

# Prediksi Status Pekerjaan Lulusan SMK Menggunakan Algoritma Random Forest: Analisis Multifaktor Akademis, Sosial, dan Keluarga

Muchamad Firmansyah Tubira<sup>1\*</sup>

Yulian Findawati<sup>2</sup>

Rohman Dijaya<sup>3</sup>

Yunianita Rahmwati<sup>4</sup>

<sup>1,2,3,4</sup> Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo, Jl. Mojopahit No.666 B, Sidowayah, Celep, Kec. Sidoarjo, Kabupaten Sidoarjo, Jawa Timur 61215, Indonesia

<sup>1</sup>muchfirmansyah96@gmail.com, <sup>2</sup> yulianfindawati@umsida.ac.id, <sup>3</sup>[rohman.dijaya@umsida.ac.id](mailto:rohman.dijaya@umsida.ac.id),

<sup>4</sup>yunianita@umsida.ac.id.

## \*Penulis Korespondensi:

Muchamad Firmansyah Tubira

muchfirmansyah96@gmail.com

## Abstrak

*Penelitian ini bertujuan untuk mengembangkan model prediksi status pekerjaan lulusan siswa Sekolah Menengah Atas (SMK) di Indonesia dengan menggunakan analisis multifaktor yang melibatkan kinerja akademik, lingkungan sosial, dan keluarga. Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma Random Forest untuk mengumpulkan data dalam jumlah besar dan memberikan prediksi yang spesifik. Model tersebut memprediksi status pekerja lulusan siswa SMK sebesar 76%, yang menunjukkan kinerja kerja yang baik. Penelitian ini juga menemukan bahwa faktor keluarga yang signifikan memengaruhi status pekerjaan lulusan siswa SMK secara signifikan (38.9%), diikuti oleh faktor sosial (36.8%) dan akademik (24.3%). Data tersebut diperkuat dari fitur dukungan orangtua terhadap Pendidikan (Parental Support for Education). Tujuan penelitian ini untuk mendorong sekolah, orang tua, dan pemerintah untuk fokus pada pendidikan holistik SMK, seperti kolaborasi antara sekolah dan industri, untuk meningkatkan status pekerjaan lulusan siswa SMK.*

**Kata kunci:** *Faktor Keluarga; Machine Learning; Prediksi; Random Forest; SMK; Status Pekerjaan.*

## Abstract

*This study aims to develop a prediction model for the employment status of high school graduates in Indonesia using multifactor analysis involving academic performance, social environment, and family. This study uses a quantitative approach with the Random Forest algorithm to collect large amounts of data and provide specific predictions. The model predicts the employment status of vocational school graduates by 76%, indicating good work performance. This study also found that significant family factors significantly affect the employment status of vocational school graduates (38.9%), followed by social factors (36.8%) and academic factors (24.3%). The data is strengthened by the Parental Support for Education feature. The purpose of this study is to encourage schools, parents, and the government to focus on holistic vocational school education, such as collaboration between schools and industry, to improve the employment status of vocational school graduates.*

**Keywords:** *Family Factors; Job Status; Machine Learning; Prediction; Random Forest; Vocational High School (SMK).*

## 1. Pendahuluan

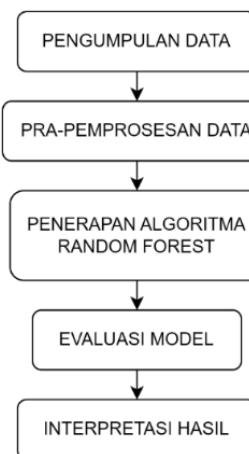
Pendidikan Tinggi di Indonesia, khususnya Sekolah Menengah Kejuruan(SMK), bertujuan untuk mencapai standar tinggi dan kepuasan kerja yang tinggi di pasar tenaga kerja[1]. Hal ini tidak hanya dicapai melalui pengetahuan teoritis tetapi juga melalui penerapan praktis dari keterampilan tersebut. Siswa dan siswa sering kali menjadi indikator utama keberhasilan mereka di pasar tenaga kerja[2]. Prestasi akademik dianggap sebagai indikator pertama keberhasilan

yang tinggi, tetapi faktor sosial seperti keterlibatan dalam kegiatan ekstrakurikuler dan lingkungan keluarga juga memainkan peran penting dalam mencapai standar yang tinggi[3]. Kondisi moral dan keuangan juga memainkan peran penting dalam menentukan status siswa di pasar tenaga kerja[4]. Penambangan data, yang menggunakan data untuk mengeksplorasi dan meningkatkan pengetahuan, digunakan untuk membuat analisis yang komprehensif[5]. Penelitian ini dilakukan di sebuah sekolah kejuruan di Sidoarjo untuk mengeksplorasi hubungan antara aspek akademik, sosial, dan keluarga dalam menentukan status sekolah menengah atas. Penelitian ini menggunakan pendekatan multidimensi, yang mengintegrasikan berbagai faktor untuk membuat analisis yang komprehensif[6].

