

Klasifikasi Prestasi Akademik Siswa Menggunakan Algoritma Random Forest

Bayu Susilo Aji Nugroho^{1*}, Claudea Cahaya Cinta², Abdul Fuad Nurdiansah³, Valen Purwaixna Saputra⁴, Muchamad Syarif Hidayatullah⁵

¹Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

^{1*} 230103053@mhs.udb.ac.id

²Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

² 230103054@mhs.udb.ac.id

³Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

³ 230103043@mhs.udb.ac.id

⁴ Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

⁴ 230103079@mhs.udb.ac.id

⁵ Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

⁵ 230103067@mhs.udb.ac.id

Abstrak— Prestasi akademik termasuk salah satu tolak ukur utama dalam menilai keberhasilan proses pembelajaran siswa. Proses identifikasi kategori prestasi akademik secara manual cenderung memakan banyak waktu dan menjadi kurang efisien, terutama ketika jumlah data siswa terus bertambah. Tujuan dari penelitian ini untuk mengimplementasikan algoritma Random Forest untuk mengklasifikasikan kategori prestasi akademik siswa (GradeClass) berdasarkan variabel StudyTimeWeekly, Absences, Tutoring, dan GPA. Data yang dipakai adalah Student Performance Dataset yang bersumber dari Kaggle dengan jumlah 2.392 data. Tahapan penelitian meliputi preprocessing data, pembagian data ke dalam 80% data latih dan 20% data uji, Pembangunan pada model Random Forest menggunakan 60 decision tree, serta evaluasi melalui Accuracy, Macro F1-Score, Confusion Matrix, dan 5-Fold Cross Validation. Pada hasil pengujian menunjukkan bahwa model Random Forest memperoleh akurasi sebesar 92,3%. Pengujian variasi jumlah decision tree menunjukkan bahwa penggunaan 60 decision tree memberikan keseimbangan yang baik antara akurasi dan waktu pelatihan, sedangkan analisis feature importance menunjukkan bahwa GPA merupakan variabel yang paling dominan dalam memprediksi GradeClass. Oleh karena itu, sistem yang dikembangkan berpotensi mendukung proses evaluasi akademik agar lebih objektif, cepat, dan efisien.

Kata kunci— Random Forest, klasifikasi, prestasi akademik, data mining, GradeClass.

Abstract— Academic achievement is one of the primary indicators used to assess the success of students' learning process. Identifying students' academic performance categories manually is often time-consuming and becomes increasingly inefficient as the amount of student data grows. This study aims to implement the Random Forest algorithm to classify students' academic performance categories (GradeClass) based on the variables StudyTimeWeekly, Absences, Tutoring, and GPA. The dataset used in this study was the Student Performance Dataset obtained from Kaggle, consisting of 2,392 records. The research procedure included data preprocessing, splitting the dataset into 80% training data and 20% testing data, developing a Random Forest model using 60 decision trees, and evaluating its performance using Accuracy, Macro F1-Score, Confusion Matrix, and 5-Fold Cross Validation. The experimental results showed that the Random Forest model achieved an accuracy of 92.3%. Furthermore, experiments with different numbers of decision trees indicated that using 60 decision trees provided a good balance between classification accuracy and training time, while the feature importance analysis revealed that GPA was the most influential variable in predicting GradeClass. Therefore, the proposed system has the potential to support academic evaluation by providing a more objective, efficient, and reliable classification process.

Keywords— Random Forest, classification, academic achievement, data mining, GradeClass.

I. PENDAHULUAN

Prestasi akademik siswa termasuk salah satu indikator penting dalam menilai keberhasilan proses belajar dan kualitas pendidikan[1][2]. Identifikasi dini terhadap siswa yang berpotensi

mengalami penurunan prestasi dapat membantu guru dan pihak sekolah dalam merancang strategi pembelajaran serta intervensi yang lebih tepat sasaran. Untuk mendukung upaya tersebut, perkembangan teknologi informasi memungkinkan institusi pendidikan mengelola

data akademik dalam jumlah besar. Data tersebut kemudian dapat dimanfaatkan melalui pendekatan Educational Data Mining (EDM) untuk menghasilkan analisis yang lebih cepat dan objektif dibandingkan metode evaluasi konvensional[3].

