

**ANALISIS REKOMENDASI JURUSAN KULIAH BAGI
PELAJAR SMA DENGAN MENGGUNAKAN
ALGORITMA RANDOM FOREST**

SKRIPSI



**MUHAMMAD ALDI RIZKY PRATAMA
2055201096**

**TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS LANCANG KUNING
PEKANBARU**

2026

HALAMAN PERSETUJUAN

Nama : MUHAMMAD ALDI RIZKY PRATAMA
NIM : 2055201096
Program Studi : Teknik Informatika
Judul : Analisis Rekomendasi Jurusan Kuliah Bagi Pelajar SMA
Dengan Menggunakan Algoritma Random Forest

Pekanbaru, 12 Februari 2026

Disetujui Oleh,
Pembimbing



Dr. SUTEJO, M.Kom., MTA
NIDN. 1002068801
Mengetahui,

Dekan

Fakultas Ilmu Komputer



Dr. Yogi Yunefri, M.Kom., MTA., MCE
NIDN. 1022068803

Ketua Program Studi

Teknik Informatika



Dr. Yogi Ersan Fadrial, M.Kom., MTA
NIDN. 1027109202

HALAMAN PENGESAHAN

Nama : MUHAMMAD ALDI RIZKY PRATAMA
NIM : 2055201096
Program Studi : Teknik Informatika
Judul : Analisis Rekomendasi Jurusan Kuliah Bagi Pelajar SMA
Dengan Menggunakan Algoritma Random Forest

Skripsi ini telah dipertahankan di depan Dewan Penguji pada tanggal sebagai persyaratan untuk mendapatkan gelar Sarjana Komputer (S.Kom)

Dewan Penguji,

Ketua



Dr. SUTEJO, M.Kom., MTA
NIDN. 1002068801

Anggota



Dr. Muhamad Sadar, S.E, M.Kom., MTA., MCF
NIDN. 1030096501

Anggota



Pandu Pratama Putra, M.Kom., MTA., MCF
NIDN. 1003069101

Mengetahui :

Dekan

Fakultas Ilmu Komputer



Dr. Yogi Yunefri, M.Kom., MTA., MCE
NIDN. 1022068803

Ketua Program Studi

Teknik Informatika



Dr. Yogi Ersan Fadrial, M.Kom., MTA
NIDN. 1027109202

PERNYATAAN KEASLIAN KARYA ILMIAH

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat Karya/Pendapat yang pernah ditulis oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka

Apabila ditemukan suatu Jiplakan/Plagiat maka saya bersedia menerima akibat berupa sanksi Akademis dan sanksi lain yang diberikan oleh yang berwenang sesuai ketentuan peraturan dan perundang-undangan yang berlaku

Pekanbaru, 12 Februari 2026



MUHAMMAD ALDI RIZKY PRATAMA

2055201096

HALAMAN PERSEMBAHAN

Dengan menyebut nama Allah Subhanahu wa ta'ala, segala puji dan syukur penulis panjatkan atas rahmat, taufik, dan hidayah-Nya sehingga skripsi ini dapat diselesaikan. Dengan penuh rasa hormat dan cinta, karya sederhana ini penulis persembahkan kepada:

1. Ayah dan Ibu tercinta Atas kasih sayang yang tidak pernah putus, doa yang tidak pernah berhenti, serta dukungan lahir dan batin yang menjadi sumber kekuatan dalam setiap langkah penulis. Segala jerih payah, pengorbanan, dan keikhlasan yang telah diberikan menjadi alasan utama penulis untuk terus berjuang hingga titik ini.
2. Adik tersayang Yang selalu memberi semangat, candaan, dan dukungan di saat penulis merasa lelah dan hampir menyerah. Kehadiranmu menjadi pengingat bahwa penulis tidak pernah berjalan sendiri.
3. Bapak dan Ibu Dosen serta Dosen Pembimbing Yang telah membimbing dengan penuh kesabaran, memberikan ilmu yang bermanfaat, membuka cara pandang baru, serta mengarahkan penulis dalam proses penyusunan skripsi ini. Setiap nasihat dan koreksi menjadi bekal berharga bagi perjalanan penulis ke depan.
4. Teman-teman seperjuangan Yang selalu hadir dalam doa, dukungan moral, diskusi, canda tawa, dan kerja sama sepanjang masa perkuliahan hingga penyusunan skripsi. Kebersamaan kita adalah bagian penting dari proses belajar yang tidak akan terlupakan.
5. Almamater tercinta, Universitas Lancang Kuning Yang telah memberikan ruang dan kesempatan bagi penulis untuk belajar, bertumbuh, dan mengembangkan diri, khususnya di lingkungan Fakultas Ilmu Komputer, Program Studi Teknik Informatika.