Algoritma Random Forest digunakan untuk mengatasi masalah ini dengan memiliki peran pengambilan keputusan dalam proses pembelajaran dan menggunakan berbagai set data untuk memprediksi hasil selama pelatihan[7]. Pendekatan ini membantu sekolah dan pemerintah membuat keputusan yang tepat tentang masa depan pendidikan dan memberikan informasi berharga tentang faktor-faktor yang berkontribusi terhadap status sekolah menengah atas[8]. Metode agregasi bootstrap membantu memisahkan kumpulan data yang berbeda, sehingga menghasilkan prediksi yang lebih akurat[5]. Penelitian ini berfokus pada lulusan dalam lima tahun terakhir, bertujuan untuk memprediksi apakah mereka akan bekerja, berwirausaha, atau melanjutkan pendidikan. Sebagaimana diungkapkan oleh (Baffa *et al.*, 2023), hasil prediksi yang akurat dapat menjadi acuan bagi sekolah dan pemerintah dalam menyusun kebijakan yang mendukung peningkatan penyerapan lulusan di dunia kerja [9]. Maka dari itu penelitian ini bertujuan untuk menciptakan tenaga kerja yang lebih siap dan sesuai dengan kebutuhan industri, sekaligus meningkatkan daya saing lulusan SMK di dunia kerja.

## 2. Metode Penelitian

Algoritma *Random Forest* merupakan teknik *machine learning* yang berbasis pada metode ensemble, di mana banyak pohon keputusan digunakan untuk meningkatkan akurasi prediksi[10]. Teknik ini mengatasi masalah overfitting dalam data kompleks dengan membuat pohon keputusan besar dan menghitung hasil melalui pemungutan suara untuk klasifikasi atau regresi[11]. Dalam konteks prediksi status pekerjaan, *Random Forest* efektif dalam menangani data dengan banyak variabel, seperti aspek demografi, sosial, dan kekeluargaan, karena dapat mengidentifikasi faktor-faktor penting di antara data yang heterogen[8]. Teknik ini juga dapat digunakan untuk memprediksi status pekerjaan secara lebih akurat dengan menyesuaikan karakteristik individu. *Random Forest* menyediakan prediksi umum dan sangat penting dalam mengidentifikasi variabel yang secara signifikan memengaruhi kinerja pekerjaan[7]. Metode penelitian meliputi pengumpulan data, pra-pemprosesan data, evaluasi data, dan interpretasi hasil[12]. Tahapan-tahapan utama dalam alur penelitian adalah sebagai berikut:



**Gambar 1.** Alur Penelitian

## 2.1 Pengumpulan Data

Langkah pertama dalam penelitian ini adalah mengumpulkan data yang relevan dari berbagai sumber seperti survei tentang kinerja sekolah dan administrasi sekolah[13]. Data ini mencakup variabel dari aspek akademis, sosial, dan keluarga, serta atribut target, seperti status kinerja sekolah[14]. Data tersebut kemudian disusun dalam format terstruktur untuk analisis yang lebih komprehensif[15]. Status kinerja sekolah bergantung pada kondisi yang dihadapi siswa setelah menyelesaikan pendidikan, seperti jadwal kerja, kemajuan, atau perpindahan ke sekolah yang lebih tinggi[16]. Status ini sering kali menjadi indikator efektivitas sistem pendidikan, terutama di departemen sekolah tingkat tinggi seperti SMK, yang bertujuan untuk menyediakan lingkungan kerja yang lebih baik bagi siswa. Dalam konteks SMK, prediksi kinerja sekolah didasarkan pada analisis multifaktor, dengan mempertimbangkan faktor akademis, sosial, dan keluarga untuk menentukan kinerja mereka di tempat kerja[17].

## 2.2 Pra-pemrosesan Data

Setelah data terkumpul, dilakukan prapemrosesan untuk memastikan kualitas data. pada tahap prapemrosesan ini meliputi pembersihan data untuk menghilangkan duplikat, memperbaiki data yang tidak konsisten, dan menghilangkan nilai yang hilang untuk mencegah prediksi model[4]. Data non-numerik, seperti; status hukum, status karyawan, dan lain-lain. Data diubah ke format numerik menggunakan teknologi pengodean, yang memungkinkannya diproses menggunakan algoritma *Random Forest*[8]. Pemilihan fitur dilakukan menggunakan metode *Random Forest* untuk mengidentifikasi variabel paling signifikan yang memengaruhi status pekerjaan, yang bertujuan untuk meningkatkan kinerja model dan mengurangi risiko overfitting[7].

## 2.3 Penerapan Algoritma Random Forest

Langkah selanjutnya adalah membuat model klasifikasi menggunakan algoritma Random Forest. Proses ini melibatkan dua langkah: pemilihan set data (80% data pelatihan) dan pemilihan model (20% data pengujian[11]). Data pelatihan digunakan untuk melatih model, yang membuat pohon keputusan berdasarkan berbagai subset data dan fitur. Pemilihan model digunakan untuk memprediksi status tenaga kerja (bekerja,kuliah, atau berwirausaha). Pada data pengujian, yang ditentukan melalui pemungutan suara mayoritas dari semua titik keputusan[10].