Proses evaluasi prestasi akademik di sekolah kebanyakan masih dilakukan secara manual berdasarkan hasil pengamatan guru dan rekapitulasi nilai siswa. Banyaknya jumlah siswa serta adanya potensi subjektivitas dalam proses penilaian menyebabkan proses evaluasi menjadi kurang efektif [4]. Oleh sebab itu, diperlukannya metode yang mampu mengklasifikasikan prestasi akademik secara otomatis agar proses pengambilan keputusan menjadi lebih efektif dan efisien.

Berbagai penelitian terdahulu telah membandingkan algoritma klasifikasi, seperti Decision Tree dan Naive Bayes, untuk memprediksi prestasi akademik siswa. Hasil penelitian menunjukkan klasifikasi menggunakan Decision Tree menghasilkan akurasi yang lebih tinggi dibandingkan klasifikasi menggunakan Naive Bayes[5]. Meskipun demikian, Decision Tree masih rentan mengalami overfitting karena hanya membangun satu pohon keputusan[6], sementara Naive Bayes menganggap setiap atribut bersifat independen sehingga kurang mampu menangkap hubungan yang kompleks antarvariabel[5]. Oleh karena itu, diperlukan algoritma yang dapat menghasilkan model klasifikasi yang lebih stabil dan tetap mempertahankan tingkat akurasi yang tinggi.

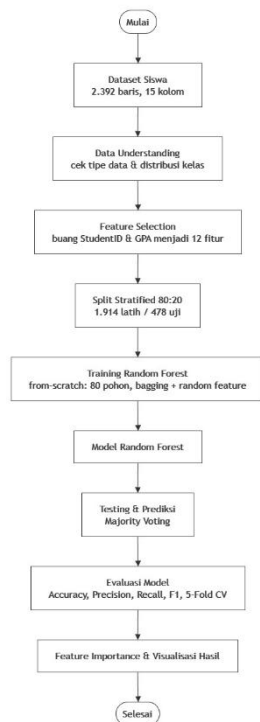
Sebagai solusi alternatif, algoritma Random Forest dipilih karena dapat menghasilkan model klasifikasi yang lebih stabil dan akurat melalui pembentukan banyak pohon keputusan (decision tree)[7][8]. Pada Penelitian sebelumnya algoritma Random Forest mampu mengklasifikasikan prestasi akademik pada mahasiswa dengan akurasi 98,2%. Selain itu, penelitian tersebut menunjukkan bahwa nilai UTS dan UAS menjadi faktor yang paling berpengaruh dalam menentukan hasil klasifikasi

prestasi akademik[9]. Oleh karena itu, algoritma Random Forest dinilai berpotensi baik untuk diterapkan dalam klasifikasi prestasi akademik siswa.

Berdasarkan keunggulan tersebut, algoritma Random Forest dipilih untuk mengklasifikasikan kategori prestasi akademik siswa (GradeClass) menggunakan beberapa atribut yang berkaitan dengan proses pembelajaran. Tujuan dari penelitian ini untuk membuat model klasifikasi prestasi akademik siswa menggunakan algoritma Random Forest yang berdasarkan parameter *StudyTimeWeekly*, *Absences*, *Tutoring*, dan *GPA*, dengan *GradeClass* sebagai variabel target. Keempat atribut tersebut dipilih karena memiliki keterkaitan dengan aktivitas belajar siswa dan telah banyak digunakan dalam penelitian sebagai faktor yang memengaruhi capaian akademik. Melalui penelitian ini diharapkan dapat menghasilkan model klasifikasi yang akurat dan dapat menjadi bahan pertimbangan bagi pihak sekolah dalam melakukan identifikasi dini terhadap siswa yang memerlukan perhatian akademik.

II. METODOLOGI PENELITIAN

Pendekatan kuantitatif dipakai dalam penelitian untuk menerapkan algoritma Random Forest agar dapat mengklasifikasikan kategori prestasi akademik siswa (GradeClass) berdasarkan Student Performance Dataset. Tahapan penelitian ini meliputi pengumpulan data, preprocessing data, pembagian data menjadi data latih dan data uji, pembangunan model menggunakan algoritma Random Forest, serta evaluasi kinerja dari model[10]. Alur tahapan penelitian ditunjukkan pada gambar 1.



Gambar 1 Flowchart Penelitian

A. Sumber Data

Pada penelitian ini menggunakan Student Performance Dataset yang diperoleh dari Kaggle. Dataset ini terdiri atas 2.392 data siswa dan 15 atribut. Sebagian sampel data yang digunakan dalam penelitian ditampilkan pada Tabel 1.

