

Prediksi Minat Siswa Berdasarkan Tes OCEAN dan *Aptitude* Menggunakan Algoritma *Random Forest*

Muhamad Sidik¹, Andri Kurniawan², Fariz Faqih³, Farizi Ilham^{4*}

¹⁻⁴Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspittek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ¹sidikwf@gmail.com, ²andrikurniawansarolangun@gmail.com, ³farizfaqih82@gmail.com, ⁴dosen02954@unpam.ac.id

(* : coressponding author)

Abstrak—Minat merupakan faktor penting dalam proses belajar karena berpengaruh terhadap motivasi dan keberhasilan siswa. Namun, di tingkat sekolah dasar, banyak siswa belum memahami kecenderungan minat mereka. Untuk itu, proyek kerja praktik di SDN 1 Talagahiang ini bertujuan mengembangkan sistem prediksi minat siswa berbasis Tes Kepribadian OCEAN dan Tes Aptitude menggunakan algoritma klasifikasi Random Forest. Data yang digunakan bersumber dari dataset publik Kaggle dan melalui tahap pra-pemrosesan sebelum dilatih menggunakan model machine learning. Model yang dibangun kemudian diintegrasikan ke dalam aplikasi web interaktif berbasis Streamlit, sehingga siswa dapat mengisi tes secara daring dan langsung memperoleh hasil prediksi minat mereka. Sistem ini diharapkan dapat membantu guru dalam memberikan bimbingan yang lebih tepat serta mendorong siswa untuk mengenali potensi dirinya sejak dini. Proyek ini juga menjadi bentuk kontribusi mahasiswa Universitas Pamulang dalam mendukung digitalisasi pendidikan di tingkat sekolah dasar.

Kata Kunci: Minat Siswa, *Random Forest*, Tes OCEAN, Tes *Aptitude*, Prediksi Minat, *Machine Learning*, *Streamlit*

Abstract—Interest is an important factor in the learning process as it affects student motivation and success. However, at the primary school level, many students do not understand their interest tendencies. For this reason, this practical work project at SDN 1 Talagahiang aims to develop a student interest prediction system based on the OCEAN Personality Test and Aptitude Test using the Random Forest classification algorithm. The data used is sourced from the Kaggle public dataset and goes through a pre-processing stage before being trained using a machine learning model. The built model is then integrated into an interactive web application based on Streamlit, so that students can fill out the test online and immediately get the results of their interest prediction. This system is expected to help teachers provide more appropriate guidance and encourage students to recognize their potential early on. This project is also a form of Pamulang University students' contribution in supporting the digitalization of education at the elementary school level.

Keywords: Student Interest, *Random Forest*, OCEAN Test, *Aptitude Test*, Interest Prediction, *Machine Learning*, *Streamlit*

1. PENDAHULUAN

Minat merupakan salah satu faktor penting dalam proses pembelajaran karena dapat memengaruhi motivasi, perhatian, dan keterlibatan siswa dalam menerima materi pendidikan (Syah, 2016). Pemetaan minat yang dilakukan sejak dini dapat membantu sekolah dan orang tua untuk mengarahkan siswa ke bidang yang sesuai dengan kecenderungan dan potensinya.

Namun, di tingkat sekolah dasar, mayoritas siswa belum dapat mengidentifikasi potensi dan minatnya secara mandiri. Oleh karena itu, dibutuhkan pendekatan sistematis yang dapat membantu proses pengenalan diri siswa. Salah satu pendekatan yang dapat diterapkan adalah melalui tes kepribadian OCEAN dan Tes *Aptitude*, yang kemudian diproses secara otomatis menggunakan algoritma *machine learning*, salah satunya *Random Forest*.

Salah satu teori kepribadian yang paling banyak digunakan dalam dunia kerja adalah Teori Sifat Kepribadian Model Lima Besar atau *Big Five Personality Traits Model* yang diperkenalkan oleh psikolog terkenal Lewis Goldberg. Teori ini mengelompokkan kepribadian ke dalam lima dimensi utama, yaitu *Openness* (keterbukaan terhadap pengalaman baru), *Conscientiousness* (kesadaran dan kedisiplinan), *Extraversion* (kecenderungan untuk bersosialisasi), *Agreeableness* (keramahan dan kerja sama), serta *Neuroticism* (kestabilan emosional). Untuk memudahkan pengingat, kelima dimensi ini sering disingkat dengan akronim *OCEAN*.