Semoga skripsi ini dapat menjadi amal jariyah ilmu yang bermanfaat, walaupun masih jauh dari sempurna. Segala doa, dukungan, dan kebaikan yang telah diberikan kepada penulis, semoga dibalas dengan pahala yang berlipat ganda oleh Allah Subhanahu wa ta'ala. Aamiin.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa ta'ala atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan laporan penelitian skripsi yang berjudul “Analisis Rekomendasi Jurusan Kuliah Bagi Pelajar SMA Dengan Menggunakan Algoritma Random Forest.” Laporan skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Lancang Kuning, Pekanbaru.

Dalam proses penyusunan dan penyelesaian skripsi ini, penulis menghadapi berbagai kendala dan keterbatasan, baik dari segi waktu, kemampuan, maupun pemahaman. Namun berkat pertolongan Allah Subhanahu wa ta'ala serta dukungan, bimbingan, dan bantuan dari berbagai pihak, skripsi ini akhirnya dapat diselesaikan. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Yogi Yunefri, M.Kom., MTA., MCE., selaku Dekan Fakultas Ilmu Komputer Universitas Lancang Kuning.
2. Bapak Afriansyah, M.Kom., MTA., selaku Wakil Dekan I Fakultas Ilmu Komputer Universitas Lancang Kuning.
3. Ibu Dr. Lucky Lhaura Van F.C., M.Kom., MTA., selaku Wakil Dekan II Fakultas Ilmu Komputer Universitas Lancang Kuning.
4. Bapak Dr. Sutejo, M.Kom., MTA., selaku Wakil Dekan III Fakultas Ilmu Komputer Universitas Lancang Kuning.
5. Bapak Dr. Yogi Ersan Fadrial, M.Kom., MTA., selaku Ketua Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning.
6. Bapak Dr. Sutejo, M.Kom., MTA, selaku Dosen Pembimbing yang dengan sabar telah memberikan bimbingan, arahan, ilmu, koreksi, dan motivasi kepada penulis sejak tahap perencanaan hingga penyelesaian skripsi ini.

7. Bapak dan Ibu Dosen Fakultas Ilmu Komputer Universitas Lancang Kuning yang telah memberikan ilmu, pengalaman, dan wawasan yang sangat bermanfaat selama masa perkuliahan.
8. Kedua orang tua tercinta serta seluruh keluarga penulis yang selalu mendoakan, memberikan dukungan moral maupun material, serta menjadi sumber semangat dalam setiap langkah penulis.
9. Sahabat-sahabat dan rekan seperjuangan di Program Studi Teknik Informatika yang telah memberikan bantuan, diskusi, dan dorongan selama proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, baik dari segi isi maupun penyajian. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Besar harapan penulis semoga skripsi ini dapat memberikan manfaat, baik bagi penulis sendiri, bagi lingkungan akademik, maupun bagi pihak-pihak yang berkepentingan dalam pengembangan sistem rekomendasi jurusan kuliah.

Pekanbaru, 12 Februari 2026

Muhammad Aldi Rizky Pratama

ABSTRAK

Kesalahan pemilihan jurusan kuliah masih menjadi masalah serius di Indonesia, dengan persentase mahasiswa yang merasa salah jurusan dilaporkan mencapai sekitar 87%, sehingga berdampak pada motivasi belajar, prestasi, hingga potensi dropout. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi jurusan kuliah bagi pelajar SMA menggunakan algoritma Random Forest serta mengevaluasi tingkat akurasi sistem dalam memberikan rekomendasi jurusan yang tepat. Data penelitian diperoleh dari 394 siswa SMA Negeri 3 Pujud melalui kuesioner daring, yang mencakup 22 fitur berupa nilai akademik, latar belakang keluarga, kegiatan favorit, dan mata pelajaran favorit, dengan 8 jurusan kuliah sebagai kelas target. Data dibagi menggunakan stratified train-test split dengan proporsi 80%:20% dan dilatih menggunakan Random Forest Classifier dengan 200 pohon, kedalaman maksimum 15, dan parameter lain yang disesuaikan secara trial-and-error. Hasil pengujian menunjukkan akurasi keseluruhan sebesar 18,99% dengan performa terbaik pada jurusan Ilmu Komunikasi (F1-Score 30,77%), sementara beberapa jurusan seperti Kedokteran dan Manajemen/Akuntansi tidak berhasil diprediksi sama sekali. Analisis feature importance mengindikasikan bahwa fitur nilai akademik mendominasi proses prediksi (sekitar 49,43%), dengan nilai Bahasa Inggris menjadi fitur paling berpengaruh, sedangkan indikator minat eksplisit hanya berkontribusi kecil. Temuan ini menunjukkan bahwa meskipun Random Forest dapat diimplementasikan sebagai prototipe sistem rekomendasi jurusan, akurasinya masih rendah sehingga lebih tepat digunakan sebagai alat bantu pengambilan keputusan bagi guru BK, dan disarankan untuk penelitian lanjutan dengan penambahan variabel psikologis, penanganan class imbalance, serta perluasan data lintas sekolah.