Tabel 1 Sampel Data

Student ID	StudyTime Weekly	Absences	Tutoring	GPA	Grade Class
1001	19,83	7	1	2,93	C
1002	15,41	0	0	3,04	B
1003	4,21	26	0	0,11	E
1004	10,03	14	0	2,05	D
.....					
.....					

B. Prapemrosesan Data

Tahap preprocessing bertujuan untuk menghasilkan data yang berkualitas sehingga dapat digunakan dalam proses pembangunan model *Random Forest* [2] [11].

Tahapan yang dilakukan meliputi:

1. Data Understanding

Pada tahap ini dilakukannya pemeriksaan karakteristik dataset, yang meliputi jumlah data, jumlah atribut, tipe data, distribusi kelas pada variabel *GradeClass*, serta pemeriksaan missing value dan data duplikat.

2. Seleksi Fitur

Tahap seleksi fitur ini dilakukan untuk menentukan variabel yang digunakan pada proses klasifikasi prestasi akademik siswa[12].

3. Pembagian Variabel

Pada tahap ini dilakukan pemisahan dataset menjadi variabel masukan (features) dan variabel target (target) lalu digunakan pada proses klasifikasi menggunakan algoritma *Random Forest*.

C. Pembagian Data

Setelah proses preprocessing, dataset dibagi menjadi data latih (training data) dan data uji (testing data) menggunakan metode hold-out dengan rasio 80:20. Melalui metode ini, sebanyak 80% data digunakan sebagai data latih, sementara 20% sisanya digunakan sebagai data uji untuk mengevaluasi kinerja model [13].

D. Pembangunan Model Random Forest

Random Forest merupakan metode ensemble learning yang membangun banyak decision tree dari sampel hasil bootstrap sampling, dengan setiap pohon menggunakan atribut yang dipilih secara acak (random feature selection) [10] [14]. Setiap decision tree menghasilkan prediksi terhadap data masukan, lalu seluruh prediksi digabungkan menggunakan *majority voting* untuk menentukan hasil klasifikasi

akhir. Pendekatan ini bertujuan meningkatkan kemampuan generalisasi model, sehingga model memiliki risiko overfitting yang lebih rendah, serta performa klasifikasi yang lebih konsisten dibandingkan penggunaan Decision Tree tunggal. [15].

Prediksi akhir pada algoritma Random Forest ditentukan menggunakan mekanisme *majority voting* yang dinyatakan pada Persamaan (1).

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_n(x)\}$$

- $\hat{y}$ = kelas prediksi akhir yang dihasilkan oleh algoritma Random Forest.
- $h_i(x)$ = prediksi yang diberikan oleh pohon keputusan ke- i terhadap data masukan x
- n = banyaknya pohon keputusan (decision tree) yang digunakan dalam proses klasifikasi.
- mode adalah kelas yang memiliki frekuensi kemunculan tertinggi dari seluruh hasil prediksi decision tree, sehingga dipilih sebagai hasil klasifikasi akhir

E. Evaluasi Model

Kinerja model dievaluasi menggunakan metrik Accuracy, Macro F1-Score, dan Confusion Matrix. Accuracy digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, Macro F1-Score untuk mengevaluasi keseimbangan performa klasifikasi pada setiap kelas, sedangkan Confusion Matrix digunakan untuk menganalisis distribusi prediksi benar dan salah pada masing-masing kelas. Selain itu, metode 5-Fold Cross Validation diterapkan untuk menguji kestabilan dan kemampuan generalisasi model terhadap variasi pembagian data. Nilai Akurasi dihitung menggunakan persamaan berikut.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

TP (True Positive) = Jumlah data yang akan diprediksi benar sebagai kelas positif.

TN (True Negative) = Jumlah data yang akan diprediksi benar sebagai kelas negatif.

FP (False Positive) = Jumlah data yang akan diprediksi positif tetapi sebenarnya negatif.

FN (False Negative) = Jumlah data yang akan diprediksi negatif tetapi sebenarnya positif.

Nilai Macro F1-Score diperoleh dari rata-rata nilai F1-Score pada seluruh kelas. Nilai F1-Score dihitung menggunakan Persamaan (2).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Dengan

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

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

Evaluasi menggunakan *Macro F1-Score* dilakukan dengan menghitung nilai rata-rata F1-Score dari setiap kelas, sehingga seluruh kelas memperoleh perlakuan yang setara dalam penilaian kinerja model.

