

Analisis Implementasi Sistem Prediksi Minat dan Jurusan SMA pada Siswa SMPN 1 Panongan Menggunakan *Random Forest* Berbasis Website

Ruskandi Dinda Anugrah^{*1}, Zurnan Alfian, S.Kom., M.Kom.²

^{1,2,3}Universitas Pamulang; JL. Raya Puspitek No. 46, Buaran, Kec. Serpong, Kota Tangerang Selatan
Banten 15310, (021) 7412566

^{1,2,3}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

e-mail: ¹ruskandidinda.a@gmail.com, ²dosen02678@unpam.ac.id

Abstrak

Pemilihan jurusan SMA merupakan keputusan kritis bagi siswa SMP yang sering menghadapi dilema karena bergantung pada preferensi pribadi dan pengaruh eksternal. Penelitian ini bertujuan mengembangkan sistem prediksi minat dan jurusan SMA berbasis website menggunakan algoritma *Random Forest* untuk membantu siswa SMPN 1 Panongan membuat keputusan yang lebih objektif. Metode penelitian menggunakan pendekatan kuantitatif dengan data 347 sampel nilai rapor semester genap kelas 9. Tahap pra-pemrosesan data meliputi rekayasa fitur melalui proyeksi linear, *oversampling SMOTE* untuk mengatasi ketidakseimbangan kelas, dan normalisasi Min-Max. Sistem dikembangkan menggunakan Python dengan framework Flask untuk backend dan Bootstrap untuk frontend. Hasil pengujian menunjukkan algoritma *Random Forest* mencapai akurasi 78%, *precision* 78%, *recall* 78%, dan *F1-Score* 78%, lebih unggul dibanding *Decision Tree* (74%), *Gaussian Naive Bayes* (60%), dan *Multinomial Naive Bayes* (59%). Pengujian *Black Box* memvalidasi semua fitur sistem berfungsi dengan baik. Berbeda dari penelitian sebelumnya yang hanya memprediksi jurusan berdasarkan minat, penelitian ini mengombinasikan nilai akademik dan minat siswa dalam sistem berbasis web. Sistem ini terbukti efektif memberikan prediksi jurusan yang objektif berbasis data akademik, meskipun keputusan akhir tetap berada di tangan siswa.

Kata kunci— *Random Forest*, *Prediksi Jurusan*, *Sistem Berbasis Website*, *Machine Learning*, *Penjurusan Siswa*

I. PENDAHULUAN

Pendidikan membentuk individu untuk beradaptasi dengan perkembangan zaman. Fase kritis dalam pendidikan adalah saat siswa SMP memasuki jenjang SMA, mereka harus menentukan arah yang selaras dengan minat, potensi, dan kemampuan pribadi (Pribadi et al. 2021). Sesuai teori Howard Gardner tentang kecerdasan majemuk, setiap individu memiliki kombinasi kecerdasan berbeda sehingga pendekatan pendidikan idealnya disesuaikan dengan potensi unik masing-masing siswa (Auralia et al. 2025).

Namun, siswa sering menghadapi dilema besar dalam memilih jurusan yang sesuai dengan minat mereka, berdampak pada perencanaan karier masa depan yang tidak optimal. Penelitian menunjukkan bahwa perencanaan karier yang baik melibatkan pemahaman diri, minat, dan kemampuan (Sutrisno et al. 2024). Dalam praktiknya, guru Bimbingan Konseling memberikan bimbingan melalui diskusi langsung dengan siswa. Namun, proses pemilihan

jurusan masih bergantung pada preferensi pribadi siswa dan pengaruh eksternal seperti keinginan orang tua atau teman, yang dapat mengakibatkan pemilihan jurusan tidak selalu sejalan dengan minat siswa.

Nilai akademik seperti nilai rapor mencerminkan kinerja siswa dan dapat menggambarkan potensi serta kemampuan mereka. Penelitian menunjukkan nilai rapor, terutama mata pelajaran tertentu seperti matematika, terbukti menjadi indikator utama dalam memprediksi prestasi akademik siswa (Gunawan et al. 2024). Dengan menggunakan data akademik, sistem dapat menghasilkan prediksi jurusan yang lebih objektif dan relevan karena mencerminkan langsung hasil belajar siswa.

