan rekomendasi olahraga.
2. Preprocessing Data
   Tahap dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses klasifikasi. Proses ini meliputi validasi data seperti pengecekan jumlah data, struktur atribut, nilai kosong (*missing value*) dan transformasi data kategorikal menjadi numerik menggunakan metode Ordinal *Encoding*. Ordinal *Encoding* digunakan karena algoritma *Decision Tree* dan *Random Forest* membutuhkan data numerik sebagai masukan lebih tepatnya menggunakan pengodean metode labe (Guntara, 2025). Proses *encoding* diterapkan pada atribut OCEAN, intensitas olahraga, preferensi sosial, lokasi olahraga, dan tujuan olahraga.
3. Pembagian Dataset
   Dataset pada penelitian ini dibagi menjadi dua bagian menggunakan metode dengan perbandingan 67% data *training* dan 33% data . Sebanyak 67% data digunakan untuk melatih model agar dapat mempelajari pola hubungan

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

antara variabel input dan kategori olahraga yang menjadi target klasifikasi, sedangkan 33% data sisanya digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang belum pernah diproses sebelumnya. Proporsi pembagian data *training* dan yang mendekati 70:30 banyak digunakan dalam penelitian klasifikasi karena mampu menyeimbangkan proses pelatihan model dengan pengujian kinerja pada data yang belum pernah digunakan sebelumny (Muningsih et al., 2026; Refindha et al., 2025).

4. Pemodelan dan Pelatihan
   Pada tahap pemodelan, algoritma *Decision Tree* diterapkan menggunakan dua kriteria pemisahan data, yaitu dan dengan parameter *max_depth* = 3 dan *random_state* = 0 untuk membandingkan performa model dalam melakukan klasifikasi kategori olahraga. Selain itu, digunakan pula algoritma *Random Forest* sebagai metode yang menggabungkan beberapa pohon keputusan untuk meningkatkan stabilitas prediksi, dengan parameter *n_estimators* = 10 dan 100 serta *random_state* = 0 guna menjaga konsistensi hasil pemodelan.
5. Evaluasi Model
   Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja klasifikasi yang dihasilkan. Selain itu, *Confusion Matrix* digunaan digunakan untuk menggambarkan performa model klasifikasi dengan melihat kesesuaian antara hasil prediksi model terhadap label atau kelas sebenarnya. Evaluasi ini terdiri dari empat komponen utama, yaitu True Positive (TP) sebagai prediksi benar pada kelas positif, True Negative (TN) sebagai prediksi benar pada kelas negatif, False Positive (FP) sebagai kesalahan prediksi ketika data negatif diklasifikasikan sebagai positif, dan False Negative (FN) sebagai kegagalan model dalam mengenali kelas positif (Setiadi & Sulistyo, 2025). Seluruh proses implementasi dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab dengan memanfaatkan *library Pandas*, *NumPy*, *Scikit-Learn*, *Category Encoders*, dan *Matplotlib*. *Google Colab* digunakan sebagai lingkungan pemrograman berbasis *cloud* yang mendukung eksekusi kode *Python* sehingga mempermudah proses pengolahan data, pembangunan model klasifikasi, dan evaluasi performa algoritma (jaenal arifin, titania dwiandini, styorini, abdulloh eizzi irsyada, 2023)

Penelitian ini menerapkan dua algoritma klasifikasi, yaitu *Decision Tree* dan *Random Forest*, karena keduanya mampu mengolah data kategorikal serta menghasilkan model yang dapat dibandingkan untuk menentukan performa terbaik. *Decision Tree* dipilih karena mudah dipahami dan diinterpretasikan, sedangkan *Random Forest* memiliki keunggulan dalam meningkatkan akurasi prediksi serta meminimalkan *overfitting* melalui metode (Rian oktafiani, arief hermawan, 2026; Yaman et al., 2024)

2.5 *Decision Tree*

Algoritma *Decision Tree* termasuk metode yang populer dalam *data mining* karena kemampuannya yang baik dalam mengklasifikasikan dan memprediksi suatu data. Algoritma ini bekerja dengan merepresentasikan sekumpulan data dan aturan keputusan ke dalam bentuk struktur pohon, sehingga hubungan antar atribut dapat dipahami dengan lebih mudah. Pohon keputusan membagi data secara bertahap menjadi kelompok-kelompok yang lebih kecil berdasarkan atribut tertentu. Dalam struktur tersebut, simpul akar dan simpul internal digunakan untuk melakukan pengujian terhadap atribut, sedangkan simpul daun merepresentasikan hasil akhir atau kelas yang diprediksi . *Gini Index* merupakan salah satu ukuran evaluasi yang digunakan untuk mengukur tingkat ketidakmurnian *(impurity)* suatu data dalam proses pemisahan pada algoritma pohon keputusan. Metrik ini membantu menentukan pemilihan atribut terbaik dengan melihat seberapa efektif suatu pembagian data dalam menghasilkan kelompok kelas yang lebih homogen (Bere et al., 2026; Septhya et al., 2023)**.**