III. HASIL DAN PEMBAHASAN

A. Hasil Prapemrosesan Data

1. Data Understanding

Tahapan pada data understanding ini dilakukan untuk mengetahui karakteristik dari Student Performance Dataset sebelum digunakan dalam proses pembangunan model Random Forest. Dataset yang digunakan terdiri atas 2.392 data siswa dengan 15 atribut yang bersumber dari Kaggle. Setelah itu dilakukan pemeriksaan terhadap

kualitas data untuk memastikan tidak terdapat permasalahan yang berpotensi memengaruhi proses klasifikasi.

Berdasarkan hasil pemeriksaan, tidak terdapat *missing value* pada dataset, sehingga seluruh data dapat digunakan untuk proses penelitian tanpa harus melakukan pembersihan data tambahan. Selain itu, dilakukan identifikasi tipe data dan distribusi kelas pada variabel target GradeClass sebagai dasar dalam proses klasifikasi.

Tabel 2 Pemeriksaan Dataset

No	Informasi Dataset	Hasil
1	Jumlah data	2.392
2	Jumlah atribut	15
3	Missing value	0
4	Data duplikat	0
5	Variabel target	Grade Class

2. Seleksi Fitur

Tahap seleksi fitur bertujuan menentukan atribut yang digunakan dalam proses klasifikasi prestasi akademik siswa. Dari keseluruhan atribut pada dataset, dipilih empat variabel yang sesuai dengan tujuan penelitian, yaitu StudyTimeWeekly, Absences, Tutoring, dan GPA, sedangkan GradeClass digunakan sebagai variabel target.

Pemilihan atribut dilakukan karena variabel-variabel tersebut merepresentasikan aktivitas belajar, tingkat kehadiran, dukungan pembelajaran, dan capaian akademik siswa yang berhubungan dengan kategori prestasi akademik.

Tabel 3 Variabel Feature dan Target

Variabel	Peran
StudyTimeWeekly	Feature
Absences	Feature
Tutoring	Feature
GPA	Feature
GradeClass	Target

3. Pembagian Variabel

Berdasarkan hasil pembagian variabel, atribut StudyTimeWeekly, Absences, Tutoring, dan GPA digunakan sebagai variabel masukan (features), Sedangkan GradeClass berperan sebagai variabel target yang digunakan dalam proses pembentukan model Random Forest. Pembagian variabel dilakukan untuk memisahkan atribut masukan dan variabel target sehingga model dapat mempelajari hubungan antara keduanya dalam proses klasifikasi prestasi akademik siswa.

B. Hasil Pembagian Data

Berdasarkan proses pembagian data menggunakan rasio 80:20, diperoleh sebanyak 1.914 data sebagai data latih (training data) dan 478 data sebagai data uji (testing data). Proporsi ini memberi porsi data latih yang lebih besar sehingga model memperoleh cukup data untuk mempelajari pola klasifikasi, sedangkan data uji dipakai untuk menilai kemampuan model dalam memprediksi data yang belum pernah dipelajari.

Tabel 4 Pembagian Data

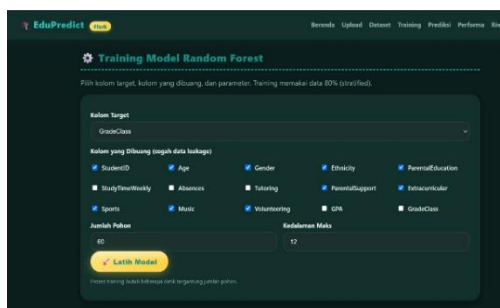
Jenis Data	Jumlah Data
Data latih	1.914
Data uji	478
Total data	2.392

C. Hasil Pembangunan Model Random Forest

Model Random Forest dibentuk menggunakan empat variabel masukan, yaitu StudyTimeWeekly, Absences, Tutoring, dan GPA, dengan GradeClass sebagai variabel target. Pembangunan model dilakukan menggunakan 60 decision tree ($n_estimators = 60$). Setelah proses pelatihan selesai, model berhasil dibangun dan selanjutnya digunakan untuk mengklasifikasikan

prestasi akademik siswa berdasarkan data masukan yang diberikan pengguna.