Kata kunci: rekomendasi jurusan, Random Forest, sistem pendukung keputusan, data siswa SMA

ABSTRACT

Major misselection remains a serious issue in Indonesia, with approximately 87% of university students reported feeling that they chose an inappropriate study program, which negatively affects learning motivation, academic performance, and dropout risk. This study aims to develop a major recommendation system for high school students using the Random Forest algorithm and to evaluate the accuracy of the system in providing suitable major recommendations. The dataset consists of 394 students from SMA Negeri 3 Pujud collected via an online questionnaire, comprising 22 features including academic grades, family background, favorite activities, and favorite subjects, with 8 undergraduate majors as target classes. The data were split using a stratified train–test split with an 80%:20% ratio and trained using a Random Forest Classifier with 200 trees, a maximum depth of 15, and other hyperparameters selected through trial-and-error. The experimental results show an overall accuracy of 18.99%, with the best performance obtained for the Communication Science major (F1-Score 30.77%), while some majors such as Medicine and Management/Accounting were not predicted correctly at all. Feature importance analysis reveals that academic grades dominate the prediction process (around 49.43%), with English scores being the most influential feature, whereas explicit interest indicators contribute only marginally. These findings indicate that although Random Forest can be implemented as a prototype for a major recommendation system, its current accuracy is low, making it more appropriate as a decision-support tool for school counselors, and future work is recommended to incorporate psychological variables, address class imbalance, and extend the dataset to multiple schools.

Keywords: major recommendation, Random Forest, decision support system, high school student data

DAFTAR ISI

HALAMAN PERSETUJUAN.....	II
HALAMAN PENGESAHAN.....	III
PERNYATAAN KEASLIAN KARYA ILMIAH.....	IV
HALAMAN PERSEMBAHAN.....	V
KATA PENGANTAR	VI
ABSTRAK	VIII
ABSTRACT.....	IX
DAFTAR ISI.....	X
DAFTAR TABEL	XIII
DAFTAR GAMBAR	XIV
DAFTAR LAMPIRAN	XVI
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Rumusan Masalah	4
C. Batasan Masalah.....	4
D. Tujuan Penelitian.....	4
E. Manfaat Penelitian.....	5
F. Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
A. Tinjauan Penelitian Terdahulu	7
B. Teori-Teori Dasar	11
1. Analisis.....	11
2. Rekomendasi	11
3. Pelajaran	12
4. Jurusan.....	13
5. Kuliah	13
6. Pelajar SMA	14
7. Algoritma Random Forest.....	15
8. Evaluasi Model <i>Random Forest</i>	17
9. Faktor-Faktor yang Mempengaruhi Pemilihan Jurusan Kuliah	18
10. <i>Data Mining dan Machine Learning</i> dalam Pendidikan	19

11. Python	20
12. <i>Google Colab (Jupyter Notebook berbasis Cloud)</i>	21
13. Library Python untuk Machine Learning	21
BAB III METODOLOGI PENELITIAN.....	24
A. Tahapan-tahapan Penelitian	24
B. Lokasi dan Waktu Penelitian.....	26
C. Data yang Digunakan	28
D. Teknik Pengumpulan Data	32
E. Alur Sistem Rekomendasi	33
F. Metode/Pemodelan yang Digunakan.....	35
BAB IV GAMBARAN UMUM OBJEK PENELITIAN	42
A. Sejarah SMA Negeri 3 Pujud	42
B. Visi Dan Misi Sekolah	43
C. Logo SMA Negeri 3 Pujud.....	44
BAB V HASIL DAN PEMBAHASAN.....	45
A. Hasil Penelitian	45
1. Sumber Data Penelitian.....	45
2. <i>Proses Load Data</i>	45
3. Proses Pengecekan <i>Missing Value</i> dan Seleksi Kelengkapan Data ($\geq 85\%$).....	46
4. Proses Deteksi dan Penanganan <i>Outlier</i> (Metode <i>IQR</i>)	49
5. Proses Normalisasi Data <i>Numerik (Min-Max Scaler)</i>	51
6. Proses Penentuan Target (Jurusan_Utama) dan Distribusi Jurusan ...	53
7. Proses <i>Encoding Data Kategorikal (Label Encoding)</i>	55
8. Proses One-Hot Encoding Variabel Minat (Kegiatan & Mapel Favorit)	59
9. Proses Pembentukan Data Final (X dan y) serta Daftar Fitur	62
10. Proses Pembagian Data Training dan Testing.....	64
11. Proses Training Model Random Forest dan Hasil Akurasi	65
12. <i>Confusion Matrix</i>	68
13. Analisis Feature Importance.....	71
14. Feature Selection (Threshold $\geq 3\%$) dan Uji Ulang Model.....	73
B. Pembahasan	76