2.6 Random Forest

Random Forest merupakan metode klasifikasi yang terdiri atas beberapa pohon keputusan yang dibangun menggunakan data dan fitur yang dipilih secara acak (Adriansyah et al., 2022; Danny akhmad zulfikar & Vivi aida fitria, 2025) Pada proses pembentukannya, setiap pohon dibangun menggunakan sampel data dan fitur yang dipilih secara acak melalui teknik *bootstrap sampling* dan *Featur baging*. Hasil prediksi akhir diperoleh berdasarkan mekanisme *majority* pada kasus klasifikasi atau nilai rata-rata pada kasus regresi. *Random Forest* memiliki keunggulan dalam menangani data dengan banyak atribut, mengurangi risiko *overfitting*, serta menyediakan informasi mengenai tingkat pentingnya setiap fitur dalam proses prediksi. Namun, algoritma ini umumnya menghasilkan model yang lebih kompleks dengan ukuran yang lebih besar dan membutuhkan waktu prediksi yang lebih lama dibandingkan penggunaan satu pohon keputusan saja (Astri & Kamal, 2025)**.**

2.7 Evaluasi Model

Pengujian dilakukan untuk mengetahui akurasi dan kinerja model pada penelitian dalam melakukan klasifikasi kategori olahraga berdasarkan data yang digunakan . Pengukuran dilakukan menggunakan beberapa metrik evaluasi yang umum diterapkan pada permasalahan klasifikasi terutama *Confusion Matrix* adalah tabel evaluasi yang menunjukkan perbandingan antara hasil prediksi dan kondisi aktual pada model klasifikasi (Septhya et al., 2023). Matriks ini berukuran n × n, di mana n menunjukkan jumlah kelas yang digunakan dalam proses klasifikasi. Melalui *Confusion Matrix*, tingkat ketepatan prediksi dan kesalahan klasifikasi pada setiap kelas dapat diketahui secara lebih rinci sehingga kinerja model dapat dievaluasi dengan lebih baik. *Confusion Matrix* untuk n=2ditampilkan dalam Tabel 1.

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

Tabel 1. Representasi Confusion Matrix

Kelas	Prediksi Positif	Prediksi Negatif
Jumlah positive sebenanrnya	Jumlah True Positives (TP)	Jumlah False Negatives (FN)
Jumlah negative sebenanrnya	Jumlah True Positives (TP)	Jumlah True Negatives (TN)

Accuracy, *precision*, *recall*, dan *F1-score* merupakan metrik evaluasi yang digunakan untuk mengukur serta membandingkan kinerja algoritma klasifikasi. Nilai dari masing-masing metrik tersebut diperoleh berdasarkan hasil confusion matriks pada fitur dataset dan dihitung menggunakan persamaan (1) sampai (4).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$f1-score = 2x\frac{Precision\ x\ Recall}{Precision+\ Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Pada bagian ini membahas hasil dari pengujian menggunakan metode *algoritma decision tree* dan *Random Forest* pada dataset yang di peroleh dari pakar olahraga. Dataset ini dimanfaatkan sebagai bahan untuk melatih sekaligus menguji model klasifikasi. Implementasi model dilakukan dengan mengacu pada kode program yang diambil dari repositori *GitHub* milik Prashant Banerjee, kemudian dimodifikasi agar sesuai dengan kebutuhan penelitian. Penyesuaian tersebut mencakup tahapan praproses data, pembangunan model, hingga evaluasi performa.

3.1 Praproses Data

Tahap praproses data dilakukan untuk mempersiapkan data sebelum digunakan dalam proses klasifikasi. Dataset yang digunakan terdiri dari 603 data dengan enam atribut, yaitu OCEAN, Intensitas Olahraga, Preferensi Sosial, Lokasi Olahraga, Tujuan Olahraga, dan Kategori Olahraga. Seluruh atribut bertipe kategorikal dan tidak ditemukan nilai kosong sehingga tidak diperlukan proses penanganan *missing value*. Selanjutnya dilakukan pemisahan atribut fitur dan label. Lima atribut digunakan sebagai fitur prediktor, sedangkan atribut Kategori Olahraga digunakan sebagai kelas target. Data kemudian dibagi menjadi data latih sebesar 67% dan data uji sebesar 33% menggunakan metode dengan *random state* 42. Karena seluruh atribut berupa data kategorikal, dilakukan proses *encoding* menggunakan Ordinal Encoder agar dapat diproses oleh algoritma *Decision Tree* dan *Random Forest*.

Selain proses pemisahan data dan *encoding*, tahap praproses memiliki peran penting dalam memastikan kualitas data yang digunakan pada proses pembelajaran mesin. Dataset yang digunakan dalam penelitian ini memiliki karakteristik yang cukup baik karena tidak ditemukan data yang hilang maupun duplikasi data. Kondisi tersebut memberikan keuntungan karena model dapat langsung mempelajari pola hubungan antar atribut tanpa harus melalui proses imputasi atau pembersihan data yang kompleks. Seluruh data yang digunakan berasal dari hasil validasi pakar olahraga sehingga setiap kombinasi atribut memiliki keterkaitan yang logis dengan kategori olahraga yang direkomendasikan.