Perkembangan teknologi *Artificial Intelligence (AI)* membuka peluang baru dalam mendukung proses pengambilan keputusan di sektor pendidikan (Ambarita dan Nurrahmatullah 2024). *Machine Learning* sebagai subbidang *AI* memungkinkan sistem belajar dan membuat keputusan berdasarkan data yang ada (Prairi et al. 2025). Salah satu metode yang dapat digunakan adalah algoritma *Random*

Forest yang menggunakan klasifikasi berbasis kumpulan pohon keputusan dengan keunggulan akurat, stabil, dan tahan terhadap overfitting (Breiman 2001).

Berdasarkan observasi pada SMPN 1 Panongan, saat ini belum terdapat sistem yang dapat membantu siswa menentukan pilihan minat dan jurusan SMA berdasarkan data akademik. Penelitian ini bertujuan menganalisis implementasi sistem prediksi minat dan jurusan SMA berbasis website menggunakan algoritma Random Forest untuk memberikan solusi yang membantu siswa memilih jurusan sesuai minat dan potensi akademik mereka.

II. METODE PELAKSANAAN

Penelitian ini menggunakan pendekatan kuantitatif dan rekayasa perangkat lunak dengan fokus penerapan algoritma Random Forest menggunakan Python. Tahapan penelitian dimulai dari pengumpulan data yang dilakukan melalui wawancara dan observasi di SMPN 1 Panongan dengan guru BK untuk memahami proses pemetaan minat siswa. Data akademik berupa nilai rapor semester genap kelas 9 dari 347 siswa digunakan sebagai dataset. Data mencakup 10 mata pelajaran yang kemudian diolah menjadi 4 fitur agregat melalui proyeksi linear, yaitu Rata-rata IPA (Matematika dan IPA), Rata-rata IPS (PKN dan IPS), Rata-rata Bahasa (Bahasa Indonesia dan Inggris), serta Rata-rata Keseluruhan.

Tahap pra-pemrosesan data meliputi tiga proses utama yang dipilih untuk meningkatkan kualitas model prediksi. Pertama, rekayasa fitur (Feature Engineering) dilakukan untuk mengondensasi 10 nilai mata pelajaran menjadi 4 fitur rata-rata yang lebih representatif terhadap karakteristik jurusan. Kedua, *oversampling* SMOTE diterapkan untuk mengatasi ketidakseimbangan distribusi kelas yang dapat menyebabkan bias prediksi terhadap kelas mayoritas, menghasilkan 105 sampel sintetis untuk kelas minoritas (IPS) sehingga meningkatkan total sampel dari 347 menjadi 452. Ketiga, normalisasi Min-Max Scaling diterapkan untuk menyeragamkan skala fitur, khususnya pada model yang sensitif terhadap perbedaan skala seperti Multinomial Naive Bayes dengan rentang nilai [1, 100].

Sistem dikembangkan menggunakan Python dengan Flask sebagai web *framework* dan scikit-learn untuk implementasi algoritma *Machine Learning*. *Frontend* dibangun menggunakan HTML, CSS, JavaScript, dan *Bootstrap* untuk menghasilkan antarmuka responsif. Database *MySQL* digunakan untuk penyimpanan data, sedangkan Visual Studio Code menjadi development tool utama. Fitur sistem mencakup registrasi dan login siswa, kuisioner minat, input nilai rapor, prediksi jurusan, riwayat tes, serta dashboard untuk admin dan superadmin.

Model Random Forest dipilih karena kemampuannya menangani data kompleks dan tahan terhadap overfitting. Evaluasi dilakukan menggunakan confusion matrix dengan metrik akurasi, precision, recall, dan F1-Score. Data dibagi dengan rasio 70:30 untuk pelatihan dan pengujian. Perbandingan dilakukan dengan Decision Tree, Gaussian Naive Bayes, dan Multinomial Naive Bayes untuk memvalidasi keunggulan Random Forest. Pengujian Black

Box dilakukan untuk memvalidasi fungsionalitas sistem dari perspektif pengguna.

