Jurnal Teknik Mesin, Industri, Elektro dan Informatika
Volume 5, Nomor 2, Juni 2026
e-ISSN: 2963-7805; p-ISSN: 2963-8208, Hal. 141-152
DOI: https://doi.org/10.55606/jtmei.v5i2.6118
Tersedia: https://ejurnal.politeknikpratama.ac.id/index.php/jtmei

Penerapan Metode *Support Vector Machine* dan *Random Forest* pada Klasifikasi Multikelas Minat Studi atau Karier Siswa Pasca Lulus SMA (Studi Kasus SMA Se-Kabupaten Kudus)

Eko Hadi Wahyono[1*], Eka Ardhianto[2]
[1,2]Magister Teknologi Informasi, Fakultas Informasi dan Industri, Universitas Stikubank Semarang, Indonesia
Email: ekohadiwahyono0016@mhs.unisbank.ac.id[1], ekaardhianto@edu.unisbank.ac.id[2]

**Penuis Korespondensi: ekohadiwahyono0016@mhs.unisbank.ac.id*

Abstract. *This study aims to compare the performance of the Support Vector Machine (SVM), Random Forest Feature Selection + SVM, and Random Forest Classifier methods in multi-class classification of post-high school students' study or career interests in Kudus Regency. The data used include academic and non-academic variables, such as subject grades, achievements, and parental support. The data processing process was carried out through a preprocessing stage that included handling missing values, categorical data transformation, and normalization. Model evaluation was carried out using the 10-Fold Cross Validation method with accuracy, precision, recall, and F1-score metrics. The results showed that the Random Forest Classifier model had the best performance with an accuracy of 47.63%, precision of 35.20%, recall of 27.52%, and F1-score of 30.9%. Meanwhile, the SVM and Random Forest Feature Selection + SVM models produced similar performance with an accuracy of 46.97% and an F1-score of 15.9%. Variable analysis showed that academic factors, especially Mathematics and Physics grades, were the most influential variables on student interest. However, the model's overall performance is still limited due to data imbalance and the lack of parameter optimization. This study shows that Random Forest is more effective in handling multiclass classification on educational data than SVM..*

Keywords*: Career; Classification; Random Forest; Student Interest; Support Vector Machine (SVM).*

Abstrak. Penelitian ini bertujuan untuk membandingkan kinerja metode Support Vector Machine (SVM), Random Forest Feature Selection + SVM, dan Random Forest Classifier dalam klasifikasi multikelas minat studi atau karier siswa pasca lulus SMA di Kabupaten Kudus. Data yang digunakan meliputi variabel akademik dan non-akademik, seperti nilai mata pelajaran, prestasi, serta dukungan orang tua. Proses pengolahan data dilakukan melalui tahap preprocessing yang mencakup penanganan missing value, transformasi data kategorikal, dan normalisasi. Evaluasi model dilakukan menggunakan metode 10-Fold Cross Validation dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Random Forest Classifier memiliki performa terbaik dengan accuracy sebesar 47,63%, precision 35,20%, recall 27,52%, dan F1-score 30,9%. Sementara itu, model SVM dan Random Forest Feature Selection + SVM menghasilkan performa yang sama dengan accuracy 46,97% dan F1-score 15,9%. Analisis variabel menunjukkan bahwa faktor akademik, khususnya nilai Matematika dan Fisika, merupakan variabel yang paling berpengaruh terhadap minat siswa. Namun demikian, performa model secara umum masih terbatas akibat ketidakseimbangan data dan belum dilakukannya optimasi parameter. Penelitian ini menunjukkan bahwa Random Forest lebih efektif dalam menangani klasifikasi multikelas pada data pendidikan dibandingkan SVM.

Kata Kunci: Karier; Klasifikasi; Minat Siswa; *Random Forest*; *Support Vector Machine (SVM).*

1. LATAR BELAKANG

Perkembangan kecerdasan buatan, khususnya machine learning, telah memberikan kontribusi signifikan dalam bidang pendidikan melalui pemanfaatan data untuk mendukung pengambilan keputusan yang lebih akurat dan berbasis bukti. Pendekatan ini dikenal dalam bidang Educational Data Mining yang mampu mengekstraksi pola dari data siswa untuk memprediksi berbagai aspek, seperti performa akademik dan minat studi atau karier.