Gambar 2 menunjukkan halaman pelatihan model Random Forest pada sistem yang dikembangkan. Halaman ini menyediakan fitur untuk menetapkan variabel target, memilih variabel yang digunakan dalam proses pelatihan, serta mengatur jumlah decision tree. Setelah seluruh parameter ditentukan, proses pelatihan dijalankan melalui tombol Latih Model hingga model berhasil dibangun.



Gambar 2 Halaman Training Model

D. Pengaruh Jumlah Decision Tree terhadap Performa Model

Untuk memperoleh jumlah *decision tree* yang optimal, maka dilakukan pengujian menggunakan beberapa variasi jumlah pohon yang disajikan pada Tabel 5.

Tabel 5 Perbandingan Kinerja pada Berbagai Jumlah Decision Tree

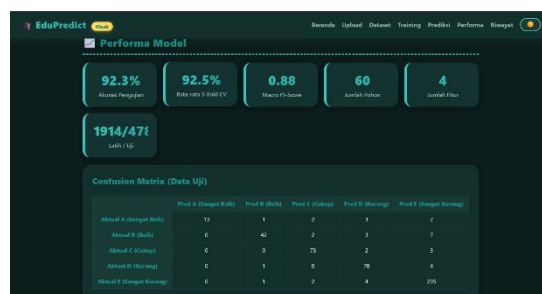
Jumlah Pohon	Accurac y	Macro F1-Score	Waktu Latih
50	92,3%	0,88	6,5 detik
60	92,3%	0,88	7,2 detik
100	92,5%	0,88	12,2 detik
200	92,3%	0,88	24,0 detik

Hasil pengujian tersebut menunjukkan bahwa akurasi model relatif stabil pada jumlah *decision tree*, yang berkisar diantara 92,3% hingga 92,5%, sedangkan waktu pelatihan meningkat seiring bertambahnya jumlah pohon. Oleh karena

itu, penggunaan 60 decision tree dipilih karena mampu memberikan akurasi yang tinggi dengan waktu pelatihan yang lebih efisien.

E. Hasil Evaluasi Model

Hasil evaluasi model Random Forest ditampilkan pada Gambar 3. Berdasarkan hasil pengujian, model memperoleh nilai Accuracy sebesar 92,3%, Macro F1-Score sebesar 0,88, dan rata-rata 5-Fold Cross Validation sebesar 92,5%. Selain itu, sistem juga menampilkan confusion matrix yang menggambarkan distribusi hasil prediksi pada setiap kategori prestasi akademik siswa.



Gambar 3 Halaman Performa Model

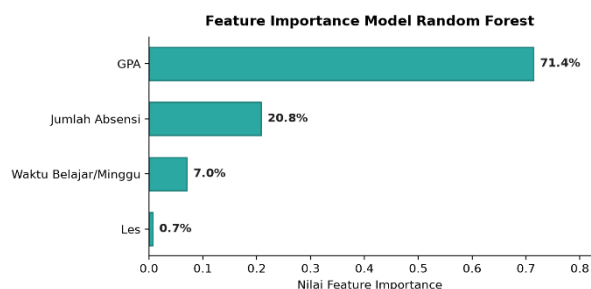
Berdasarkan hasil evaluasi tersebut, model Random Forest mampu mengklasifikasikan prestasi akademik siswa dengan tingkat akurasi yang tinggi. Nilai Macro F1-Score sebesar 0,88 menunjukkan bahwa model memiliki performa yang baik dalam mengenali setiap kategori prestasi akademik secara seimbang. Sementara itu, nilai rata-rata 5-Fold Cross Validation sebesar 92,5% yang menunjukkan jika model memiliki tingkat kestabilan yang baik pada berbagai pembagian data. Hasil confusion matrix memperlihatkan sebagian besar data berhasil diklasifikasikan ke dalam kelas yang sesuai, sehingga menunjukkan bahwa model mampu membedakan kategori prestasi akademik siswa dengan baik.

F. Analisis Feature Importance

Nilai *feature importance* pada Random Forest digunakan untuk mengukur kontribusi setiap variabel dalam memprediksi GradeClass. Hasil perhitungannya disajikan pada Tabel 6 dan Gambar 4.