Distribusi data pada setiap kategori olahraga juga menunjukkan kondisi yang seimbang. Masing-masing kategori olahraga memiliki jumlah data yang sama sehingga potensi bias model terhadap kelas tertentu dapat diminimalkan. Keseimbangan data merupakan faktor penting dalam proses klasifikasi karena model tidak akan cenderung mempelajari satu kelas secara berlebihan. Dengan demikian, hasil evaluasi yang diperoleh dapat mencerminkan kemampuan model yang sebenarnya dalam mengenali seluruh kategori olahraga yang tersedia.

Penggunaan Ordinal Encoder dipilih karena seluruh atribut pada dataset berbentuk kategorikal. Proses *encoding* mengubah setiap kategori menjadi representasi numerik sehingga dapat diproses oleh algoritma *Decision Tree* maupun *Random Forest*. Meskipun representasi angka diberikan pada setiap kategori, algoritma berbasis pohon keputusan tidak mengasumsikan adanya hubungan matematis antar nilai tersebut. Oleh karena itu, penggunaan *Ordinal Encoder* tetap dapat menghasilkan performa yang baik pada penelitian ini. Hasil praproses yang optimal menjadi fondasi penting bagi proses pembentukan model karena kualitas data yang baik akan memudahkan algoritma dalam menemukan pola klasifikasi yang tepat.

3.2 Pembangunan Model

Pada penelitian ini dilakukan pembangunan dua model klasifikasi, yaitu *Decision Tree* dan *Random Forest*, untuk membandingkan kemampuan keduanya dalam mengklasifikasikan kategori olahraga berdasarkan karakteristik pengguna. Model *Decision Tree* dibangun menggunakan kriteria pemisahan dengan parameter *max_depth* sebesar 3 dan *random_state* 0. Pembatasan kedalaman pohon diterapkan untuk mengurangi risiko serta menghasilkan model yang lebih sederhana dan mudah diinterpretasikan. Sementara itu, model *Random Forest* dibangun menggunakan pendekatan *ensemble* dengan parameter *n_estimators* = 300, *max_depth* = 10, *min_samples_split* = 5, *min_samples_leaf*

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

= 2, *class_weight* = balanced, dan *random_state* = 0. Penggunaan banyak pohon keputusan pada *Random Forest* memungkinkan proses prediksi dilakukan melalui mekanisme voting dari beberapa model, sehingga dapat meningkatkan stabilitas, kemampuan generalisasi, dan akurasi klasifikasi dibandingkan penggunaan satu pohon keputusan saja.

3.3 Evaluasi Model

Tabel 2. Hasil Evaluasi Model

Model	Accuracy	Precission	Recall	F1-Score
Random Forest	82,91%	0,89%	0,83%	0,81%
Decission Tree	77,89%	0,65%	0,78%	0,70%

Tabel 2 tersebut menunjukkan bahwa *Random Forest* menghasilkan performa yang lebih baik dibandingkan *Decision Tree* pada seluruh metrik evaluasi. Akurasi *Random Forest* mencapai 82,91%, lebih tinggi sekitar 5,02% dibandingkan *Decision Tree* yang memperoleh akurasi 77,89%. Selain itu, nilai *precission* sebesar 0,89 menunjukkan bahwa prediksi yang dihasilkan lebih tepat dibandingkan *Decision Tree* yang hanya memperoleh *precision* sebesar 0,65. Dari sisi *recall*, *Random Forest* juga memiliki kemampuan yang lebih baik dalam mengenali data pada setiap kelas dengan nilai 0,83, sedangkan *Decision Tree* memperoleh nilai 0,78. Peningkatan performa yang paling terlihat terdapat pada nilai *F1_Score*, yaitu 0,81 pada dan 0,70 pada *Decision Tree*. Hasil ini menunjukkan bahwa *Random Forest* memiliki keseimbangan yang lebih baik antara *precision* dan *recall* sehingga mampu memberikan performa klasifikasi yang lebih optimal. Penggunaan banyak pohon keputusan pada *Random Forest* memungkinkan model menangkap pola data yang lebih kompleks dan mengurangi kesalahan klasifikasi yang masih sering terjadi pada *Decision Tree*, terutama pada kelas *Cardio* dan *Mind-Body*.