Naskah Masuk: 17 Maret 2026; Revisi: 15 April 2026; Diterima:20 Mei 2026; Tersedia: 26 Mei 2026

Dalam konteks pendidikan, pendekatan machine learning dapat membantu institusi pendidikan dalam meningkatakan kualitas pembelajaran serta perencanaan akademik secara lebih efektif dan memberikan rekomendasi secara personal (Tiwari & Jain, 2024; Rahman et al., 2025).

Penentuan minat studi atau karier siswa merupakan keputusan kompleks yang dipengaruhi oleh berbagai faktor, baik akademik maupun non-akademik, seperti nilai, motivasi, serta latar belakang keluarga (Chen & Ding, 2023). Kompleksitas ini menyebabkan pendekatan konvesional krang mampu memberikan hasil yang optimal. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu mengidentifikasikan hubungan antar variabel secara lebih objektid dan sistematis. Hal tersebut menunjukkan bahwa kombinasi faktor akademik dan demografi memiliki peran penting dalam memprediksi perilaku dan arah pilihan siswa di masa depan (Rahman et al., 2025; Tiwari & Jain, 2024).

Metode machine learning seperti Support Vector Machine (SVM) dan Random Forest telah banyak digunakan dalam hal klasifikasi maupun prediksi di data pendidikan karena kemampuannya dalam menagani data kompleks dan berdimansi tinggi (Yagci, 2022; Agung et al., 2026;) Qur'ani et al., 2025). SVM dikenal efektif dalam membangun hyperplane optimal untuk pemisah kelas, sedangkan random forest sebagai metode ensemle mampu meningkatkan akurasi serta memberikan informasi penting terkait kontribusi masing-masing variabel. Pada penelitian Ismail Setiawan et al. 2025, menunjukkan bahwa metode Random Forest memiliki kemampuan untuk mengurang masalah overfitting yang sering terjadi pada decision tunggal sehingga dapat meningkatkan akurasi prediksi.

Meskipun demikian, hasil penelitian sebelumnya menunjukkan hasil yang berbeda seperti Budi Prayoga & Ermatita, 2024 dalam penelitiannya menunjukkan Random Forest mampu memprediksi minat siswa berdsarkan data demografi dan prestasi akademik secara efektif dalam proses klasifikasi. Sedangkan Agung et al., 2026 dalam penelitiannya menunjukkan bahwa Support Vector Machine (SVM) menunjukkan nilai akurasi yang lebih unggul dari Random Forest dalam klasifikasi dataset siswa. Oleh karena itu, diperlukan kajian lebih lanjut untuk membandingkan performa beberapa model klasifikasi serta mengevaluasi pengaruh seleksi fitur dalam meningkatkan kinerja model.

Berdasarkan hal tersebut, Kabupaten Kudus adalah kabupaten yang berada di Jawa Tengah memiliki populasi pelajar yang cukup besar. Tingkat pendidikan di kabupaten ini juga terus mengalami peningkatan dari tahun ke tahun, ditandai dengan semakin banyaknya siswa yang melanjutkan studi ke jenjang yang lebih tinggi.

Namun demikian, terdapat siswa yang memillih untuk langsung bekerja setelah lulus SMA. Penelitian ini bertujuan untuk membandingkan kinerja metode Support Vector Machine (SVM) tanpa seleksi fitur, kombinasi Random Forest sebagai metode seleksi fitur dengan SVM dengsn classifier, serta Random Forest sebagai classifier dalam klasifikasi multikelas minat studi atau karir siswa pasca lulus SMA. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi berbasis data yang dapat membantu pihak sekolah dan pemangku kebijakan dalam ememberikan rekomendasi yang lebih tepat dan terarah.

2. KAJIAN TEORITIS

Machine learning merupakan cabang kecerdasan buatan yang memungkinkan sistem untuk mempelajari pola dari data dan menghasilkan prediksi secara otomatis (Yagci, 2022). Dalam bidang pendidikan, penerapan machine learning dikenal sebagai Educational Data Mining, yang digunakan untuk menganalisis data siswa guna mendukung pengambilan keputusan akademik, seperti prediksi performa dan minat studi atau karier (Kharis & Zili, 2022). Salah satu teknik yang banyak digunakan adalah klasifikasi, yaitu proses pengelompokan data ke dalam kategori tertentu berdasarkan atribut yang dimiliki. Pada penelitian ini digunakan klasifikasi multikelas karena variabel target terdiri dari beberapa kategori, seperti kuliah, kerja, kedinasan, dan tidak berminat.