1. Pembahasan Hasil Preprocessing Data.....	76
2. Pembahasan Kinerja Model Random Forest pada Berbagai Skenario Split	77
3. Analisis Kesalahan Prediksi Berdasarkan Confusion Matrix.....	78
4. Interpretasi Feature Importance (Skenario Terbaik 80:20)	78
5. Pembahasan Hasil Feature Selection (Threshold $\geq 3\%$)	79
6. Alasan Umum Akurasi Masih Rendah dan Implikasi untuk Sistem Rekomendasi	80
7. Saran Perbaikan dan Pengembangan.....	81
C. Keterbatasan Penelitian	81
1. Keterbatasan Variabel dan Representasi Minat.....	81
2. Penentuan Target “Jurusan_Utama” Bersifat Penyederhanaan.....	81
3. Distribusi Data per Kelas Tidak Sepenuhnya Merata	82
4. Evaluasi Masih Terbatas pada Pembagian Data <i>Train-Test Split</i>	82
5. Optimasi Hyperparameter Belum Dilakukan Secara Ekstensif	82
BAB IV PENUTUP	83
A. Kesimpulan.....	83
B. Saran.....	84
DAFTAR PUSTAKA	85
LAMPIRAN.....	88

DAFTAR TABEL

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	24
Gambar 3. 2 Flowchart Sistem Rekomendasi Jurusan Hybrid.....	33
Gambar 4. 1 Logo Sekolah.....	44
Gambar 5. 1 Script Load Dataset form.csv.....	45
Gambar 5. 2 Script Missing Value & Seleksi Kelengkapan Data.....	47
Gambar 5. 3 Ringkasan Missing Value pada Kolom Penting.....	48
Gambar 5. 4 Ringkasan Seleksi Kelengkapan Data ($\geq 85\%$).....	48
Gambar 5. 5 Visualisasi Hasil Pembersihan Data.....	49
Gambar 5. 6 Script Deteksi Outlier Metode IQR.....	50
Gambar 5. 7 Ringkasan Outlier (Metode IQR).....	51
Gambar 5. 8 Script Normalisasi Data (Min-Max Scaler).....	51
Gambar 5. 9 Data Sebelum Normalisasi.....	52
Gambar 5. 10 Data Sesudah Normalisasi.....	52
Gambar 5. 11 Script Ekstraksi Jurusan_Utama.....	54
Gambar 5. 12 Distribusi Jurusan Utama (N = 394).....	54
Gambar 5. 13 Grafik Distribusi Jurusan Utama.....	55
Gambar 5. 14 Script Label Encoding Fitur Kategorikal.....	56
Gambar 5. 15 Mapping Label Encoding – Pekerjaan Ayah.....	57
Gambar 5. 16 Mapping Label Encoding – Pekerjaan Ibu.....	58
Gambar 5. 17 Mapping Label Encoding – Penghasilan Orang Tua.....	58
Gambar 5. 18 Mapping Label Encoding – Kelas.....	59
Gambar 5. 19 Script One-Hot Encoding Minat.....	60
Gambar 5. 20 One-Hot Encoding – Kegiatan Favorit.....	61
5. 21 One-Hot Encoding – Mata Pelajaran Favorit.....	61
Gambar 5. 22 Script Pembentukan Feature Columns, X dan y.....	62
Gambar 5. 23 Daftar Fitur Model (Total 22 Fitur).....	63
Gambar 5. 24 Script Train-Test Split Multi-Skenario.....	64
Gambar 5. 25 Hasil Pembagian Data Training dan Testing.....	65
Gambar 5. 26 Script Training Random Forest dan Evaluasi Akurasi.....	66

Gambar 5. 27 Ringkasan Akurasi Model Random Forest (Berbagai Skenario Split)	67
Gambar 5. 28 Grafik Perbandingan Akurasi Model Random Forest	68
Gambar 5. 29 Script Confusion Matrix Semua Skenario	69
Gambar 5. 30 Confusion Matrix Semua Skenario	70
Gambar 5. 31 Script Feature Importance Random Forest	71
Gambar 5. 32 Top 15 Feature Importance Random Forest (80:20)	72
Gambar 5. 33 Grafik Feature Importance	73
Gambar 5. 34 Script Feature Selection Berdasarkan Threshold	74
Gambar 5. 35 Fitur Terpilih (Importance $\geq 3\%$)	74
Gambar 5. 36 Visualisasi Fitur Terpilih	75

DAFTAR LAMPIRAN

Lampiran 1 Surat Fakultas dan Sekolah	88
Lampiran 2 Survei Langsung	90
Lampiran 3 Wawancara.....	97
Lampiran 4 Tempat Pengambilan Wawancara.....	106
Lampiran 5 Kode Program	107

BAB I

PENDAHULUAN

A. Latar Belakang

Era revolusi industri 4.0 dan transformasi digital telah mengubah lanskap pendidikan tinggi dan dunia kerja secara fundamental. Pendidikan tinggi memiliki peran strategis dalam mengembangkan sumber daya manusia yang berkualitas dan kompetitif untuk menghadapi tantangan global. Perguruan tinggi sebagai jenjang lanjutan dalam sistem pendidikan formal diharapkan mampu menghasilkan lulusan yang tidak hanya menguasai pengetahuan teoretis, tetapi juga keterampilan praktis yang relevan dengan kebutuhan pasar kerja yang dinamis. Dalam konteks ini, pemilihan jurusan yang tepat menjadi pondasi krusial bagi keberhasilan akademik dan karir mahasiswa di masa depan.