Perbedaan performa antara *Random Forest* dan *Decision Tree* menunjukkan adanya pengaruh yang signifikan dari pendekatan *ensemble* terhadap kualitas klasifikasi. *Decision Tree* bekerja dengan membangun satu pohon keputusan berdasarkan atribut yang dianggap paling informatif pada setiap proses pemisahan data. Pendekatan ini menghasilkan model yang sederhana dan mudah dipahami, namun memiliki keterbatasan dalam menangkap variasi pola yang kompleks pada dataset. Ketika terdapat beberapa kelas dengan karakteristik yang saling berdekatan, model cenderung menghasilkan batas keputusan yang kurang optimal sehingga meningkatkan kemungkinan terjadinya kesalahan klasifikasi. Sebaliknya, Random Forest membangun ratusan pohon keputusan yang dilatih menggunakan subset data dan atribut yang berbeda. Setiap pohon memberikan prediksi secara independen, kemudian hasil akhir ditentukan melalui mekanisme voting. Pendekatan ini membuat model lebih *robust* terhadap variasi data dan mampu mengurangi pengaruh kesalahan yang mungkin terjadi pada satu pohon keputusan. Oleh karena itu, Random Forest mampu menghasilkan nilai accuracy, precision, recall, dan F1-score yang lebih tinggi dibandingkan Decision Tree. Nilai *precision* sebesar 0,89 pada *Random Forest* menunjukkan bahwa sebagian besar prediksi yang diberikan model sesuai dengan kelas sebenarnya. Hal ini penting dalam sistem rekomendasi olahraga karena kesalahan rekomendasi dapat menyebabkan pengguna memperoleh kategori olahraga yang kurang sesuai dengan karakteristiknya. Sementara itu, nilai *recall* sebesar 0,83 menunjukkan bahwa sebagian besar data pada setiap kategori olahraga berhasil dikenali dengan baik oleh model. Kombinasi *precision* dan *recall* yang tinggi menghasilkan nilai *F1-score* sebesar 0,81 yang menunjukkan keseimbangan performa klasifikasi secara keseluruhan. Hasil tersebut mengindikasikan bahwa *Random Forest* lebih mampu memanfaatkan informasi yang terdapat pada atribut *OCEAN*, intensitas olahraga, preferensi sosial, lokasi olahraga, dan tujuan olahraga dalam menentukan kategori olahraga yang tepat. Kemampuan ini sangat penting karena hubungan antar atribut dalam dataset tidak selalu bersifat linear dan sering kali melibatkan kombinasi beberapa karakteristik sekaligus. Dengan demikian, penggunaan Random Forest menjadi pilihan yang lebih tepat ketika tujuan utama penelitian adalah memperoleh akurasi prediksi yang tinggi dan kemampuan generalisasi yang baik.

3.4 *Confusion Matrix*

Untuk memberikan gambaran kinerja masing-masing model, hasil *Confusion Matrix* dari empat algoritma yang digunakan disajikan pada Tabel 3.

Tabel 3. Hasil Confusion Matrix

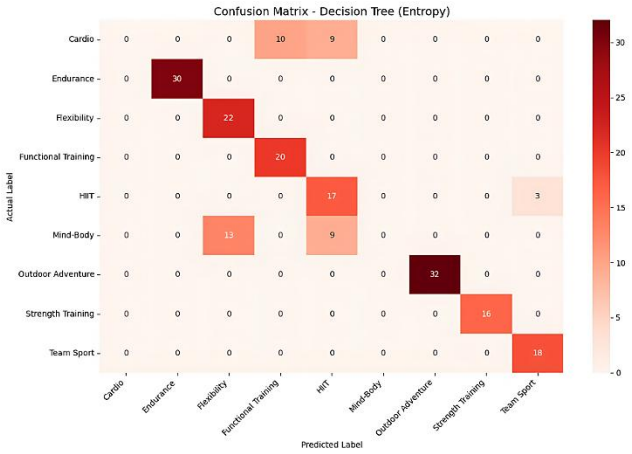
Model	True Positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)
Random Forest	155	34	34	1558
Decission Tree	165	44	44	1548

3.4.1 Decission Tree

Dari hasil pengujian pada Gambar 2, beberapa kelas seperti *Endurance*, *Flexibility*, *Functional Training*, *Outdoor Adventure*, *Strength Training*, dan Team Sport berhasil diprediksi dengan sangat baik. Hal ini terlihat dari banyaknya data yang berada pada diagonal utama *Confusion Matrix*, yang menunjukkan prediksi sesuai dengan kelas sebenarnya. Namun, model masih mengalami kesulitan dalam mengidentifikasi kelas *Cardio* dan *Mind-Body*. Pada kelas *Cardio*, seluruh data uji salah diklasifikasikan ke kelas *Functional Training* dan *HIIT*. Sementara itu, pada kelas *Mind-Body*

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

sebagian besar data diprediksi sebagai *Flexibility* dan *HIIT*. Kondisi ini menunjukkan bahwa karakteristik data pada kedua kelas tersebut memiliki kemiripan dengan kelas lain sehingga sulit dipisahkan oleh model *Decision Tree*.