Beberapa algoritma klasifikasi telah banyak diterapkan dalam data pendidikan, di antaranya Support Vector Machine (SVM) dan Random Forest. SVM merupakan metode yang bekerja dengan membentuk hyperplane optimal untuk memisahkan data antar kelas. Algoritma ini memiliki keunggulan dalam menangani data berdimensi tinggi serta mampu bekerja dengan baik pada data non-linear melalui penggunaan kernel, seperti Radial Basis Function (RBF) (Yagci, 2022). Namun demikian, performa SVM sangat dipengaruhi oleh pemilihan parameter dan distribusi data, sehingga cenderung kurang optimal pada dataset yang tidak seimbang.

Di sisi lain, Random Forest merupakan metode ensemble learning yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan stabilitas model. Metode ini memiliki keunggulan dalam mengurangi overfitting serta mampu menangani data kompleks dengan baik (Hapsari et al., 2024). Selain itu, Random Forest dapat digunakan untuk mengukur tingkat kepentingan variabel (feature importance), sehingga dapat dimanfaatkan dalam proses seleksi fitur. Penelitian sebelumnya menunjukkan bahwa Random Forest sering memberikan performa yang lebih stabil dibandingkan metode tunggal dalam klasifikasi data pendidikan (Dervenis et al., 2022).

Beberapa penelitian juga membandingkan Random Forest dan SVM sebagai classifier dengan tujuan untuk mengetahui performa model mana yang paling baik. Meskipun demikian, hasil yang diperoleh Random forest tidak selalu lebih unggul dari SVM (Agung et al., 2026).

Meskipun berbagai metode telah digunakan, masih terdapat tantangan dalam klasifikasi minat siswa, terutama terkait ketidakseimbangan distribusi kelas (class imbalance) yang dapat menyebabkan model bias terhadap kelas tertentu. Selain itu, pemilihan fitur yang kurang optimal juga dapat mempengaruhi performa model secara keseluruhan. Oleh karena itu, diperlukan analisis lebih lanjut untuk membandingkan performa beberapa metode klasifikasi serta mengevaluasi efektivitas seleksi fitur dalam meningkatkan hasil prediksi.

Berdasarkan kajian tersebut, penelitian ini berfokus pada perbandingan tiga pendekatan, yaitu SVM tanpa seleksi fitur, Random Forest sebagai seleksi fitur yang dikombinasikan dengan SVM, serta Random Forest sebagai classifier langsung. Penelitian ini diharapkan dapat memberikan kontribusi dalam menentukan model yang paling efektif untuk klasifikasi multikelas minat siswa serta mengidentifikasi variabel yang paling berpengaruh dalam pengambilan keputusan siswa pasca lulus SMA

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif untuk membandingkan performa beberapa algoritma machine learning dalam klasifikasi multikelas minat studi atau karier siswa pasca lulus SMA. Pendekatan eksperimen dilakukan dengan membangun dan menguji beberapa model klasifikasi, sedangkan pendekatan komparatif digunakan untuk mengevaluasi perbedaan kinerja antar model berdasarkan metrik tertentu.

Data yang digunakan merupakan data siswa SMA di Kabupaten Kudus, Jawa Tengah, yang diperoleh dari sumber primer dan sekunder. Data primer dikumpulkan melalui kuesioner dan wawancara singkat, sedangkan data sekunder berasal dari data akademik sekolah seperti nilai raport dan dokumentasi lainnya.

Variabel input terdiri dari faktor akademik dan non-akademik, meliputi jenis kelamin, jurusan, nilai Matematika, nilai Fisika, nilai Bahasa Indonesia, nilai PPKN, nilai raport, nilai ijazah, prestasi, dan dukungan orang tua. Variabel target adalah minat siswa yang diklasifikasikan ke dalam empat kategori, yaitu kedinasan, kerja, kuliah, dan tidak berminat. Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses pemodelan.

Tahapan yang dilakukan meliputi: 1) Data cleaning, yaitu penanganan missing value, data duplikat, dan inkonsistensi data, 2) Transformasi data kategorikal, yaitu mengubah data kategorikal menjadi numerik menggunakan teknik *one-hot encoding* dan label encoding, 3) Normalisasi data, khususnya untuk algoritma SVM, menggunakan metode standardisasi (*z-transformation*) agar skala antar fitur menjadi seragam.