Tes *aptitude* adalah jenis tes yang dibuat untuk menilai kemampuan individu dalam area tertentu, seperti kemampuan berpikir, pemecahan masalah, atau keterampilan teknis. Tes ini

biasanya digunakan dalam proses seleksi kerja, pendidikan, serta pengembangan karier guna mengevaluasi potensi dan kecocokan seseorang dengan bidang atau pekerjaan tertentu.

Random Forest merupakan algoritma klasifikasi berbasis *ensemble learning* yang menggabungkan banyak *decision tree* untuk meningkatkan akurasi prediksi (Breiman, 2001). Algoritma ini efektif digunakan untuk pemodelan data pendidikan karena mampu menangani banyak fitur dan bekerja baik pada dataset berukuran kecil hingga sedang (Rokach, 2016).

Penelitian ini bertujuan untuk membangun sistem prediksi minat siswa dengan menggabungkan hasil tes *OCEAN* dan *Aptitude* sebagai input, yang kemudian diklasifikasikan menggunakan algoritma *Random Forest*. Sistem dikembangkan dalam bentuk aplikasi web berbasis *Streamlit* agar dapat diakses secara mudah oleh siswa dan guru di lingkungan sekolah.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik yang tersedia di situs Kaggle. Dataset tersebut berisi hasil Tes Kepribadian OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism) dan Tes Aptitude yang mencakup kemampuan verbal, numerik, logika, serta spasial. Data ini digunakan sebagai dasar pelatihan dan pengujian dalam membangun sistem klasifikasi minat siswa.

2.2 Pra-pemrosesan Data

Tahapan pra-pemrosesan data dilakukan untuk memastikan kualitas data sebelum digunakan dalam pelatihan model. Langkah pertama adalah membersihkan data dari nilai kosong atau tidak relevan. Hal ini dilakukan dengan menghapus baris atau kolom yang memiliki banyak nilai hilang (missing values) serta menghapus data duplikat yang dapat menurunkan kualitas dataset. Selanjutnya, data kategorikal diubah ke bentuk numerik karena sebagian besar algoritma machine learning hanya dapat memproses data numerik. Proses ini dilakukan menggunakan teknik label encoding, di mana setiap kategori diubah menjadi nilai numerik.

Untuk mengatasi ketidakseimbangan jumlah data antar kelas, digunakan teknik oversampling bernama Synthetic Minority Over-sampling Technique (SMOTE). SMOTE bekerja dengan cara membuat sampel sintetis baru dari kelas minoritas, bukan hanya menyalin data yang ada. Teknik ini memilih sampel secara acak dari kelas minoritas, lalu menciptakan data baru dengan interpolasi terhadap tetangga terdekatnya (k-nearest neighbors). Dengan demikian, variasi data meningkat dan distribusi antar kelas menjadi lebih seimbang tanpa memicu overfitting. Setelah semua proses tersebut, data dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji. Pembagian ini bertujuan untuk mengevaluasi performa model secara objektif terhadap data yang belum pernah digunakan sebelumnya.

2.3 Pembangunan Model

Model klasifikasi dibangun menggunakan algoritma Random Forest yang termasuk dalam metode ensemble learning berbasis decision tree. Algoritma ini dipilih karena mampu menangani fitur yang kompleks, memiliki akurasi tinggi, dan tidak mudah overfitting. Proses pelatihan dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka scikit-learn.

2.4 Evaluasi Model

Evaluasi model dilakukan dengan menghitung akurasi menggunakan data uji. Hasil prediksi model dibandingkan dengan label aktual untuk mengetahui seberapa akurat model dalam mengklasifikasikan minat siswa. Pengukuran dilakukan menggunakan confusion matrix dan metrik akurasi sebagai indikator utama.

2.5 Implementasi Sistem

Model yang telah dilatih kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Streamlit. Aplikasi ini memungkinkan siswa mengisi Tes OCEAN dan Aptitude secara daring dan langsung mendapatkan hasil prediksi minat setelah tes selesai. Tampilan antarmuka dibuat sederhana agar mudah digunakan oleh siswa dan guru di sekolah dasar.