Namun, realitas di lapangan menunjukkan kondisi yang sangat memprihatinkan. Berdasarkan data penelitian terbaru (Alfaiz Maulana, 2024), sekitar 87% mahasiswa di Indonesia merasa telah salah memilih program studi, dengan kurangnya pemahaman mengenai karakteristik dan prospek karir sebagai faktor utama. Angka ini menunjukkan bahwa hampir 9 dari 10 mahasiswa mengalami ketidaksesuaian antara pilihan jurusan dengan minat, bakat, dan kemampuan mereka. Fenomena ini semakin mengkhawatirkan mengingat dari 3,7 juta lulusan SMA/SMK/MA setiap tahunnya, sekitar 1,8-1,9 juta siswa yang melanjutkan ke perguruan tinggi harus membuat keputusan krusial mengenai pilihan jurusan kuliah mereka (detik.com, 2021).

Kesalahan pemilihan jurusan ini menimbulkan dampak yang sangat signifikan, baik dari aspek akademik maupun ekonomi. Dampak akademik meliputi penurunan motivasi belajar mahasiswa yang cenderung kehilangan semangat dalam mengikuti perkuliahan karena merasa program studi tidak sesuai dengan *passion* mereka, rendahnya prestasi akademik akibat ketidaksesuaian antara kemampuan dan minat dengan jurusan yang dipilih, serta tingginya angka *dropout* yang mencapai 15-20% di berbagai perguruan tinggi Indonesia. Dari aspek ekonomi, kesalahan pemilihan jurusan menyebabkan kerugian finansial yang mencapai triliunan rupiah per tahun akibat biaya

pendidikan yang terbuang, waktu yang hilang, dan beban psikologis serta finansial keluarga.

Akar permasalahan ini terletak pada keterbatasan sistem bimbingan konseling yang ada saat ini. Sebagian besar SMA di Indonesia masih menggunakan metode manual dan konvensional dalam memberikan konseling pemilihan jurusan, dengan rasio konselor terhadap siswa yang sangat tidak ideal (1:150-200 siswa). Sistem rekomendasi yang tersedia masih mengandalkan tes minat bakat sederhana yang tidak terintegrasi dengan data akademik komprehensif siswa. Penelitian (Ulfa *et al.*, 2020) menunjukkan bahwa algoritma C4.5 yang digunakan dalam sistem rekomendasi saat ini menghasilkan akurasi yang relatif rendah, sementara penelitian lain (Fitriawanti *et al.*, 2018) dengan algoritma Naive Bayes hanya mencapai akurasi 82%.

Keterbatasan teknologi dalam sistem rekomendasi jurusan menjadi masalah krusial yang harus segera diatasi. Sistem yang ada belum mampu mengintegrasikan data multidimensional secara optimal, seperti nilai akademik siswa, minat dan bakat, kondisi sosio-ekonomi keluarga, tren pasar kerja, dan kebutuhan industri masa depan. Kompleksitas data ini memerlukan pendekatan teknologi yang lebih canggih untuk menghasilkan rekomendasi yang akurat dan personal. Selain itu, kurangnya validasi empiris terhadap efektivitas sistem rekomendasi yang digunakan menyebabkan tingkat kepercayaan yang rendah dari siswa dan orang tua terhadap rekomendasi yang diberikan.

Di era digital saat ini, negara-negara maju telah mengimplementasikan sistem rekomendasi berbasis *artificial intelligence* dan *machine learning* untuk membantu siswa dalam pemilihan jalur karir dan pendidikan. Sementara itu, Indonesia masih tertinggal dalam pemanfaatan teknologi ini untuk sektor pendidikan. *Gap* teknologi ini menjadi semakin krusial mengingat percepatan transformasi digital dan tuntutan personalisasi dalam layanan pendidikan yang semakin tinggi.

Mengingat besarnya dampak pemilihan jurusan yang tepat terhadap masa depan pendidikan dan karir siswa, pengembangan sistem rekomendasi jurusan yang lebih akurat menggunakan teknologi *machine learning* menjadi

sangat *urgent*. *Random Forest*, sebagai algoritma pembelajaran ansambel yang menggabungkan beberapa *decision tree*, memiliki keunggulan signifikan dalam menangani data kompleks dan menghasilkan prediksi yang lebih akurat dibandingkan algoritma tunggal. Penelitian (Priantama *et al.*, 2022) menunjukkan bahwa *Random Forest Classifier* dapat mencapai akurasi hingga 97,22% dalam prediksi performa akademik mahasiswa ketika dioptimasi dengan pemilihan fitur berbasis korelasi.