Tahapan pereprocessing ini bertujuan untuk memastikan data dalam kondisi optimal sehingga dapat meningkatkan performa model klasifikasi. Selanjutnya penelitian ini menggunakan aplikasi Altair AI Studio sebagai tools utama dalam proses pengolahan data dan pemodelan. Aplikasi ini menyediakan workflow berbasis visual yang mendukung tahapan preprocessing, seleksi fitur, pemodelan klasifikasi, serta evaluasi performa model secara sistematis. Tiga model klasifikasi dilakukan sebagai pembanding

Support Vector Machine (SVM)

Model SVM digunakan tanpa seleksi fitur sebagai model dasar dalam klasifikasi multikelas. SVM bekerja dengan mencari hyperplane optimal untuk memisahkan data antar kelas. Parameter yang digunakan meliputi kernel (linear dan RBF), parameter regularisasi (C), dan gamma.

Random Forest Feature Selection + SVM

Pada model ini, Random Forest digunakan untuk melakukan seleksi fitur berdasarkan nilai *feature importance*. Fitur yang memiliki kontribusi terbesar kemudian digunakan sebagai input pada model SVM. Pendekatan ini bertujuan untuk meningkatkan efisiensi dan performa model dengan mengurangi dimensi data.

Random Forest Classifier

Random Forest digunakan secara langsung sebagai model klasifikasi multikelas. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan dan menghasilkan prediksi berdasarkan mekanisme voting. Parameter utama yang digunakan meliputi jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah minimum sampel untuk pemisahan (*min_samples_split*).

Evaluasi model dilakukan menggunakan metode 10-Fold Cross Validation untuk mengukur kemampuan generalisasi model terhadap data. Metrik evaluasi yang digunakan meliputi: 1) Accuracy, untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan, 2) Precision, untuk mengukur ketepatan prediksi pada masing-masing kelas, 3) Recall, untuk mengukur kemampuan model dalam mendeteksi data yang relevan, 4) F1-score, sebagai ukuran keseimbangan antara precision dan recall.

Penggunaan beberapa metrik evaluasi ini penting terutama pada klasifikasi multikelas untuk memberikan gambaran performa model secara lebih komprehensif.

4. HASIL DAN PEMBAHASAN

Penerapan Support Vector Machine (SVM)

Penerapan metode Support Vector Machine (SVM) dalam penelitian ini dilakukan tanpa menggunakan seleksi fitur. Seluruh variabel input yang telah melalui tahap preprocessing digunakan sebagai dasar pembentukan model klasifikasi.

Model SVM dibangun dengan pendekatan multiclass menggunakan strategi One-vs-Rest (OvR) atau One-vs-One (OvO). Parameter kernel yang digunakan meliputi kernel linear dan Radial Basis Function (RBF), dengan tujuan untuk mengakomodasi kemungkinan pola linear maupun non-linear dalam data.

Berdasarkan hasil evaluasi menggunakan 10-Fold Cross Validation dengan bantuan aplikasi Altair AI Studio, diperoleh nilai performa sebagai berikut:

accuracy: 46.97% +/- 0.19% (micro average: 46.97%)

weighted_mean_recall: 25.00% +/- 0.00% (micro average: 25.00%), weights: 1, 1, 1, 1

weighted_mean_precision: 11.74% +/- 0.05% (micro average: 11.74%), weights: 1, 1, 1, 1

	true Tidak Minat	true Kuliah	true Kedinasan	true Kerja	class precision
pred. Tidak Minat	0	0	0	0	0.00%
pred. Kuliah	47	921	544	449	46.97%
pred. Kedinasan	0	0	0	0	0.00%
pred. Kerja	0	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	0.00%	

Gambar 1. Hasil Penerapan Support Vector Machine (SVM).

Hasil tersebut menunjukkan nilai accuracy 46.97%, precision 11.74%, recall 25%, dan F1-secore 15%, bahwa model SVM memiliki kemampuan klasifikasi yang masih terbatas. Nilai precision yang rendah mengindikasikan bahwa banyak prediksi yang tidak tepat (false positive tinggi), sedangkan nilai recall yang rendah menunjukkan bahwa model belum mampu mengenali seluruh data secara optimal.