## 2.4 Evaluasi Model

Untuk mengevaluasi kinerja suatu model, digunakan beberapa metrik evaluasi seperti; Akurasi, Presisi dan Recall, F1-Score, dan AUC (Area Under Curve). Akurasi mengukur kemungkinan suatu model memprediksi hasil tertentu, sedangkan recall mengukur jumlah kasus yang berhasil diprediksi oleh model tersebut[4]. F1-Score merupakan rasio harmonis antara presisi dan recall, yang digunakan untuk memberikan gambaran umum tentang kinerja model jika dibandingkan dengan model lain. Hasil evaluasi ini akan sangat penting untuk menentukan kesesuaian model, seperti parameter penyetelan atau pengumpulan data. Hasil evaluasi digunakan untuk mengidentifikasi kelemahan model dan melakukan perbaikan jika diperlukan[8].

## 2.5 Interpretasi Hasil

Penelitian ini menggunakan data historis untuk memprediksi kondisi kerja di masa mendatang, membantu memahami berbagai variabel yang terkait dengan kondisi kerja[18]. Faktor-faktor seperti jumlah populasi, lingkungan sosial, dan kinerja akademis dapat memberikan pemahaman yang lebih baik tentang kondisi kerja di masa mendatang, yang memungkinkan intervensi yang lebih efektif[19]. Metode prediktif, seperti Random Forest, digunakan di berbagai bidang, termasuk pendidikan, untuk membuat keputusan yang tepat. Tindakan prediktif dapat mengoptimalkan pengembangan kurikulum dan meningkatkan kinerja siswa untuk mencapai kinerja kerja yang tinggi[20]. Penelitian ini dilakukan di sebuah sekolah menengah atas di Sidoarjo, sebuah lembaga pendidikan setempat, untuk menganalisis berbagai faktor yang memengaruhi status pekerjaan saat ini, termasuk faktor akademis, sosial, dan kekeluargaan. Penelitian ini juga menggunakan data dari arsip sekolah dan lembaga terkait, seperti administrasi

sekolah dan kondisi sosial ekonomi siswa. Menggabungkan data ini dengan kumpulan data pertama dan kedua menghasilkan pemahaman yang komprehensif tentang masa depan pekerjaan, baik di bidang pendidikan maupun lingkungan[21]n. Untuk memperjelas atribut yang digunakan dalam penelitian ini, berikut adalah tabel yang merangkum kategori, atribut, dan deskripsi variabel:

**Tabel 1.** Kategori, atribut dan deskripsi Dataset

Kategori	Atribut	Deskripsi
<b>Akademis</b>	Average_Report_Score	Nilai rata-rata rapor siswa.
	Vocational_Major	Jurusan keahlian siswa di SMK.
	Attendance_Percentage	Percentase status pekerjaan lulusan siswa.
	Extracurricular_Activities	Keikutsertaan dalam ekstrakurikuler.
	Academic_Achievements	Prestasi akademis yang diraih siswa.
	Graduation_Time	Waktu kelulusan siswa.
	Additional_Training	Pelatihan tambahan setelah lulus.
	Certification	Kepemilikan sertifikasi tambahan.
<b>Sosial</b>	Organization_Participation	Keaktifan dalam organisasi sekolah.
	Social_Media_Use_Hours	Lama penggunaan media sosial per hari.
	Social_Circle	Jumlah teman dekat siswa.
	Self_Confidence_Level	Tingkat kepercayaan diri siswa.
	Community_Involvement	Keterlibatan dalam kegiatan keluarga.
<b>Keluarga</b>	Parent_Job_Status	Pekerjaan utama orang tua.
	Parent_Education	Tingkat pendidikan orang tua.
	Family_Income_Rp	Pendapatan bulanan keluarga.
	Family_Size	Jumlah anggota keluarga.
	Family_Economic_Condition	Penilaian umum mengenai kondisi ekonomi keluarga, seperti mampu atau tidaknya memenuhi kebutuhan dasar.
	Parental_Support_for_Education	Bentuk dukungan orang tua terhadap pendidikan anak, baik secara moral (motivasi) maupun finansial (biaya pendidikan).
	Residence	Tempat tinggal siswa.
	Job_Status_After_Graduation	Status pekerjaan setelah lulus.
<b>Target</b>		

**Tabel 2.** Sampel Dataset\_1

No	Average_Report_Score	Vocational_Major	Attendance_Percentage (%)	Extracurricular_Activities	Academic_Achievements	Graduation_Time	Organization_Participation
1	87	Rekayasa Perangkat Lunak	0,054861	Ya	0	Tepat waktu	Ya
2	89	Teknik Komputer dan Jaringan	0,060417	Tidak	1	Tepat waktu	Tidak
3	88	Rekayasa Perangkat Lunak	0,052083	Ya	0	Tepat waktu	Tidak
4	86	Rekayasa Perangkat Lunak	0,065278	Tidak	0	Tepat waktu	Tidak