Gambar 2. Matriks Decission Tree

Tabel 4. Hasil Training dan Testing

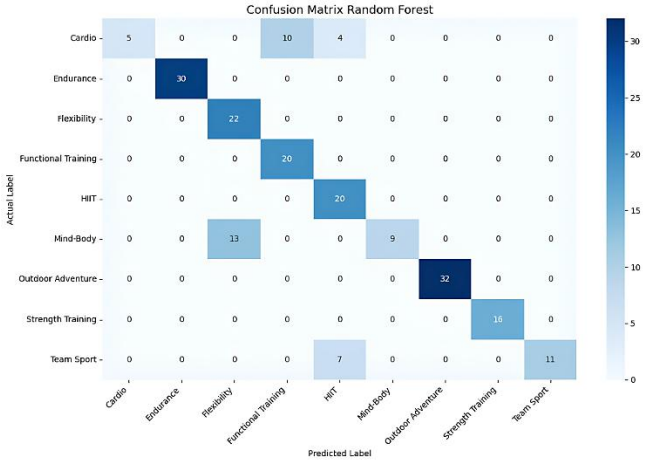
	Training set Score	*Testing set Score*
Gini Index	0.6584	0.6332
Entropy	0.7475	0.7789

Tabel 4 menunjukkan perbandingan performa *Decision Tree* menggunakan kriteria *Gini Index* dan *Entropy*. Pada metode *Gini Index* diperoleh nilai *training set score* sebesar 65,84% dan *set score* sebesar 63,32%. Selisih yang relatif kecil antara kedua nilai tersebut menunjukkan bahwa model tidak mengalami , tetapi kemampuan klasifikasinya masih tergolong rendah.

Sementara itu, metode *Entropy* menghasilkan *training set score* sebesar 74,75% dan *set score* sebesar 77,89%. Hasil ini menunjukkan bahwa *Entropy* mampu menghasilkan pemisahan data yang lebih baik dibandingkan *Gini Index*. Bahkan nilai *set score* sedikit lebih tinggi daripada *training set score*, yang mengindikasikan bahwa model memiliki kemampuan generalisasi yang cukup baik terhadap data yang belum pernah dilihat sebelumnya. Oleh karena itu, kriteria *Entropy* dipilih sebagai model *Decision Tree* terbaik pada penelitian ini.

Rendahnya performa pada kelas *Cardio* dan *Mind-Body* disebabkan oleh tingginya kemiripan karakteristik kedua kelas tersebut dengan kategori olahraga lainnya. Kelas *Cardio* memiliki pola yang mirip dengan *Functional Training* dan *HIIT* karena sama-sama berfokus pada peningkatan kebugaran dan daya tahan tubuh, sehingga model sering salah mengklasifikasikan data *Cardio* ke dalam kedua kategori tersebut. Sementara itu, kelas *Mind-Body* memiliki karakteristik yang serupa dengan *Flexibility* karena sama-sama menekankan aspek relaksasi, keseimbangan, dan fleksibilitas tubuh, sehingga banyak data *Mind-Body* diprediksi sebagai *Flexibility*. Selain itu, keterbatasan atribut yang digunakan dalam proses klasifikasi menyebabkan model kurang mampu menangkap perbedaan spesifik antar kategori olahraga, yang pada akhirnya berdampak pada rendahnya performa prediksi pada kedua kelas tersebut.

3.4.2 Random Forest



Gambar 3. Matriks Random Forest

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

Pada Gambar 3 Sebagian besar data pada setiap kategori olahraga berhasil diklasifikasikan dengan benar. Kelas *Endurance*, *Outdoor Adventure*, dan *Strength Training* memperoleh hasil prediksi yang sangat baik dengan nilai mencapai 100%, yang menunjukkan bahwa seluruh data pada kelas tersebut berhasil dikenali oleh model. Meskipun masih terdapat kesalahan klasifikasi pada kelas *Cardio* dan *Mind-Body*, jumlah kesalahan tersebut lebih sedikit dibandingkan *Decision Tree*. Pada kelas *Cardio*, model berhasil mengidentifikasi 5 data dengan benar dari total 19 data uji. Sedangkan pada kelas *Mind-Body*, model berhasil mengklasifikasikan 9 data dengan benar dari total 22 data uji. Hasil ini menunjukkan bahwa penggunaan banyak pohon keputusan pada mampu meningkatkan kemampuan model dalam mengenali pola data yang kompleks dan mengurangi kesalahan prediksi.

Tabel 5. Hasil Akurasi Random Forest

	Accuracy score with 10 decision-trees	*Accuracy score* with 100 decision-trees
Random Forest	0.8422	0.8422

Tabel 5 menunjukkan hasil pengujian menggunakan dua jumlah pohon keputusan yang berbeda, yaitu 10 dan 100 *decision tree*s. Kedua konfigurasi menghasilkan *acuracy score* yang sama sebesar 84,42%. Hasil ini menunjukkan bahwa model telah mencapai performa yang stabil bahkan ketika hanya menggunakan 10 pohon keputusan. Penambahan jumlah pohon hingga 100 tidak memberikan peningkatan akurasi yang signifikan. Hal ini mengindikasikan bahwa pola dalam dataset sudah dapat dipelajari dengan baik oleh model sejak jumlah pohon yang lebih sedikit. Meskipun demikian, penggunaan 100 pohon tetap memberikan keuntungan berupa kestabilan prediksi yang lebih baik karena keputusan akhir diperoleh dari agregasi lebih banyak pohon keputusan.