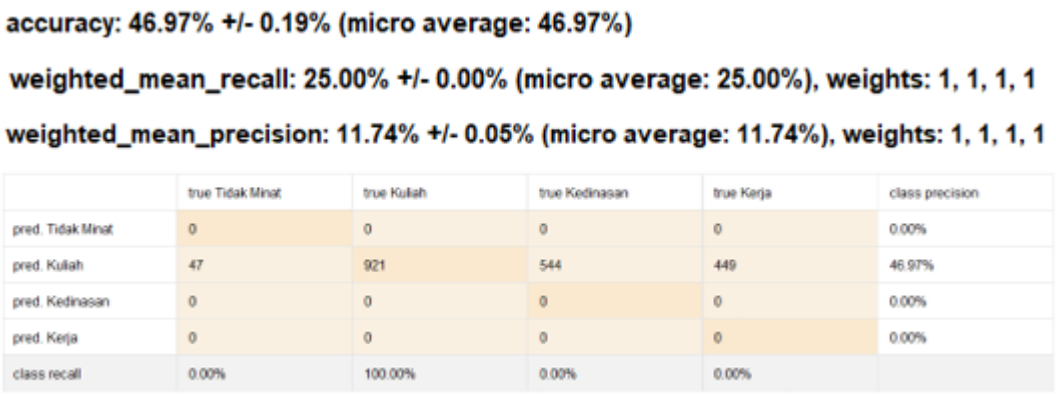
Penerapan Random Forest (Feature Selection) + SVM

Pada model kedua, dilakukan kombinasi antara Random Forest sebagai metode seleksi fitur dan SVM sebagai classifier. Dimana pada tahap seleksi fitur, data siswa dilakukan dengan menerapkan Random Forest yang bertujuan mengetahui nilai importance tertinggi, sehingga dapat diketahu hasil penilaian importance pada setiap variabel atau atribut data siswa sebagai berikut.

Tabel 1. Nilai Importance Atribut.

No	Attribute	Weight
1	NILAI MATEMATIKA	0.219
2	NILAI FISIKA	0.186
3	NILAI BAHASA INDONESIA	0.151
4	NILAI PPKN	0.140
5	NILAI RAPORT	0.117
6	NILAI IJASAH	0.106
7	JENIS KELAMIN = P	0.020
8	JENIS KELAMIN = L	0.015
9	PRESTASI	0.014
10	JURUSAN = IPS	0.010
11	DUKUNGAN ORANG TUA	0.009
12	JURUSAN = Bahasa dan Budaya	0.007
13	JURUSAN = IPA	0.005

Berdasarkan tabel 1, diketahui atribut mana saja yang tergolong tinggi. Pada pemilihan atribut penelitian ini memilih 5 atribut tertinggi antara lain adalah atribut Nilai Matematika, Nilai Fisika, Nilai Bahasa Indonesia, Nilai PPKN, dan Nilai Raport. Setelah atribut sudah dilakuan seleksi atau pemilihan atribut maka dilanjutkan dengan menerapkan model klasifikasi SVM, sehingga hasil yang di dapat sebagai berikut:



accuracy: 46.97% +/- 0.19% (micro average: 46.97%)

weighted_mean_recall: 25.00% +/- 0.00% (micro average: 25.00%), weights: 1, 1, 1, 1

weighted_mean_precision: 11.74% +/- 0.05% (micro average: 11.74%), weights: 1, 1, 1, 1

	true Tidak Minat	true Kuliah	true Kedinasan	true Kerja	class precision
pred. Tidak Minat	0	0	0	0	0.00%
pred. Kuliah	47	921	544	449	46.97%
pred. Kedinasan	0	0	0	0	0.00%
pred. Kerja	0	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	0.00%	

Gambar 2. Hasil Penerapan Random Forest (Featrue Selection) + SVM.

Hasil ini menunjukkan nilai accuracy 46.97%, recall 25%, precision 11.74%, dan F1-score 15.9%, bahwa tidak terjadi peningkatan performa dibandingkan model SVM tanpa seleksi fitur. Fenomena ini menunjukkan bahwa seleksi fitur tidak selalu meningkatkan akurasi model.

Permasalahan utama kemungkinan bukan pada jumlah fitur, tetapi pada distribusi data yang tidak seimbang. Fitur yang dipilih belum cukup mampu membedakan antar kelas secara signifikan. Dengan demikian, penggunaan Random Forest sebagai metode seleksi fitur dalam penelitian ini belum memberikan kontribusi yang signifikan terhadap peningkatan performa model.

Penerapan Random Foresrt Classifier

Model ketiga menggunakan Random Forest secara langsung sebagai classifier tanpa kombinasi dengan metode lain. Random Forest bekerja dengan membangun banyak decision tree dan melakukan voting untuk menentukan hasil akhir klasifikasi.

accuracy: 47.63% +/- 2.37% (micro average: 47.63%)

weighted_mean_recall: 27.51% +/- 1.60% (micro average: 27.52%), weights: 1, 1, 1, 1

weighted_mean_precision: 35.37% +/- 5.51% (micro average: 35.20%), weights: 1, 1, 1, 1

	true Tidak Minat	true Kuliah	true Kedinasan	true Kerja	class precision
pred. Tidak Minat	0	0	0	0	0.00%
pred. Kuliah	46	848	495	383	47.86%
pred. Kedinasan	0	21	30	10	49.18%
pred. Kerja	1	52	19	56	43.75%
class recall	0.00%	92.07%	5.51%	12.47%	

Gambar 3. Hasil Penerapan Random Forest Classifier.