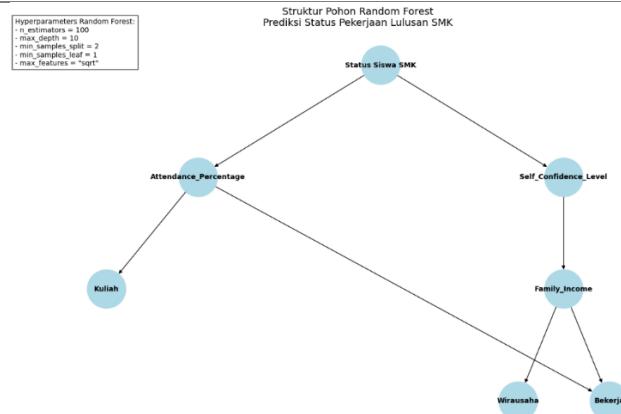
No	Average Report Score	Vocational Major	Attendance (%)	Extracurricular	Academic Achievements	Graduation Time	Organization Participation
5	93	Teknik Komputer dan Jaringan	0,048611	Tidak	1	Tepat waktu	Tidak

Tabel 3. Sampel Dataset\_2

No	Social Media Use	Social Circle	Self-Confidence	Community Involvement	Parent Job	Parent Education	Family Income (Rp)
1	5+ jam	9	1	Ya	Petani	SMP	2-5 juta
2	3-4 jam	1	7	Tidak	Petani	SMA	< 2 juta
3	1-2 jam	6	5	Ya	Petani	Sarjana	2-5 juta
4	3-4 jam	9	6	Ya	Buruh	SMP	2-5 juta
5	3-4 jam	6	9	Tidak	Petani	Sarjana	2-5 juta

Tabel 4. Sampel Dataset\_3

No	Family Size	Economic Condition	Parental Support	Residence	Time to Get Job (Months)	Additional Training	Certification	Job After Graduation	Status
1	6	Menengah	5	Rumah sendiri	3	Tidak	Tidak	Kuliah	
2	2	Kurang Mampu	4	Kos	10	Tidak	Ya	Wirausaha	
3	8	Menengah	6	Rumah sendiri	11	Tidak	Ya	Bekerja	
4	2	Menengah	6	Tinggal dengan kerabat	9	Ya	Ya	Wirausaha	
5	6	Menengah	6	Tinggal dengan kerabat	9	Tidak	Ya	Wirausaha	



Gambar 2 Hyperparameter

### 3 Hasil

#### 3.1 Proses Klasifikasi

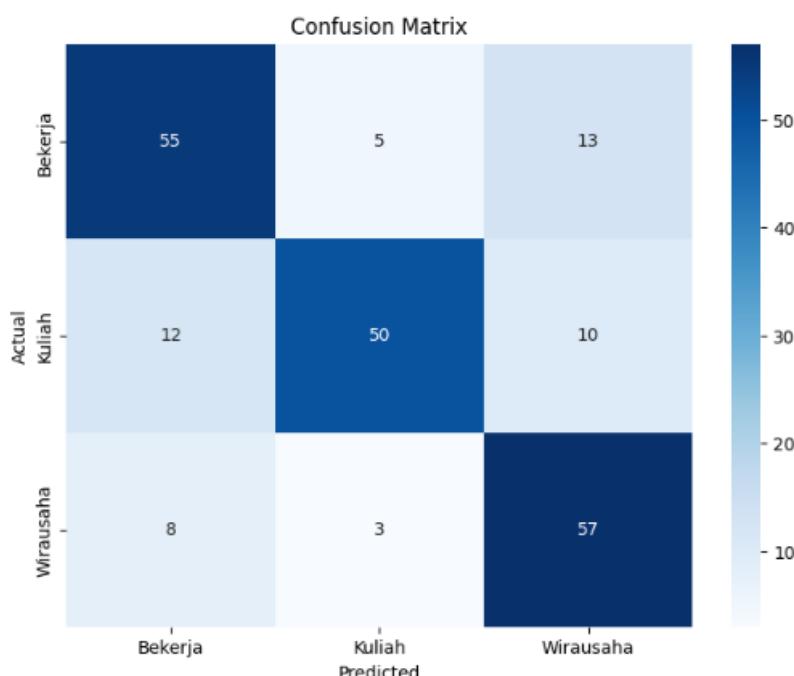
Classification Report:

	precision	recall	f1-score	support
0	0.73	0.75	0.74	73
1	0.86	0.69	0.77	72
2	0.71	0.84	0.77	68
accuracy			0.76	213
macro avg	0.77	0.76	0.76	213
weighted avg	0.77	0.76	0.76	213

Gambar 3 Data Klasifikasi

Setelah dilakukan pelatihan dan pengujian menggunakan algoritma *Random Forest*, menghasilkan tingkat akurasi sebesar 76%. Artinya dari jumlah keseluruhan data set hanya 76% lulusan berhasil diprediksi dengan benar, yang menunjukkan akurasi yang baik dalam mengidentifikasi status pekerjaan, baik itu pekerja maupun non-pekerja. *Precision* dan *Recall* mencapai tingkat akurasi yang tinggi sebesar 86% dalam mengidentifikasi alumni yang tidak bekerja, dan 84% dalam mengidentifikasi alumni yang bekerja. Skor F1 yakni 0,77 yang menunjukkan konsistensi yang baik dalam mengidentifikasi status pekerjaan dan meminimalkan kesalahan klasifikasi. Potensi interpretatif model menunjukkan bahwa model *Random Forest* dapat memberikan prediksi yang akurat dan konsisten untuk berbagai status pekerjaan, menjadikannya alat yang berharga untuk pengambilan keputusan dalam perencanaan pendidikan dan strategi peningkatan pekerjaan.