Keunggulan *Random Forest* terletak pada kemampuannya dalam mengatasi *overfitting*, menangani nilai yang hilang (*missing values*), dan memberikan *ranking importance* terhadap fitur-fitur yang digunakan. Hal ini menjadikannya pilihan yang tepat untuk mengolah data siswa yang seringkali tidak lengkap atau bervariasi. Penelitian (Syahputra *et al.*, 2022) juga menunjukkan efektivitas algoritma *machine learning* dalam mengklasifikasi calon mahasiswa untuk sistem rekomendasi program studi. Studi perbandingan (Novianto *et al.*, 2023) mengenai metode klasifikasi *Random Forest* dan *Support Vector Machine* dalam memprediksi capaian studi mahasiswa memberikan landasan empiris yang kuat untuk pemilihan algoritma ini.

Urgensi penelitian ini juga didorong oleh kebutuhan akan solusi teknologi yang dapat mengurangi tingkat kesalahan pemilihan jurusan dari 87% menjadi lebih rendah, perlunya sistem yang dapat mengintegrasikan *multiple criteria* dalam pengambilan keputusan, serta kebutuhan akan sistem yang dapat diadaptasi untuk kondisi lokal pendidikan Indonesia. Implementasi sistem rekomendasi berbasis *Random Forest* diharapkan dapat memberikan dampak transformatif dalam meningkatkan akurasi pemilihan jurusan, mengurangi angka *dropout*, dan pada akhirnya menghasilkan lulusan perguruan tinggi yang lebih siap menghadapi tantangan dunia kerja di era *industry 4.0*.

Berdasarkan identifikasi masalah dan urgensi di atas, penelitian ini berfokus pada analisis rekomendasi jurusan kuliah bagi pelajar SMA menggunakan algoritma *Random Forest*. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam mengurangi tingkat kesalahan pemilihan jurusan,

meningkatkan kepuasan akademik mahasiswa, mengoptimalkan investasi pendidikan keluarga, dan pada akhirnya menghasilkan lulusan perguruan tinggi yang lebih kompeten dan siap menghadapi tantangan dunia kerja masa depan.

B. Rumusan Masalah

Berdasarkan rumusan masalah yang telah diuraikan tersebut, maka perumusan masalah dalam penelitian ini adalah :

1. Bagaimana Menerapkan Algoritma *Random Forest* Untuk Rekomendasi Jurusan Kuliah Bagi Pelajar SMA?
2. Bagaimana Akurasi Sistem Dalam Memberikan Rekomendasi Jurusan Kuliah Yang Tepat?

C. Batasan Masalah

Dalam penelitian ini, Batasan masalah dalam penelitian ini ialah sebagai berikut:

1. Objek penelitian difokuskan pada data akademik pelajar SMA kelas XII
2. Menerapkan algoritma *Random Forest* untuk menyelesaikan proses klasifikasi rekomendasi jurusan
3. Jurusan kuliah yang direkomendasikan terbatas pada 8 program studi populer di Indonesia
4. Data yang digunakan mencakup nilai rapor, hasil tes minat bakat, dan karakteristik demografis siswa
5. Implementasi sistem menggunakan bahasa pemrograman *Python* dengan library *Scikit-learn*

D. Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan penelitian adalah:

1. Untuk menerapkan algoritma *Random Forest* dalam sistem rekomendasi jurusan kuliah bagi pelajar SMA
2. Untuk mengetahui tingkat akurasi sistem dalam memberikan rekomendasi jurusan kuliah yang tepat

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang signifikan bagi berbagai pihak yang terlibat dalam ekosistem pendidikan

1. Bagi Peneliti

Penelitian ini memberikan kontribusi akademik dalam pengembangan kompetensi riset algoritma *Random Forest* untuk menyelesaikan permasalahan nyata di bidang pendidikan.

2. Bagi Sekolah

Sekolah dapat mengimplementasikan sistem rekomendasi berbasis *Random Forest* untuk meningkatkan akurasi dan efisiensi layanan bimbingan konseling, mengoptimalkan alokasi sumber daya sesuai potensi siswa, meningkatkan reputasi institusi melalui inovasi teknologi.

3. Bagi Pembaca

Penelitian ini memberikan pemahaman komprehensif bagi siswa dan orang tua tentang pentingnya pemilihan jurusan yang tepat, panduan praktis bagi praktisi pendidikan untuk implementasi teknologi *machine learning*.

F. Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada Bab ini berisi tentang teori-teori yang mendukung penelitian, meliputi konsep sistem rekomendasi, algoritma *Random Forest*, *machine learning*, pemilihan jurusan kuliah, dan penelitian terkait sebelumnya.

BAB III METODE PENELITIAN

Pada bab ini berisi tentang metode penelitian yang digunakan, meliputi jenis penelitian, teknik pengumpulan data, populasi dan

sampel, variabel penelitian, dan tahapan implementasi algoritma *Random Forest*.

BAB IV JADWAL PELAKSANAAN

Pada bab ini berisi tentang rencana jadwal kegiatan penelitian dari tahap awal hingga penyelesaian penelitian.

DAFTAR PUSTAKA

Pada bagian ini berisi sumber-sumber referensi yang digunakan dalam penelitian ini.