Hasil pemodelan random forest menunjukkan nilai accuracy 47.63%, precision: 35.20%, recall 27.52%, dan F1-score 30.9%. berdasarkan nilai akurasi tersebut, adanya peningkatan performa dibandingkan model sebelumnya, khususnya pada nilai precision dan F1-score. Keunggulan Random Forest dalam penelitian ini disebabkan oleh kemampuan menangani data yang kompleks dan non-linear, tidak terlalu sensitif terhadap noise, dan mampu mengurangi overfitting melalui metode ensemble. Namun demikian, model masih memiliki keterbatasan, terutama dalam mengenali kelas minoritas seperti "Tidak Minat", yang menunjukkan bahwa permasalahan ketidakseimbangan data masih belum teratasi.

Perbandingan Ketiga Model Klasifikasi

Berdasarkan hasil pengujian, dapat dilakukan perbandingan performa pada tabel 2, sebagai berikut:

Tabel 2. Perbandingan Model Klasifikasi.

Model	Accuracy	Precision	Recall	F1-score
SVM	46.97%	11.74%	25%	15.9%
RF (Feature Selection) + SVM	46.97%	11.74%	25%	15.9%
Random Forest	47.63%	35.20%	27.52%	30.9%

Dari table 2 menunjukkan bahwa Random Forest memberikan performa terbaik, SVM dan RF (Feature Selection) +SVM memiliki performa yang sama, setra peningkatan terbesar terjadi pada precision dan F1-score yang lebih tinggi pada Random Forest menunjukkan bahwa model ini memiliki keseimbangan yang lebih baik antara precision dan recall.

Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa ketiga model klasifikasi yang digunakan dalam penelitian ini memiliki performa yang relatif rendah, dengan nilai akurasi di bawah 50%. Meskipun demikian, model Random Forest menunjukkan performa terbaik dibandingkan metode lainnya, khususnya pada nilai precision dan F1-score.

Hasil ini menunjukkan bahwa metode machine learning yang digunakan telah mampu menangkap pola dalam data, namun belum optimal dalam melakukan klasifikasi multikelas secara akurat. Salah satu faktor utama yang memengaruhi performa model adalah ketidakseimbangan distribusi data (class imbalance). Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga kelas minoritas sulit dikenali. Hal ini sejalan dengan penelitian oleh Rifqi Fitriadi & Deni Mahdiana (2023) yang menyatakan bahwa ketidakseimbangan data merupakan salah satu tantangan utama dalam machine learning karena dapat menurunkan kemampuan model dalam mengenali seluruh kelas secara merata.

Pada model Support Vector Machine (SVM), performa yang rendah menunjukkan bahwa algoritma ini kurang optimal dalam menangani data multiclass yang tidak seimbang. Meskipun SVM dikenal memiliki kemampuan generalisasi yang baik, performanya sangat bergantung pada pemilihan parameter dan distribusi data. Temuan ini sejalan dengan penelitian oleh Yagci (2022) yang menyatakan bahwa SVM memiliki performa yang baik juga setelah Random Forest pada data pendidikan, namun memerlukan tuning parameter yang tepat serta distribusi data yang seimbang agar dapat menghasilkan akurasi yang optimal.

Sementara itu, penerapan Random Forest sebagai metode seleksi fitur tidak memberikan peningkatan performa pada model SVM. Hal ini menunjukkan bahwa seleksi fitur tidak selalu berdampak positif terhadap akurasi model. Menurut Hariyanti et al., (2024) efektivitas seleksi fitur sangat bergantung pada karakteristik data dan permasalahan yang dihadapi. Jika permasalahan utama bukan pada banyaknya fitur, melainkan pada distribusi data atau kualitas data, maka seleksi fitur tidak akan memberikan peningkatan signifikan.