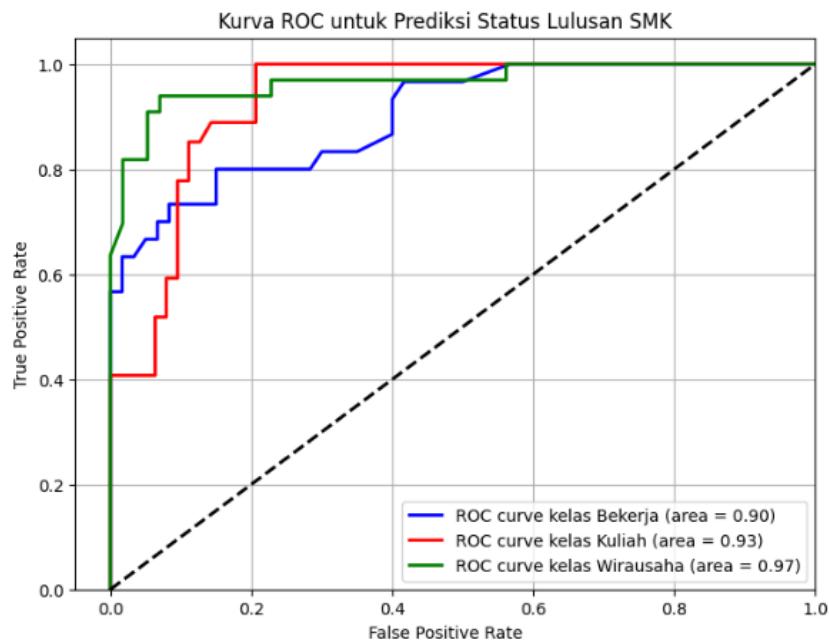
#### 3.2 Confusion Matrix



Gambar 4 Cofusion Matrix

Confusion matrix pada Gambar X menunjukkan bahwa model mampu memprediksi status pekerjaan lulusan SMK dengan ketepatan yang relatif tinggi. Model dapat mengklasifikasikan lulusan yang bekerja, kuliah, maupun berwirausaha dengan cukup akurat, meskipun masih terdapat kesalahan klasifikasi antar kategori, khususnya antara bekerja dan wirausaha. Hal ini mungkin disebabkan oleh overlap karakteristik fitur pada siswa yang bekerja dan berwirausaha, seperti tingkat kepercayaan diri dan pelatihan tambahan.

### 3.3 Kurva ROC untuk Prediksi Status Lulusan SMK

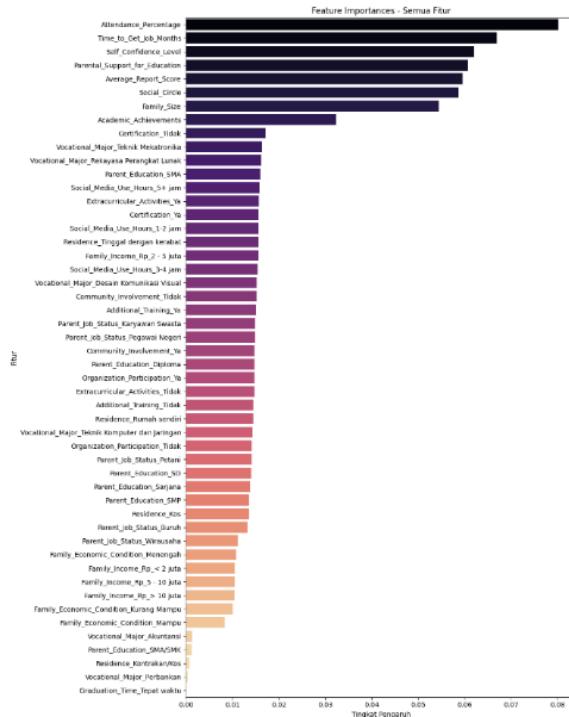


**Gambar 5** Kurva ROC

Model Random Forest dievaluasi menggunakan kurva *Receiver Operating Characteristic* (ROC). Ketiga kelompok memiliki AUC sebesar 0,90, yang menunjukkan kinerja model yang baik dalam membandingkan berbagai status pekerjaan SMK. Kurva ROC menunjukkan kemampuan prediksi model yang stabil dan akurat dalam konteks praktis, seperti evaluasi kinerja karyawan dan implementasi program intervensi.