Berdasarkan hasil confusion matrix, Random Forest mampu mempertahankan performa klasifikasi yang baik pada sebagian besar kategori olahraga. Kelas Endurance, Outdoor Adventure, dan Strength Training memperoleh nilai recall sempurna, yang menunjukkan bahwa seluruh data pada kategori tersebut berhasil dikenali dengan tepat oleh model. Hal ini menunjukkan bahwa kombinasi atribut pada kategori tersebut memiliki pola yang lebih jelas sehingga lebih mudah dibedakan oleh algoritma.

Namun, masih terdapat beberapa kesalahan klasifikasi pada kategori Cardio dan Mind-Body. Pada kelas Cardio, sebagian data diprediksi sebagai Functional Training dan HIIT karena ketiga kategori tersebut memiliki karakteristik yang serupa, terutama dalam tujuan meningkatkan kebugaran fisik dan intensitas latihan. Meskipun demikian, Random Forest tetap mampu mengenali sebagian data Cardio dengan benar, sedangkan Decision Tree mengalami kesulitan dalam membedakan kategori tersebut.

Kesalahan klasifikasi juga terjadi pada kelas Mind-Body yang sebagian diprediksi sebagai Flexibility karena kedua kategori memiliki kemiripan dalam aspek keseimbangan tubuh, relaksasi, dan fleksibilitas gerakan. Perbedaan hasil antara kedua algoritma menunjukkan bahwa mekanisme ensemble pada Random Forest mampu menangkap pola data yang lebih kompleks dibandingkan Decision Tree tunggal.

Secara keseluruhan, penggunaan Random Forest memberikan performa yang lebih stabil dalam mengklasifikasikan kategori olahraga. Meskipun peningkatan jumlah pohon keputusan tidak menghasilkan perubahan akurasi yang signifikan, pendekatan ensemble tetap mampu mengurangi variasi model dan meningkatkan kemampuan generalisasi. Hal ini menunjukkan bahwa Random Forest lebih efektif dalam memanfaatkan hubungan antara karakteristik kepribadian, preferensi olahraga, dan kategori olahraga dibandingkan Decision Tree.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* menghasilkan performa klasifikasi yang lebih baik dibandingkan *Decision Tree* dalam sistem rekomendasi olahraga berdasarkan karakteristik kepribadian. *Random Forest* memperoleh nilai *accuracy* sebesar 82,91%, *precision* sebesar 0,89, *recall* sebesar 0,83, dan *F1-score* sebesar 0,81, sedangkan *Decision Tree* memperoleh *accuracy* sebesar 77,89%, *precision* sebesar 0,65, *recall* sebesar 0,78, dan *F1-score* sebesar 0,70. Perbedaan performa tersebut menunjukkan bahwa pendekatan *ensemble* pada *Random Forest* mampu memberikan kemampuan generalisasi yang lebih baik dalam menangani hubungan kompleks antara karakteristik kepribadian, preferensi olahraga, dan kategori olahraga.

Hasil penelitian ini sejalan dengan penelitian (Yaman et al., 2024) yang menunjukkan bahwa *Random Forest* memiliki performa lebih unggul dibandingkan *Decision Tree* karena mampu menggabungkan beberapa pohon keputusan sehingga menghasilkan prediksi yang lebih stabil. Hal serupa juga ditemukan oleh (Utami et al., 2024), dimana *Random Forest* mampu meningkatkan kinerja klasifikasi melalui mekanisme *ensemble* yang dapat mengurangi risiko *overfitting* dan mengenali pola data yang lebih kompleks. Selain itu, penelitian (Azmi & Voutama, 2024) menunjukkan bahwa kedua algoritma tersebut dapat diterapkan pada permasalahan klasifikasi dengan evaluasi *confusion matrix*, namun *Random Forest* cenderung memberikan hasil yang lebih konsisten pada dataset dengan banyak atribut yang saling berhubungan.

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada penerapan algoritma *Decision Tree* dan Random Forest untuk sistem rekomendasi olahraga berbasis *Big Five Personality*. Sebagian besar penelitian terdahulu menerapkan algoritma tersebut pada bidang kesehatan, prediksi, maupun klasifikasi umum, sedangkan penelitian ini menggunakannya untuk menghasilkan rekomendasi kategori olahraga berdasarkan kombinasi faktor psikologis dan preferensi aktivitas pengguna. Meskipun Random Forest memberikan hasil yang lebih baik, beberapa kategori seperti


TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

Cardio dan *Mind-Body* masih mengalami kesalahan klasifikasi akibat kemiripan karakteristik dengan kategori lain seperti *Functional Training*, *HIIT*, dan *Flexibility*.