Di sisi lain, model Random Forest Classifier menunjukkan performa yang lebih baik dibandingkan model lainnya. Hal ini disebabkan oleh kemampuan Random Forest dalam menangani data yang kompleks dan mengurangi overfitting melalui pendekatan ensemble learning. Hasil ini sejalan dengan penelitian oleh Dervenis et al., (2022) serta Maurya et al., (2021) yang menunjukkan bahwa Random Forest memiliki performa yang stabil dan cenderung unggul dalam klasifikasi data pendidikan dibandingkan algoritma lainnya. Selain itu, analisis variabel menunjukkan bahwa faktor akademik seperti nilai matematika dan fisika memiliki pengaruh paling besar terhadap minat siswa.

Temuan ini konsisten dengan penelitian oleh Chen & Ding, (2023) serta Rahman et al., (2025) yang menyatakan bahwa variabel akademik memiliki kontribusi signifikan dalam memprediksi keputusan dan performa siswa di masa depan.

Namun demikian, hasil penelitian ini juga menunjukkan bahwa model belum mampu mengenali seluruh kelas secara optimal, khususnya pada kelas minoritas seperti "Tidak Minat". Hal ini mengindikasikan perlunya pendekatan tambahan, seperti teknik balancing data (misalnya SMOTE) serta optimasi parameter model, untuk meningkatkan performa klasifikasi.

Secara keseluruhan, penelitian ini memperkuat temuan sebelumnya bahwa Random Forest merupakan metode yang lebih robust dalam menangani data pendidikan yang kompleks. Namun, tanpa penanganan ketidakseimbangan data dan optimasi parameter, performa model tetap terbatas. Oleh karena itu, pengembangan lebih lanjut diperlukan agar model dapat digunakan secara lebih optimal dalam mendukung pengambilan keputusan di bidang pendidikan.

5. KESIMPULAN

Berdasarkan hasil penelitian mengenai penerapan metode klasifikasi multikelas untuk memprediksi minat studi atau karier siswa pasca lulus SMA, dapat disimpulkan bahwa ketiga model yang diuji, yaitu Support Vector Machine (SVM), Random Forest Feature Selection + SVM, dan Random Forest Classifier, mampu digunakan untuk melakukan klasifikasi, namun dengan tingkat performa yang masih terbatas. Model Support Vector Machine (SVM) tanpa seleksi fitur menghasilkan nilai accuracy sebesar 46,97%, precision 11,74%, recall 25%, dan F1-score sebesar 15,9%. Hasil ini menunjukkan bahwa model belum mampu mengklasifikasikan data secara optimal, terutama dalam membedakan antar kelas, serta cenderung bias terhadap kelas mayoritas. Temuan ini konsisten dengan hasil pembahasan yang menunjukkan bahwa SVM sensitif terhadap distribusi data yang tidak seimbang.

Penerapan Random Forest sebagai metode seleksi fitur yang dikombinasikan dengan SVM tidak menunjukkan peningkatan performa dibandingkan model SVM tanpa seleksi fitur. Nilai evaluasi yang dihasilkan tetap sama, yaitu accuracy 46,97% dan F1-score 15,9%. Hal ini menunjukkan bahwa seleksi fitur dalam penelitian ini belum memberikan kontribusi signifikan terhadap peningkatan kinerja model, sebagaimana telah dibahas bahwa permasalahan utama terletak pada distribusi data, bukan pada jumlah atau relevansi fitur. Sementara itu, model Random Forest Classifier memberikan performa terbaik dibandingkan kedua model lainnya, dengan nilai accuracy sebesar 47,63%, precision 35,20%, recall 27,52%, dan F1-score sebesar 30,9%.

Meskipun peningkatan accuracy tidak terlalu signifikan, nilai precision dan F1-score yang lebih tinggi menunjukkan bahwa model ini lebih mampu menghasilkan prediksi yang seimbang dan lebih baik dalam mengurangi kesalahan klasifikasi. Hasil ini sejalan dengan pembahasan yang menunjukkan bahwa Random Forest lebih robust dalam menangani data kompleks dan multiclass.

Selain itu, hasil analisis menunjukkan bahwa variabel akademik, khususnya nilai matematika, fisika, dan bahasa Indonesia, merupakan faktor yang paling berpengaruh dalam menentukan minat siswa. Temuan ini konsisten dengan pembahasan sebelumnya yang menyatakan bahwa faktor akademik memiliki peran dominan dibandingkan faktor non-akademik.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa performa model masih belum optimal, yang dipengaruhi oleh ketidakseimbangan data serta belum dilakukannya optimasi parameter pada masing-masing algoritma. Hal ini juga telah ditegaskan dalam pembahasan bahwa tanpa penanganan terhadap permasalahan tersebut, model cenderung mengalami bias dan kesulitan dalam mengenali seluruh kelas secara merata.