LAMPIRAN

Pada bagian ini berisi tentang data pendukung penelitian seperti surat penelitian, survei langsung, kuesioner, dan wawancara.

BAB VI

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan mengenai sistem rekomendasi jurusan kuliah menggunakan algoritma *Random Forest* pada siswa SMA Negeri 3 Pujud, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Penerapan algoritma *Random Forest* untuk rekomendasi jurusan kuliah

Penelitian ini berhasil menerapkan algoritma *Random Forest* sebagai model klasifikasi multi-kelas untuk merekomendasikan delapan jurusan kuliah berdasarkan 22 fitur yang mencakup nilai akademik, latar belakang keluarga, kegiatan favorit, dan mata pelajaran favorit pada 394 siswa SMA Negeri 3 Pujud. Proses pemodelan meliputi tahapan *preprocessing* (*label encoding* dan *one-hot encoding*), pembagian data dengan empat skenario *train-test split* (70:30, 80:20, 90:10, dan 60:40) secara *stratified*, pelatihan model *Random Forest Classifier* dengan 200 pohon keputusan, serta evaluasi menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Model ini kemudian diintegrasikan ke dalam alur sistem rekomendasi hibrida yang menggabungkan *machine learning*, *rule-based system*, dan peran guru BK sehingga dapat digunakan sebagai alat bantu pendukung keputusan dalam pemilihan jurusan, bukan pengganti proses konseling tatap muka

2. Tingkat akurasi sistem dalam memberikan rekomendasi jurusan kuliah

Hasil evaluasi menunjukkan bahwa akurasi tertinggi diperoleh pada skenario pembagian data 80:20 dengan nilai akurasi 18,99%, sedangkan skenario 70:30, 90:10, dan 60:40 masing-masing menghasilkan akurasi 14,29%, 15,00%, dan 12,03%. Pada skenario terbaik (80:20), jurusan Ilmu Komunikasi memiliki kinerja paling tinggi dengan *F1-score* sebesar 30,77%, sementara jurusan Kedokteran

dan Manajemen/Akuntansi tidak berhasil diprediksi sama sekali (F1-score 0%) karena jumlah sampel yang terbatas dan karakteristik yang tumpang tindih dengan jurusan lain.

Analisis *feature importance* menunjukkan bahwa nilai akademik mendominasi proses prediksi dengan kontribusi total 49,43%, di mana Bahasa Inggris (10,71%), Nilai Rata-rata Rapor (10,11%), IPA (9,71%), Matematika (9,62%), dan Bahasa Indonesia (9,28%) menjadi lima fitur terpenting. Latar belakang keluarga berkontribusi 21,73% yang mencakup kelas siswa (6,06%), pekerjaan ayah (5,65%), pekerjaan ibu (5,32%), dan penghasilan orang tua (4,70%), sedangkan minat eksplisit siswa melalui kegiatan favorit dan mata pelajaran favorit hanya menyumbang sekitar 15%.

Secara keseluruhan, tingkat akurasi yang rendah ini mengindikasikan bahwa model *Random Forest* dalam konfigurasi dan ruang fitur saat ini belum layak digunakan sebagai satu-satunya penentu rekomendasi jurusan, tetapi lebih tepat dimanfaatkan sebagai salah satu masukan kuantitatif dalam proses bimbingan konseling dengan tetap mempertimbangkan aspek psikologis, sosial, dan kontekstual yang tidak dapat diukur oleh algoritma.

B. Saran

1. Memanfaatkan sistem rekomendasi ini sebagai salah satu bahan pertimbangan dalam layanan bimbingan konseling, bukan sebagai satu-satunya penentu jurusan
2. Mengintegrasikan data akademik, minat, bakat, serta informasi mengenai kondisi keluarga dalam proses pemberian rekomendasi jurusan agar keputusan yang diambil siswa lebih tepat sasaran
3. Menyelenggarakan program pendukung seperti seminar karier, campus visit, pengenalan jurusan kuliah, dan tes minat bakat terstandar untuk membantu siswa mendapatkan gambaran yang lebih jelas tentang pilihan jurusan