Secara keseluruhan, hasil penelitian memperkuat temuan sebelumnya bahwa *Random Forest* merupakan algoritma yang efektif untuk klasifikasi dengan pola data yang kompleks. Penerapan metode ini mampu meningkatkan kemampuan model dalam mengenali hubungan antara karakteristik kepribadian, preferensi olahraga, dan kategori olahraga yang direkomendasikan. Penambahan atribut pengguna seperti usia, kondisi kesehatan, tingkat kebugaran, dan pengalaman olahraga dapat menjadi pengembangan selanjutnya untuk meningkatkan akurasi sistem rekomendasi.

4. KESIMPULAN

Penelitian ini berhasil membandingkan performa algoritma *Decision Tree* dan *Random Forest* dalam mengklasifikasikan kategori olahraga berdasarkan karakteristik kepribadian *Big Five Personality*, intensitas olahraga, preferensi sosial, lokasi olahraga, dan tujuan olahraga. Hasil pengujian menunjukkan bahwa *Random Forest* memiliki kinerja yang lebih baik dibandingkan *Decision Tree* dengan nilai *accuracy* sebesar 82,91%, *precision* 0,89, *recall* 0,83, dan *F1-score* 0,81, sedangkan *Decision Tree* dengan kriteria *Entropy* memperoleh *accuracy* 77,89%, *precision* 0,65, *recall* 0,78, dan *F1-score* 0,70. Temuan ini menunjukkan bahwa pendekatan *ensemble* pada *Random Forest* lebih efektif dalam menangkap pola data yang kompleks sehingga menghasilkan kemampuan klasifikasi yang lebih baik untuk sistem rekomendasi olahraga berbasis kepribadian. Meskipun demikian, penelitian ini masih memiliki keterbatasan, yaitu jumlah dataset yang relatif terbatas sebanyak 603 data, sumber data yang berasal dari pengetahuan pakar olahraga, serta atribut yang masih berfokus pada faktor kepribadian dan preferensi olahraga sehingga belum sepenuhnya merepresentasikan karakteristik pengguna yang beragam. Selain itu, beberapa kategori seperti *Cardio* dan *Mind-Body* masih menunjukkan tingkat kesalahan klasifikasi yang cukup tinggi akibat kemiripan karakteristik dengan kategori olahraga lainnya. Oleh karena itu, penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan beragam, menambahkan atribut seperti usia, jenis kelamin, kondisi kesehatan, tingkat kebugaran, dan riwayat aktivitas olahraga, serta mengeksplorasi algoritma lain seperti *XGBoost*, *LightGBM*, maupun metode *ensemble* yang lebih kompleks dengan optimasi hyperparameter untuk meningkatkan performa model klasifikasi.

REFERENCES

Adriansyah, I., Mahendra, M. D., Rasywir, E., & Pratama, Y. (2022). Perbandingan Metode Random Forest Classifier dan SVM Pada Klasifikasi Kemampuan Level Beradaptasi Pembelajaran Jarak Jauh Siswa. *Bulletin of Informatic and Data Scince*, *1*(2), 98–103. https://doi.org/https://doi.org/10.61944/bids.v1i2.49

Astri, R., & Kamal, A. (2025). *Pengembangan Sistem Rekomendasi Program Studi Multikelas Menggunakan Algoritma Random Forest TIN : Terapan Informatika Nusantara*. *6*(5), 567–577. https://doi.org/10.47065/tin.v6i5.8369

Azhari. (2024). Teori Big Five Personality Dalam Ilmu Psikologi Dan Relevansinya Dengan Konsep Kepribadian Manusia Perspektif Al-Qur'an. *Jurnal An-Nur*, *13*(1), 50–59. https://doi.org/https://doi.org/10.24014/an-nur.v13i1.28258

Azmi, A. F., & Voutama, A. (2024). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix Azel. *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika*, *13*(1). https://doi.org/https://doi.org/10.34010/komputa.v13i1.12639

Bagas, M., Darmawan, A., Dewanta, F., & Astuti, S. (2023). Analisis Perbandingan Algoritma Decision Tree , Random Forest , dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot Comparative Analysis of Decision Tree , Random Forest , and Naïve Bayes Algorithm for Flood Prediction at Dayeuhkolot Village. *TELKA - Telekomunikasi Elektronika Komputasi Dan Kontrol*, *9*(1), 52–61. https://doi.org/https://doi.org/10.15575/telka.v9n1.52-61

Bere, Y. R., Informatika, T., & Timur, M. J. (2026). Perbandingan Metode Decision Tree dan Logistic Regression dalam Klasifikasi Tingkat Obesitas Berdasarkan Gaya Hidup. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, *5*(2), 746–756. https://doi.org/http://dx.doi.org/10.35889/jutisi.v15i2.3591