III. HASIL DAN PEMBAHASAN

Sistem berhasil dikembangkan dengan antarmuka berbasis website yang responsif dan mudah digunakan. Halaman utama menyediakan akses untuk tiga jenis pengguna yaitu siswa, admin, dan superadmin, masing-masing dengan fitur sesuai perannya.

Tabel 1. Fitur Sistem Berdasarkan Peran Pengguna

Peran	Fitur Utama
Siswa	Registrasi, Login, Dashboard, Kuisioner Minat, Input Nilai Rapor, Hasil Prediksi, Riwayat Tes, Profile, Ubah Password
Admin	Login, Dashboard, Data Siswa, Laporan Analisis, Ubah Password
Superadmin	Login, Dashboard, Manajemen Akun, Manajemen Siswa, Kelola Klaster Bobot, Pengaturan Kuisioner, Kelola Prediksi, Laporan

Proses prediksi dimulai dengan siswa mengisi kuisioner minat yang menghasilkan analisis kluster minat mencakup Saintek, Soshum, dan Bahasa. Selanjutnya siswa menginput nilai rapor yang diproses menggunakan model Random Forest untuk menghasilkan prediksi jurusan berupa IPA, IPS, atau Bahasa beserta skor kepercayaan prediksi.

Evaluasi kinerja dilakukan pada empat algoritma menggunakan dataset yang sama dengan pembagian 70% data latih dan 30% data uji.

Tabel 2. Komparasi Kinerja Algoritma

Algoritma	Akurasi	Precision (Avg)	Recall (Avg)	F1-Score (Avg)
Random Forest	78%	78%	78%	78%
Decision Tree	74.73%	74.5%	74.5%	75%
Gaussian Naive Bayes	60%	60.5%	60.5%	60%
Multinomial Naive Bayes	59.34%	59.5%	59.5%	59%

Random Forest menunjukkan kinerja terbaik dengan akurasi 78% dan metrik evaluasi yang konsisten. Analisis per kelas menunjukkan performa yang berbeda untuk masing-masing kategori jurusan.

Tabel 3. Metrik Evaluasi Random Forest Per Kelas

Kelas	Precision	Recall
IPA	80%	73%
IPS	76%	82%

Confusion matrix Random Forest menunjukkan dari 91 sampel uji, 71 sampel diprediksi benar yang terdiri dari 33 sampel IPA dan 38 sampel IPS, dengan 20 sampel salah prediksi yang terdiri dari 12 *false negative* dan 8 *false positive*. Model lebih tepat memprediksi jurusan IPA dengan precision tinggi namun lebih sensitif mendeteksi jurusan IPS dengan recall tinggi.

Pengujian *Black Box* dilakukan pada 20 skenario yang mencakup semua fitur sistem untuk tiga jenis pengguna.

Tabel 4. Ringkasan Hasil Pengujian Black Box

Kategori Pengujian	Jumlah Skenario	Hasil Valid
Fitur Siswa	10	10 (100%)
Fitur Admin	4	4 (100%)
Fitur Superadmin	6	6 (100%)

Semua skenario pengujian menunjukkan hasil valid dengan tingkat keberhasilan 100%, membuktikan sistem berfungsi sesuai spesifikasi dan memberikan pengalaman pengguna yang intuitif.

Hasil ini sejalan dengan penelitian Haiza et al. (2023) yang menunjukkan keunggulan *Random Forest* untuk klasifikasi penjurusan siswa dengan akurasi mencapai 93,14%, serta penelitian Mu'Alim dan Hidayati (2022) yang mencapai akurasi 94,38%. Keunggulan *Random Forest* dalam penelitian ini konsisten dengan temuan tersebut, meskipun akurasi 78% masih dapat ditingkatkan melalui beberapa cara. Pertama, penambahan fitur dengan mengintegrasikan data ekstrakurikuler, tes psikologi, atau minat siswa sebagai fitur tambahan. Kedua, optimasi hyperparameter dengan melakukan tuning parameter seperti jumlah pohon (*n_estimators*), kedalaman maksimal (*max_depth*), dan kriteria split untuk meningkatkan kinerja model. Ketiga, penambahan data dengan memperluas dataset melalui pelibatan lebih banyak sekolah atau angkatan untuk meningkatkan generalisasi model.