Dengan demikian, penelitian ini menegaskan bahwa Random Forest merupakan metode yang paling efektif di antara model yang diuji dalam konteks klasifikasi multikelas minat siswa. Namun, untuk memperoleh hasil yang lebih optimal, diperlukan pengembangan lebih lanjut, khususnya dalam penanganan ketidakseimbangan data dan optimasi model, agar dapat menghasilkan prediksi yang lebih akurat dan dapat diandalkan dalam mendukung pengambilan keputusan di bidang pendidikan.

DAFTAR REFERENSI

Agung, G. M., Zuama, R. A., & Budi, E. S. (2026). Analysis of Student Academic Performance Using Random Forest and Support Vector Machines. Computer Science, 6(1).

Budi Prayoga, M. H., & Ermatita, E. (2024). Analisis Pemilihan Jurusan pada Calon Siswa SMK Negeri 4 Palembang Pada Faktor Penentu Pemilihan Jurusan Menggunakan Association Rule dan Random Forest. Jurnal Pendidikan Dan Teknologi Indonesia, 4(12). https://doi.org/10.52436/1.jpti.449

Chen, S., & Ding, Y. (2023). A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools. Social Sciences, 12(3). https://doi.org/10.3390/socsci12030118

Dervenis, C., Stoufis, S., Kyriatzis, V., & Fitsilis, P. (2022). Predicting Students' Performance Using Machine LearningAlgorithms. Proceedings of the 6th International Conference on Algorithms, Computing and Systems, 132.

Hapsari, A., Nursuwanda, A. S., Zuhriyah, H., & Vresdian, D. J. (2024). Klasifikasi Kesehatan Mental Mahasiswa Model TMAS dengan Algoritma Decision Tree, Logistic Regression, dan Random Forest. INTEK : Jurnal Informatika Dan Teknologi Informasi, 7(2). https://doi.org/10.37729/intek.v7i2.5690

Hariyanti, E., Hostiadi, D. P., Anggreni, Yohanes Priyo Atmojo, I Made Darma Susila, & Tangkawarow, I. (2024). Analisis Perbandingan Metode Seleksi Fitur pada Model Klasifikasi Decission Tree untuk Deteksi Serangan di Jaringan Komputer. Jurnal Sistem Dan Informatika (JSI), 18(2). https://doi.org/10.30864/jsi.v18i2.615

Ismail Setiawan, Fatah Yasin, I., & Tri Desianti, Y. (2025). Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naïve Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa menggunakan Student Depression Dataset. Jurnal Ilmu Komputer Dan Teknologi, 6(1). https://doi.org/10.35960/ikomti.v6i1.1756

Kharis, S. A. A., & Zili, A. H. A. (2022). Learning Analytics dan Educational Data Mining pada Data Pendidikan. JURNAL RISET PEMBELAJARAN MATEMATIKA SEKOLAH, 6(1). https://doi.org/10.21009/jrpms.061.02

Maurya, L. S., Hussain, M. S., & Singh, S. (2021). Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance. Applied Artificial Intelligence, 35(6). https://doi.org/10.1080/08839514.2021.1901032

Qur'ani, M. D., Setiawan, H., & Kautsar, I. A. (2025). PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MEMPREDIKSI PEMILIHAN KARIR BAGI ALUMNI UMSIDA. Ilmiah Penelitian Dan Pembelajaran Informatika, 10 no 4.

Rahman, N. F. A., Wang, S. L., Ng, T. F., & Ghoneim, A. S. (2025). Artificial Intelligence in Education: A Systematic Review of Machine Learning for Predicting Student Performance. Journal of Advanced Research in Applied Sciences and Engineering Technology, 54(1). https://doi.org/10.37934/araset.54.1.198221

Rifqi Fitriadi, & Deni Mahdiana. (2023). SYSTEMATIC LITERATURE REVIEW OF THE CLASS IMBALANCE CHALLENGES IN MACHINE LEARNING. Jurnal Teknik Informatika (Jutif), 4(5). https://doi.org/10.52436/1.jutif.2023.4.5.970

Tiwari, M., & Jain, N. (2024). STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING ALGORITHMS. ShodhKosh: Journal of Visual and Performing Arts, 5(6). https://doi.org/10.29121/shodhkosh.v5.i6.2024.4552

Yagci, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1). https://doi.org/10.1186/s40561-022-00192-z