Brons, A., Wang, S., Visser, B., Kr, B., & Bakkes, S. (2024). Machine Learning Methods to Personalize Persuasive Strategies in mHealth Interventions That Promote Physical Activity: Scoping Review and Categorization Overview. *Journal Of Medical Internet Research*, *26*(5), 1–20. https://doi.org/10.2196/47774

Danny akhmad zulfikar, & Vivi aida fitria. (2025). Detecting Online Gambling Promotion on Social Media with Random Forest. *International Conference on Business, Innovation, Technology & Science (IcoBits)*, *01*(02), 410–421. https://doi.org/https://doi.org/10.32664/icobits.v1.50

Farihah, L., & Subekti, P. (2025). ObeCheck Sebagai Platform Penerapan Metode K-Nearest Neighbors untuk Klasifikasi Obesitas Berbasis Website. *Building of Informatics, Technology and Science (BITS)*, *6*(4), 2311–2321. https://doi.org/10.47065/bits.v6i4.6726

Guntara, M. (2025). Komparasi Kinerja Label-Encoding dengan One-Hot-Encoding pada Algoritma K-Nearest Neighbor menggunakan Himpunan Data Campuran Performance Comparison of Label-Encoding with One-Hot-Encoding Methode on K-Nearest Neighbor Algorithm with Mixed Type Data Set. *JIKO (Jurnal Informatika Dan Komputer)*, *9*(2), 352–360. https://doi.org/https://doi.org/10.26798/jiko.v9i2.1605

Hardinsyah, Sulistiyono, & Nugroho, S. (2024). Pengaruh Olahraga Dalam Pembentukan Karakter Remaja: Literature


TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 813-822
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10343

Review. *JURNAL DUNIA PENDIDIKAN*, *5*(1), 244–255. https://doi.org/https://doi.org/10.55081/jurdip.v5i1.2609

jaenal arifin, titania dwiandini, styorini, abdulloh eizzi irsyada, rina dewi indahsari. (2023). Pelatihan Pemrograman Bahasa Python Pada Jurusan Perangkat Lunak Dan Gim. *Jurnal Pengabdian Masyarakat - Teknologi Digital Indonesia*, *2*(2), 42–48. https://doi.org/10.26798/jpm.v2i2.880

Muningsih, E., Handayani, V. R., & Galuhwardani, C. (2026). Optimasi Metode Klasifikasi dengan Particle Swarm Optimization (PSO) dan Perbandingan Split Data untuk Prediksi Penyakit Diabetes. *Journal of Artificial Intelligence and Technology Information*, *4*(2), 259–268. https://doi.org/https://doi.org/10.23917/khif.v4i2.6825

Nurmala Hindun, Wilyanti Agustin, S. (2022). Sosialisasi Pentingnya Aktivitas Fisik untuk Meningkatkan Kesehatan Mental Para Pekerja PT. Global Collection Malang. *Anfatama: Jurnal Pengabdian Masyarakat*, *1*(3), 34–38. https://doi.org/https://doi.org/10.572349/anfatama.v1i3

Refindha, F., Harianto, A., Alawi, Z., & Aristia, I. (2025). Pengaruh Komposisi Split Data Pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning. *Jurnal Sistem Informasi Dan Informatika (Simika)*, *8*(1), 36–44. https://doi.org/https://doi.org/10.47080/simika.v8i1.3663

Rian oktafiani, arief hermawan, donny avianto. (2026). Max Depth Impact on Heart Disease Classification: Decision Tree and Random Forest. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(158), 160–168. https://doi.org/https://doi.org/10.29207/resti.v8i1.5574

Septhya, D., Rahayu, K., Rabbani, S., & Fitria, V. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru Dhini. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(1), 15–19. https://doi.org/https://doi.org/10.57152/malcom.v3i1.591

Setiadi, E., & Sulistyo, D. A. (2025). Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning. *Jurnal Fasilkom*, *15*(3), 507–513. https://doi.org/https://doi.org/10.37859/jf.v15i3.10546

Utami, N., Baihaqi, K. A., Awal, E. E., & Wahiddin, D. (2024). Analisis Kinerja Algoritma Decision Tree Dan Random Forest Dalam Klasifikasi Penyakit Kardiovaskular. *Building of Informatics, Technology and Science (BITS)*, *6*(2), 970–980. https://doi.org/https://doi.org/10.47065/bits.v6i2.5722

Yaman, N. I., Juwita, A. R., Arum, S., Lestari, P., & Faisal, S. (2024). Perbandingan Kinerja Algoritma Decision Tree dan Random Forest untuk Klasifikasi Nutrisi pada Makanan Cepat Saji. *Jurnal Algoritma*, *21*(2), 184–195. https://doi.org/10.33364/algoritma/v.21-2.1649

Zulkarnaen, M. I., & Novita, P. (2025). Analisis Pilihan Jenis Olahraga Favorit di Kalangan Siswa Kelas IX SMPN 23 Tangerang Selatan. *Semnasfip*, *2*(2), 2539–2546. https://jurnal.umj.ac.id/index.php/SEMNASFIP/issue/view/1086