### 3.4 Feature Importances

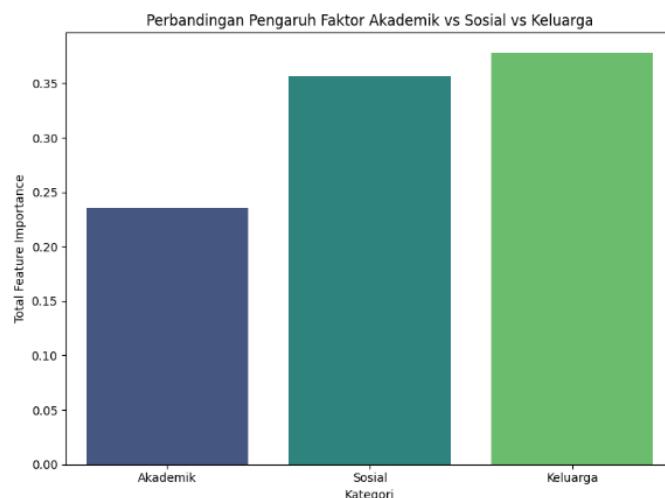
Algoritma *Random Forest* pada penelitian ini tidak hanya memprediksi status pekerjaan tetapi juga memberikan informasi tentang kontribusi relatif setiap fitur (pentingnya fitur) terhadap hasil pekerjaan. Lima fitur yang paling signifikan meliputi persentase status pekerjaan lulusan, waktu untuk mendapatkan pekerjaan, tingkat kepercayaan diri, dukungan orangtua, nilai rata-rata rapot, dan lingkungan sosial. Fitur lain, seperti jurusan kejuruan, pendidikan orang tua, perubahan tempat tinggal, dan waktu kelulusan tidak secara signifikan memengaruhi hasil pekerjaan.



**Gambar 6. Feature Importances**

Dari hasil analisis menggunakan algoritma *Random Forest* menyarankan beberapa rekomendasi untuk meningkatkan pendidikan anak seperti; sekolah harus lebih fokus pada kesejahteraan siswa, disiplin dalam status pekerjaan lulusan di sekolah dapat menunjukkan kurangnya keterlibatan siswa dalam kinerja pekerjaan, keterampilan lunak seperti pelatihan kerja dapat menjadi program yang berharga dalam kurikulum umum serta keterlibatan orangtua juga sangat penting dalam dukungan minat siswa. Sekolah dan pemerintah dapat bekerjasama untuk menyediakan program sosialisasi atau kerja untuk meningkatkan dukungan anak-anak mereka dalam Pendidikan.

### 3.5 Perbandingan Pengaruh Faktor Akademik, Sosial, dan Keluarga.



**Gambar 7. Perbandingan pengaruh faktor Akademik, Sosial, dan Keluarga**

Penelitian ini menyoroti pentingnya faktor-faktor dalam menentukan status pekerjaan siswa SMK, dengan fokus pada tiga kategori utama: akademis, sosial, dan keluarga. Faktor-faktor yang berkontribusi signifikan terhadap status pekerjaan siswa SMK meliputi status sekolah, tingkat pendidikan, keterlibatan keluarga, kondisi ekonomi, lingkungan emosional dan finansial, dan

partisipasi dalam kegiatan analitis. Faktor-faktor sosial, seperti perawatan pribadi, organisasi sosial, penggunaan media sosial, dan partisipasi keluarga, juga berkontribusi terhadap keberhasilan siswa SMK. Interaksi sosial dalam lingkungan SMK memengaruhi pengembangan soft skill, seperti komunikasi, kepemimpinan, dan kerja sama tim. Faktor akademis berkontribusi terhadap kinerja siswa SMK, kinerja akademis, dan kompetensi verbal, tetapi kinerja sekolah lebih penting. Penelitian ini menyarankan bahwa pekerjaan siswa SMK harus berfokus pada aspek kognitif dan sosial, dengan sekolah dan organisasi keluarga berfokus pada konseling, pengembangan keterampilan non-teknis, dan keterlibatan keluarga dalam proses pendidikan untuk meningkatkan kinerja alumni. Pendekatan pendidikan yang lebih holistik dapat menciptakan pekerjaan siswa SMK yang dinamis, mudah beradaptasi, dan kompetitif.

**Tabel 5. Hasil Perbandingan kategori kategori Akademik, Sosial, dan Keluarga**

Kategori	Total Importance	Makna
Keluarga	38,9	Faktor keluarga paling besar pengaruhnya terhadap status pekerjaan lulusan SMK keluarga sangat menentukan.
Sosial	36,8	Faktor sosial (social circle, self-confidence, kegiatan organisasi, dll.) juga sangat signifikan. Ini menunjukkan pentingnya keterlibatan siswa di luar akademik.
Akademik	34,3	Faktor akademik memang berpengaruh, tapi tidak sekuat dua faktor lainnya. Status pekerjaan lulusan dan nilai rapor penting, tapi bukan satu-satunya penentu keberhasilan kerja.

#### 4 Pembahasan

Penelitian ini menggunakan pendekatan kuantitatif dengan model prediktif berbasis algoritma *Random Forest* untuk memprediksi status pekerjaan lulusan SMK. Tahapan penelitian dilakukan sebagai berikut:

##### 4.1 Pre-processing Data:

Pre-processing data merupakan langkah penting dalam penelitian ini untuk memastikan kualitas dataset sebelum digunakan dalam model prediksi. Langkah-langkah pre-processing data meliputi:

- A. Data Cleaning: Mengidentifikasi dan mengatasi masalah data seperti duplikasi, data kosong, atau nilai yang tidak sesuai.