3. ANALISA DAN PEMBAHASAN

Setelah proses pengumpulan dan pra-pemrosesan data, serta pelatihan model klasifikasi menggunakan algoritma Random Forest, tahap selanjutnya adalah melakukan evaluasi terhadap kinerja model dan implementasi sistem dalam bentuk aplikasi web. Evaluasi ini mencakup pengukuran akurasi, analisis klasifikasi setiap kelas minat, visualisasi confusion matrix, serta tampilan antarmuka dari sistem berbasis Streamlit yang telah dibangun.

3.1 Hasil Evaluasi Model

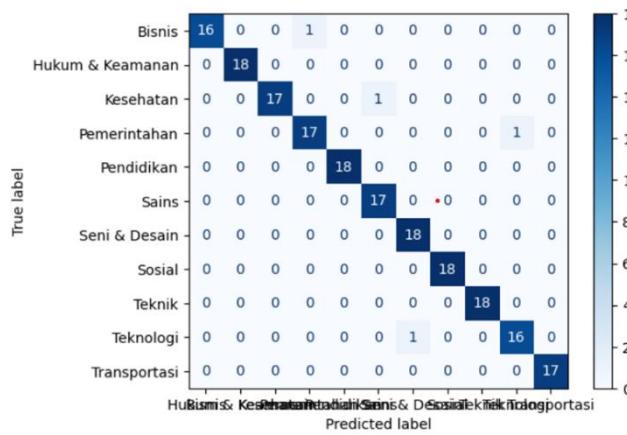
Model klasifikasi minat siswa yang dibangun menggunakan algoritma Random Forest menunjukkan performa yang sangat baik. Berdasarkan hasil evaluasi, model memperoleh akurasi sebesar 97,42%, yang menunjukkan bahwa sebagian besar prediksi sesuai dengan label sebenarnya. Evaluasi lebih rinci terhadap masing-masing kelas dapat dilihat pada Gambar 1 yang memuat nilai *precision*, *recall*, dan *f1-score* untuk setiap kategori minat. Kategori seperti Pendidikan, Sosial, Hukum & Keamanan, dan Transportasi memiliki nilai evaluasi sempurna (1.00) di semua metrik, menandakan model sangat akurat dalam mengenali pola dari kategori tersebut. Beberapa kelas seperti Teknologi, Kesehatan, dan Bisnis memiliki sedikit kesalahan prediksi, namun tetap menunjukkan nilai *f1-score* di atas 0.90, yang tergolong tinggi.

Laporan Klasifikasi:					
	precision	recall	f1-score	support	
Bisnis	1.00	0.88	0.94	17	
Hukum & Keamanan	1.00	1.00	1.00	18	
Kesehatan	1.00	0.94	0.97	18	
Pemerintahan	0.94	0.94	0.94	18	
Pendidikan	1.00	1.00	1.00	18	
Sains	0.94	1.00	0.97	17	
Seni & Desain	0.95	1.00	0.97	18	
Sosial	1.00	1.00	1.00	18	
Teknik	1.00	1.00	1.00	18	
Teknologi	0.89	0.94	0.91	17	
Transportasi	1.00	1.00	1.00	17	
accuracy			0.97	194	
macro avg	0.98	0.97	0.97	194	
weighted avg	0.98	0.97	0.97	194	

Gambar 1. Evaluasi Model

3.2 Confusion Matrix

Confusion matrix adalah metode evaluasi yang digunakan untuk memvisualisasikan performa model klasifikasi. Pada matrix ini, baris (horizontal) mewakili true label atau kelas sebenarnya dari data, sementara kolom (vertikal) menunjukkan predicted label atau hasil prediksi yang diberikan oleh model.



Gambar 2. Confusion Matrix

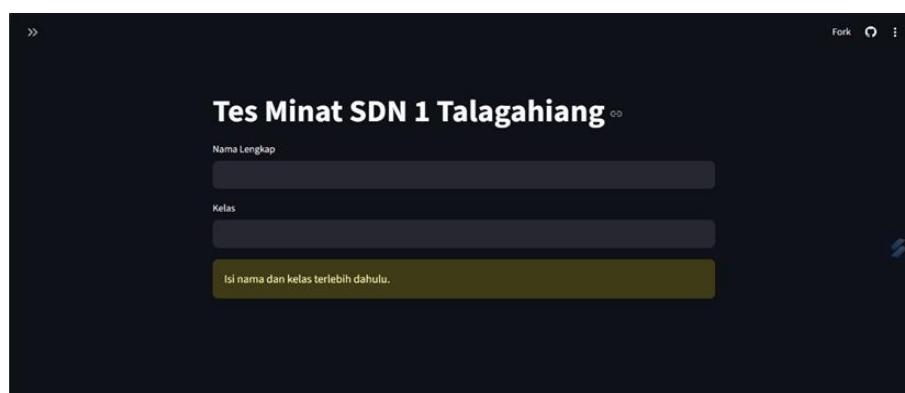
Jika model melakukan prediksi dengan benar, maka nilai akan muncul pada diagonal utama dari kiri atas ke kanan bawah. Sebaliknya, nilai di luar diagonal menunjukkan jumlah kesalahan prediksi, yaitu ketika model memprediksi label yang tidak sesuai dengan kenyataannya. pada Gambar 2 terlihat bahwa:

- a. Sebanyak 16 data dari kelas "Bisnis" berhasil diklasifikasikan dengan benar sebagai Bisnis, namun 1 data diklasifikasikan keliru sebagai "Kesehatan".
- b. Kelas "Teknologi" memiliki 1 data yang salah diklasifikasikan sebagai "Seni & Desain".
- c. Sementara kelas lain seperti "Sosial", "Teknik", dan "Transportasi" menunjukkan hasil klasifikasi yang sempurna (tanpa kesalahan).

Secara keseluruhan, distribusi prediksi yang dominan berada pada diagonal utama menunjukkan bahwa model Random Forest yang digunakan mampu mengklasifikasikan minat siswa dengan tingkat akurasi tinggi dan kesalahan yang sangat minim, sesuai dengan hasil evaluasi sebelumnya.

3.3 Tampilan Antarmuka Sistem

Sistem prediksi minat yang dikembangkan diimplementasikan dalam bentuk aplikasi web menggunakan *framework Streamlit*, sehingga dapat diakses secara daring oleh siswa dan guru. Antarmuka dirancang dengan sederhana agar mudah digunakan oleh siswa sekolah dasar. Terdapat tiga komponen utama dalam tampilan sistem.



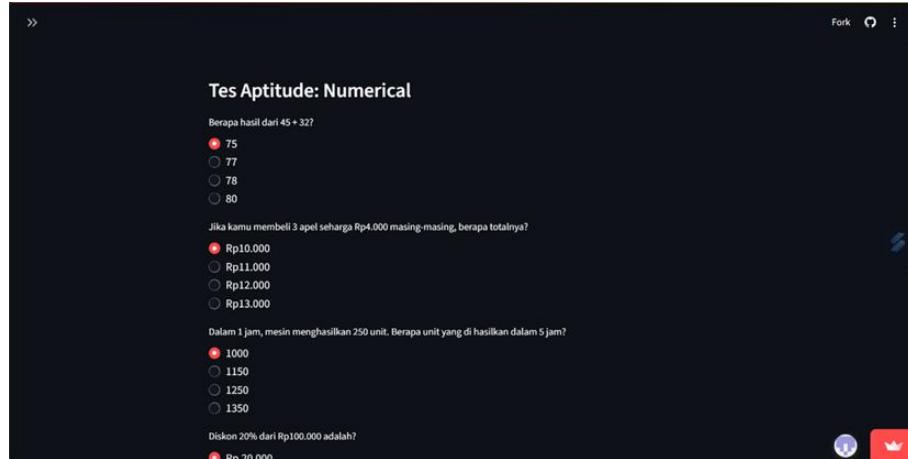
Gambar 3. Halaman Utama

Gambar 3 menampilkan halaman utama aplikasi, di mana siswa diminta untuk mengisi nama lengkap dan kelas sebelum memulai tes. Input ini menjadi identitas awal dalam proses prediksi.