DAFTAR PUSTAKA

- Alfaiz Maulana. (2024). *Sistem Rekomendasi Pemilihan Jurusan Menggunakan Algoritma Naive Bayes*.
<https://repository.arraniry.ac.id/id/eprint/39071/1/Maulana%20Alfaiz-190212071-Skripsi..pdf>
- Erwin Dariyanto (2021). 3,7 Juta pelajar SMA, Hanya 1,8 Juta yang bisa kuliah. Detik.com <https://www.detik.com/edu/sekolah/d-5623865/setiap-tahun-3-7-juta-pelajar-lulus-sma-hanya-1-8-juta-yang-bisa-kuliah>
- Anderson, J., & Lee, M. (2023). Sistem rekomendasi dalam pendidikan tinggi: Systematic review dan meta-analisis. *Educational Technology Research Journal*.
- Anugerah, R., Pratama, D., & Sari, L. (2025). Sistem rekomendasi jurusan kuliah bagi calon mahasiswa baru Universitas BSI Margonda Fakultas Teknik dan Informatika menggunakan algoritma C4.5. *Jurnal Teknologi Informasi dan Pendidikan*. <https://ejournal.itn.ac.id/index.php/jati/article/view/12849/7327>
- Fatmawati, R., & Purnomo, H. D. (2023). Optimasi parameter Naive Bayes untuk sistem rekomendasi jurusan kuliah berdasarkan minat dan kemampuan siswa. *Jurnal Ilmiah Teknologi Informasi*, 11(2), 210–225.
- Firmansyah, D., & Nurhasanah, N. (2021). Pengukuran performa sistem rekomendasi: Metrik dan implementasi. *Jurnal Teknologi Informasi dan Sistem Komputer*, 8(3), 145–158.
- Fitriawanti, R., Cholissodin, I., & Dewi, R. K. (2018). *Klasifikasi dan Rekomendasi Jurusan Kuliah Bagi Pelajar SMA Menggunakan Algoritme Naïve Bayes-WP* (Vol. 2, Issue 11). <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3129/1223>
- Hartono, B., & Dewi, R. (2023). Pengaruh bimbingan konseling terhadap ketepatan pemilihan jurusan kuliah siswa SMA. *Jurnal Bimbingan dan Konseling Indonesia*, 8(1), 12-20.

- Novianto, A., Wijaya, K., & Sari, D. (2023). Perbandingan metode klasifikasi Random Forest dan Support Vector Machine untuk prediksi capaian studi mahasiswa. *Indonesian Journal of Machine Learning*, 5(2), 78-89.
- Novianto, F., Setiawan, A., & Purnama, B. (2023). Perbandingan metode klasifikasi random forest dan support vector machine dalam memprediksi capaian studi mahasiswa. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(3), 287-295.
<https://jurnal.stkipppgritulungagung.ac.id/index.php/jipi/article/view/5423>
- Priantama, A., Gunawan, R., & Sari, D. P. (2022). Optimasi correlation-based feature selection untuk perbaikan akurasi random forest classifier dalam prediksi performa akademik mahasiswa. *JIKO (Jurnal Informatika dan Komputer)*, 5(2), 78-85.
<https://ejournal.akakom.ac.id/index.php/jiko/article/view/651/0>
- Priantama, F., Kusuma, H., & Lestari, P. (2022). Prediksi performa akademik mahasiswa menggunakan Random Forest dengan feature selection. *Jurnal Teknologi Pendidikan Indonesia*, 4(3), 134-145.
- Rahayu, S., & Permana, D. (2022). Faktor-faktor yang mempengaruhi pemilihan jurusan kuliah pada siswa SMA di era digital. *Jurnal Bimbingan dan Konseling*, 9(2), 134–147.
- Rahman, S., & Sari, M. (2021). Sistem pendukung keputusan pemilihan program studi menggunakan data mining. *Journal of Information Systems and Computer Science*, 6(1), 23-34.
- Rohmah, N., & Saputro, D. (2022). Pengembangan sistem rekomendasi jurusan berbasis web menggunakan algoritma Naive Bayes. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(3), 245–258.
- Sari, K., Wulandari, T., & Pratama, J. (2022). Faktor psikologis dan sosio-ekonomi dalam pemilihan jurusan kuliah: Analisis structural equation modeling. *Jurnal Psikologi Pendidikan Indonesia*, 7(2), 89-102.
- Syahputra, H., Rahman, A., & Ningsih, S. (2022). Model klasifikasi calon mahasiswa baru untuk sistem rekomendasi program studi sarjana berbasis

- machine learning. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(4), 725-734. <https://jtiik.ub.ac.id/index.php/jtiik/article/view/4311>
- Syahputra, R., Andika, M., & Fitriani, L. (2022). Klasifikasi calon mahasiswa untuk sistem rekomendasi program studi menggunakan machine learning. *Indonesian Journal of Artificial Intelligence*, 4(1), 56-67.
- Ulfa, A., Winarso MKom, D., & Arribe MMSi, E. (2020). SISTEM REKOMENDASI JURUSAN KULIAH BAGI CALON MAHASISWA BARU MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus : Universitas Muhammadiyah Riau). *Jurnal Fasilkom*, 10, 21–26.
- Wibowo, A., & Putri, N. (2023). Dampak ketidaksesuaian jurusan terhadap prestasi akademik: Studi longitudinal. *Jurnal Penelitian Pendidikan Tinggi*, 9(3), 201-215.
- Wijaya, A., & Sari, R. P. (2023). *Sistem rekomendasi: Teori dan perkembangan terkini*. Penerbit Informatika.
- Yuliani, I., & Suryani, E. (2020). Perancangan sistem rekomendasi jurusan menggunakan algoritma Naive Bayes pada siswa SMA. *Jurnal Teknologi dan Sistem Informasi*, 8(2), 145–152.