Teknik SMOTE yang diterapkan berhasil mengatasi ketidakseimbangan kelas dengan meningkatkan kelas minoritas dari 121 menjadi 226 sampel. Namun, penelitian lanjutan dapat mengeksplorasi teknik oversampling lain atau kombinasi dengan undersampling untuk hasil optimal. Sistem berbasis website memudahkan akses bagi siswa dan pihak sekolah. Pengembangan aplikasi mobile di masa depan dapat meningkatkan fleksibilitas akses sistem.

IV. SIMPULAN

Penelitian ini berhasil mengembangkan sistem prediksi minat dan jurusan SMA berbasis website menggunakan algoritma *Random Forest* di SMPN 1 Panongan. Sistem menyediakan fitur lengkap mulai dari registrasi, kuisioner

minat, input nilai rapor, hingga prediksi jurusan dengan antarmuka yang intuitif.

Algoritma *Random Forest* terbukti unggul dibanding algoritma pembandingan dengan akurasi 78%, precision 78%, recall 78%, dan F1-Score 78%. Model menunjukkan kinerja baik dalam memprediksi kedua kelas jurusan dengan precision IPA 80% dan IPS 76%, serta recall IPA 73% dan IPS 82%.

Pengujian *Black Box* memvalidasi semua fitur sistem berfungsi sesuai spesifikasi. Sistem berbasis data yang dikembangkan mampu memberikan prediksi jurusan objektif berdasarkan analisis nilai akademik, membantu siswa membuat keputusan yang lebih terinformasi mengenai pilihan jurusan SMA mereka.

DAFTAR PUSTAKA

- Ambarita N, Nurrahmatullah MF. 2024. Impacts of Artificial Intelligence on Student Learning: A Systematic Literature Review. *Jurnal VARIDIKA*. 36(1):13-30.
- Auralia FA, Ismayani, Lesmana G. 2025. Mengoptimalkan Potensi Siswa melalui Pendekatan Multiple Intelligences dalam Pembelajaran. *Al-Watzikhoebillah: Kajian Islam Dan Pendidikan*. 11(1):236-243.
- Breiman L. 2001. *Random Forest*. *Machine Learning*. 45(1):5-32.
- Gunawan MT, Tine JY, Murwaningtyas CE. 2024. Prediksi Prestasi Akademik Menggunakan *Random Forest*. *Jurnal Pendidikan Informatika dan Sains*. 13(2):141-153.
- Haiza, Elmayati, Zulius, Oktafia H. 2023. Penerapan Algoritma *Random Forest* Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo. *Jurnal Teknik Informatika*. 8(2):115-124.
- Mu'Alim F, Hidayati R. 2022. Implementasi Metode *Random Forest* Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang. *Jupiter*. 14(1):116-125.
- Prairie MZ, Alfian Z, Harefa K. 2025. Kecocokan Keputusan Pohon Algoritma pada Kimia Organik: Perbandingan ROC AUC Keputusan Pohon dan Ketetangaan. *Journal of Innovative and Creativity*. 5(2):10689-10700.
- Pribadi AS, Erlangga E, Wangge MY. 2021. Hubungan Antara Konsep Diri Akademik dengan Pengambilan Keputusan Karier pada Siswa SMP. *PHILANTHROPY: Journal of Psychology*. 5(1):157-167.
- Sutrisno, Afrinaldi, Efrienni. 2024. Pentingnya Perencanaan Karir bagi Siswa Kelas IX di SMP Kamang Manggek dengan Menggunakan Bimbingan Klasikal. *JISPENDIORA Jurnal Ilmu Sosial Pendidikan Dan Humaniora*. 3(1):53-58.