A.1 Menghapus Duplikasi: Untuk memastikan setiap baris data unik, digunakan fungsi logika sederhana:

$$D = Unique(X)$$

Di mana:

X: Dataset awal

D: Dataset setelah duplikasi dihapus

A.2. Mengatasi Data Kosong: Nilai yang hilang NULL diimputasi menggunakan nilai rata-rata mean untuk data numerik:

$$x_i = \frac{\sum_{j=1}^n x_j}{n}$$

Di mana:

$x_i$  : Nilai yang diimputasi

$x_j$  : Nilai pada kolom yang valid

n: Jumlah baris valid

- B. Data Transformation: Data non-numerik diubah menjadi format numerik menggunakan teknik encoding agar dapat diolah oleh algoritma.

B.1 Label Encoding: Memberikan setiap kategori angka integer unik :

$$E(X) = i, \text{ dengan } x \in C, i \in \{0, 1, \dots, |C| - 1\}$$

Dimana :

$\mathcal{C}$  : Himpunan Kategori

$i$  : Indeks Kategori

B.2 One-Hot Encoding: Mengonversi kategori menjadi vektor biner :

$$o(x) = \begin{cases} 1 & \text{jika } x = Ci \\ 0 & \text{lainnya} \end{cases}$$

Dimana :

$Ci$  : kategori ke- $i$

- C. Seleksi Fitur: Memilih fitur yang paling relevan untuk mencegah overfitting. Teknik seperti feature importance pada algoritma *Random Forest* digunakan untuk menentukan fitur-fitur yang memiliki kontribusi signifikan terhadap prediksi. Dua metode utama yang digunakan adalah :

- C.1 Mean Decrease Impurity (MDI): Pentingnya fitur diukur berdasarkan penurunan impurity pada split decision tree:

$$I(Fj) = \sum_{t \in T} p(t) \Delta I(t, Fj)$$

Dimana :

$Fj$  : Fitur ke- $j$

$T$  : Node pada tree tempat fitur  $Fj$  digunakan

$p(t)$  : proporsi data pada node  $t$

$\Delta I(t, Fj)$  : Penurunan impurity pada node  $t$

- C.2 Permutasi Feature Importance: Pentingnya fitur diukur berdasarkan dampaknya terhadap akurasi prediksi setelah fitur diacak:

$$FI(Fj) = \frac{1}{n} \sum_{i=1}^n (A - A^1)_j$$

Dimana :

$FI(Fj)$  : Importance fitur ke- $j$

$A$  : Akurasi model awal

$A^1$  : Akurasi setelah fitur  $Fj$  diacak

## 4.2 Evaluasi Model

Model dievaluasi menggunakan data training dan testing yang dibagi dengan rasio 80:20. Beberapa metrik evaluasi yang digunakan adalah:

1. Akurasi: Mengukur persentase prediksi yang benar terhadap total data.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Presisi: Mengukur ketepatan prediksi positif dibandingkan dengan total prediksi positif.

$$\text{Presisi} = \frac{TP}{TP + FP}$$

3. Recall: Mengukur seberapa baik model dalam menemukan data positif yang sebenarnya.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F1-Score: Rata-rata harmonis dari presisi dan recall.

$$\text{F1 - Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

Keterangan:

- TP (True Positive): Data positif yang terprediksi benar.
- TN (True Negative): Data negatif yang terprediksi benar.
- FP (False Positive): Data negatif yang terprediksi sebagai positif.
- FN (False Negative): Data positif yang terprediksi sebagai negatif.

Model ini diharapkan menghasilkan prediksi yang akurat terkait status pekerjaan lulusan SMK, mencakup kategori bekerja, berwirausaha, atau melanjutkan pendidikan.

## 5 Penutup

Penelitian ini bertujuan untuk menentukan status siswa SMK menggunakan Random Forest, sebuah metode pembelajaran mesin ensemble learning. Model tersebut mengidentifikasi faktor-faktor signifikan yang memengaruhi kinerja siswa setelah lulus SMK. Hasil penelitian menunjukkan bahwa faktor keluarga memengaruhi status pekerjaan siswa SMK secara signifikan, dengan total dampak sebesar 76%. Hasil penelitian ini memiliki implikasi bagi berbagai pemangku kepentingan, termasuk sekolah dan orangtua, yang harus fokus pada prestasi akademik dan secara aktif mengembangkan program *soft skills*. Dari hasil penelitian keluarga memiliki pengaruh yang tinggi sebesar 38.9% dan dikuti oleh faktor sosial sebesar 36.8% memiliki selisih sebesar 2.1%, artinya bahwa tidak hanya dukungan dari lingkungan keluarga tapi sikap sosial anak juga mempengaruhi pekerjaan mereka. Data-data tersebut didukung dari 2 fitur yakni *Parental Support for Education* (Dukungan Orangtua) dan *Self Confidence level* (Tingkat kepercayaan diri). Orangtua hendaknya menyadari peran aktif keluarga dalam pendidikan anak-anaknya, baik moral maupun material, dalam memberikan kontribusi terhadap minat status pekerjaan anak mereka. Tingkat kepercayaan diri pada siswa juga hendak lebih dilatih, hal tersebut mendukung jiwa *Teamwork* di dunia kerja. Keterkaitan antara sekolah dan industri perlu diperhatikan untuk memastikan kompetensi para alumni dalam status pekerjaan mereka. Namun, penelitian ini memiliki keterbatasan, seperti generalisasi karena hanya menggunakan data dari Sidoarjo dan jangka waktu yang singkat (2021-2024). Faktor eksternal seperti kondisi kerja dan kondisi pandemi tidak dipertimbangkan dalam analisis. Penelitian ini juga menyarankan penggunaan algoritma lain seperti Gradient Boosting atau XGBoost untuk model prediksi yang lebih optimal dan komprehensif.