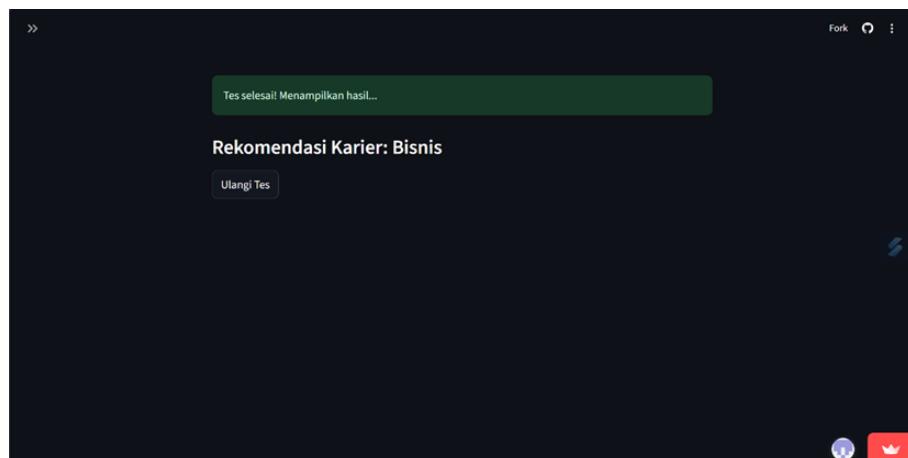
Gambar 4. Soal Tes OCEAN

Setelah mengisi data diri, siswa diarahkan ke halaman soal seperti terlihat pada Gambar 4. Pada halaman ini, siswa mengisi Tes Kepribadian *OCEAN* dengan memilih skala *Likert* dari “Sangat Tidak Setuju” hingga “Sangat Setuju” untuk setiap pernyataan.



Gambar 5. Tes Soal Aptitude

Setelah Mengisi tes *OCEAN* siswa melanjutkan mengisi tes soal *Aptitude* seperti Gambar 5. Proses ini berlanjut hingga seluruh pernyataan selesai dijawab.



Gambar 6. Hasil Tes

Setelah semua soal diisi, sistem akan secara otomatis memproses jawaban melalui model klasifikasi *Random Forest* yang telah dilatih. Hasil akhir ditampilkan pada Gambar 6, berupa rekomendasi kategori minat atau karier, misalnya "Bisnis", "Teknik", atau "Seni & Desain". Tombol "Ulang Tes" juga tersedia untuk melakukan pengujian ulang.

4. KESIMPULAN

sistem prediksi minat siswa berbasis Tes OCEAN dan Aptitude berhasil dikembangkan dan diimplementasikan dalam bentuk aplikasi web menggunakan Streamlit. Model klasifikasi Random Forest yang digunakan mampu memberikan hasil prediksi dengan akurasi tinggi, mencapai 97,42%, berdasarkan data uji dan mampu mengklasifikasikan minat siswa ke dalam 11 kategori secara konsisten dan presisi. Evaluasi menggunakan confusion matrix dan classification report membuktikan bahwa sebagian besar prediksi berada pada kelas yang tepat dengan kesalahan yang sangat minimal. Implementasi sistem dalam bentuk aplikasi berbasis Streamlit memberikan kemudahan akses bagi siswa dan guru untuk melakukan tes dan mendapatkan hasil secara langsung. Sistem ini diharapkan dapat digunakan secara aktif oleh pihak sekolah dalam membantu proses bimbingan dan pengenalan potensi siswa sejak dini. Selain itu, proyek ini juga menjadi bentuk kontribusi mahasiswa Universitas Pamulang dalam mendukung transformasi digital di bidang pendidikan dasar.

REFERENCES

- Syah, M. (2016). *Psikologi Pendidikan dengan Pendekatan Baru*. Remaja Rosdakarya.
- Nugroho, A. (2020). Penerapan Machine Learning untuk Klasifikasi Data Pendidikan. *Jurnal Teknologi Informasi dan Pendidikan*, 13(2), 111–117.
- Sari, D. P., & Lestari, M. (2021). Sistem Rekomendasi Minat Belajar Menggunakan Algoritma KNN. *Jurnal Teknologi dan Sistem Komputer*, 9(3), 145–150.
- Yuliana, N., & Suprapto, B. (2022). Penerapan Tes Aptitude dalam Menentukan Minat Siswa. *Jurnal Psikologi Pendidikan*, 7(2), 90–97.
- Suryani, D., & Mahendra, F. (2021). Analisis Penerapan Streamlit dalam Visualisasi Data Machine Learning. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 112–119.
- Prasetyo, R. P., & Anindita, S. (2023). Implementasi Algoritma Random Forest untuk Prediksi Karier Berdasarkan Tes Minat. *Jurnal Ilmiah Komputer dan Informatika (JIKI)*, 10(2), 55–63.
- Santosa, H., & Fajrina, N. (2020). Big Five Personality Test dan Aplikasinya dalam Rekomendasi Jurusan Kuliah. *Jurnal Psikologi dan Teknologi*, 3(1), 25–33.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.