## 6 Referensi

- [1] R. Bakri, N. P. Astuti, and A. S. Ahmar, "Machine Learning Algorithms with Parameter Tuning to Predict Students' Graduation-on-time: A Case Study in Higher Education," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 4, no. 2, pp. 259–265, Dec. 2022, doi: 10.35877/454ri.asci1581.
- [2] R. Nagovitsyn, "Predicting Student Employment in Teacher Education Using Machine Learning Algorithms," *Education and Self Development*, vol. 18, no. 2, pp. 133–148, Jun. 2023, doi: 10.26907/esd.18.2.10.
- [3] M. S. Hassan and A. Goyal, "Predicting Student Placements Using Machine Learning Algorithms [1]," 2023.
- [4] Muhammad Hadiza Baffa, Muhammad Abubakar Miyim, and Abdullahi Sani Dauda, "Machine Learning for Predicting Students' Employability," *UMYU Scientifica*, vol. 2, no. 1, pp. 001–009, Feb. 2023, doi: 10.56919/usci.2123\_001.
- [5] R. A. Safitri and R. Hidayati, "Komparasi Metode K-Nearest Neighbor dan Naïve Bayes untuk Mengklasifikasi Resiko Diabetes Di Posbindu Desa Bulupitu," *SMATIKA JURNAL*, vol. 14, no. 02, pp. 297–303, Dec. 2024, doi: 10.32664/smatika.v14i02.1350.
- [6] N. M. Andini, Y. Findawati, I. R. I. Astutik, and A. Eviyanti, "Implementasi Convolutional Neural Network (CNN) Untuk Mendeteksi Ujaran Kebencian Dan Emosi Di Twitter," *SMATIKA JURNAL*, vol. 14, no. 02, pp. 314–325, Dec. 2024, doi: 10.32664/smatika.v14i02.1346.
- [7] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1817/1/012009.
- [8] S. Dass, K. Gary, and J. Cunningham, "Predicting student dropout in self-paced mooc course using random forest model," *Information (Switzerland)*, vol. 12, no. 11, Nov. 2021, doi: 10.3390/info12110476.
- [9] M. H. Baffa, M. A. Miyim, and A. S. Dauda, "Machine Learning for Predicting Students' Employability," *UMYU Scientifica*, vol. 2, no. 1, pp. 001–009, Feb. 2023, doi: 10.56919/USCI.2123\_001.

- [10] S. Y. Ong, C. Y. Ting, H. N. Goh, A. Quek, and C. L. Cham, "Workplace Preference Analytics Among Graduates," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 233–248, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.17.
- [11] T. T. Huynh-Cam, L. S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, Nov. 2021, doi: 10.3390/a14110318.
- [12] S. Tripathi, "Predicting Unpaid Care Work in India Using Random Forest: An Analysis of Socioeconomic and Demographic Factors," *Social Policy and Society*, 2024, doi: 10.1017/S1474746424000447.
- [13] X. Lu and L. Wei, "Teaching reform and innovation of vocational development and employment guidance courses in colleges and universities based on random forest model," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns.2023.2.00229.
- [14] S. Dass, K. Gary, and J. Cunningham, "Predicting student dropout in self-paced mooc course using random forest model," *Information (Switzerland)*, vol. 12, no. 11, Nov. 2021, doi: 10.3390/info12110476.
- [15] T. A. Cardona, "Predicting Student Degree Completion using Random Forest Predicting Student Degree Completion using Random Forests."
- [16] Muhammad Hadiza Baffa, Muhammad Abubakar Miyim, and Abdullahi Sani Dauda, "Machine Learning for Predicting Students' Employability," *UMYU Scientifica*, vol. 2, no. 1, pp. 001–009, Feb. 2023, doi: 10.56919/usci.2123\_001.
- [17] B. Saini, G. Mahajan, H. Sharma, and Ziniya, "An Analytical Approach to Predict Employability Status of Students," *IOP Conf Ser Mater Sci Eng*, vol. 1099, no. 1, p. 012007, Mar. 2021, doi: 10.1088/1757-899x/1099/1/012007.
- [18] T. T. Huynh-Cam, L. S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, Nov. 2021, doi: 10.3390/a14110318.
- [19] X. Lin, "WITHDRAWN: College student employment data platform based on FPGA and machine learning." [Online]. Available: <https://www.elsevier.co>
- [20] G. Elsharkawy, Y. Helmy, and E. Yehia, "Employability Prediction of Information Technology Graduates using Machine Learning Algorithms." [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [21] R. Bakri, N. P. Astuti, and A. S. Ahmar, "Machine Learning Algorithms with Parameter Tuning to Predict Students' Graduation-on-time: A Case Study in Higher Education," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 4, no. 2, pp. 259–265, Dec. 2022, doi: 10.35877/454ri.ascii1581.