Tabel 6 Nilai Feature Importance Setiap Variabel

Fitur	Nilai Importance	Kontribusi (%)
GPA	0,714	71,4%
Jumlah Absensi (Absences)	0,208	20,8%
Waktu Belajar/Minggu (StudyTime Weekly)	0,071	7,1%
Les (Tutoring)	0,007	0,7%



Gambar 4 Grafik Feature Importance Model Random Forest

Hasil analisis menunjukkan bahwa GPA merupakan fitur dengan nilai feature importance tertinggi, yaitu sebesar 71,4%. Hal ini terjadi karena GradeClass dibentuk berdasarkan rentang nilai GPA, sehingga kedua variabel memiliki hubungan yang sangat kuat. Selanjutnya, Jumlah Absensi memberikan kontribusi sebesar 20,8%, diikuti oleh Waktu Belajar sebesar 7,1% dan Les sebesar 0,7%. Temuan ini menunjukkan bahwa selain GPA, tingkat kehadiran merupakan faktor yang paling berpengaruh terhadap klasifikasi prestasi akademik siswa.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Random Forest untuk mengklasifikasikan kategori prestasi akademik siswa (GradeClass) berdasarkan empat variabel masukan, yaitu StudyTimeWeekly, Absences, Tutoring, dan GPA. Sistem yang dikembangkan mampu melakukan proses pelatihan model, prediksi kategori prestasi akademik, evaluasi performa model, serta penyimpanan riwayat hasil prediksi melalui aplikasi berbasis web.

Berdasarkan pengujian yang telah dilakukan, model Random Forest menghasilkan akurasi sebesar 92,3%, nilai Macro F1-Score sebesar 0,88, serta rata-rata 5-Fold Cross Validation sebesar 92,5%. Capaian tersebut menunjukkan bahwa model mampu mengklasifikasikan kategori prestasi akademik siswa secara akurat dan memiliki tingkat konsistensi yang baik pada dataset penelitian. Selain itu, pengujian variasi jumlah decision tree menunjukkan bahwa penggunaan 60 decision tree memberikan keseimbangan yang baik antara akurasi dan waktu pelatihan.

Penelitian ini masih memiliki beberapa keterbatasan karena hanya menggunakan satu dataset dan empat variabel masukan, termasuk GPA yang memiliki keterkaitan erat dengan pembentukan GradeClass. Oleh karena itu, penelitian selanjutnya disarankan untuk mengevaluasi kembali pemilihan fitur, membandingkan performa model dengan dan tanpa variabel GPA, dan melakukan perbandingan dengan algoritma lain, seperti Support Vector Machine (SVM), XGBoost, atau Gradient Boosting, memanfaatkan dataset yang memiliki variasi karakteristik lebih luas sehingga model mampu beradaptasi dengan lebih baik terhadap data baru.

REFERENSI

- [1] S. Apandi, "Pemodelan Klasifikasi Siswa Berprestasi dengan Random Forest: Studi Kasus pada Bimbingan

- Belajar,” *Faktor Exacta*, vol. 18, no. 1, p. 63, Jul. 2025, doi: 10.30998/faktorexacta.v18i1.27163.
- [2] M. Firmansyah Tubira, Y. Findawati, R. Dijaya, and Y. Rahmwati, “Prediksi Status Pekerjaan Lulusan SMK Menggunakan Algoritma Random Forest: Analisis Multifaktor Akademis, Sosial, dan Keluarga,” *SMATIKA*, vol. 15, no. 2, p. 15, 2025, doi: 10.32664/smatika.v15i02.1804.
- [3] A. H. Fajar, S. B. Jeffersen, R. Fadilla, A. R. Zaini, A. Sucipto, and B. O. Lubis, “PREDIKSI KELAYAKAN SISWA SMA NEGERI JAKARTA SELATAN MENGGUNAKAN ALGORITMA C.45 BERDASARKAN ZONASI, PRESTASI, DAN RAPOR,” Jakarta, Apr. 2025. [Online]. Available: <https://arsip.siap-ppdb.com>.
- [4] S. Matogu and B. Silalahi, “Optimasi Random Forest untuk Identifikasi Dini Siswa Berisiko di Sekolah Menengah OPTIMIZING RANDOM FOREST ALGORITHM FOR EARLY IDENTIFICATION OF AT-RISK STUDENTS IN SECONDARY SCHOOLS,” *Jurnal Inovasi Informatika (JII)*, vol. 08, no. 01, pp. 34–48, 2026, doi: 10.51170/jii.v8i1.330.
- [5] A.- Husaini, I. Hariyanti, and A. R. Raharja, “Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa,” *Technologia : Jurnal Ilmiah*, vol. 15, no. 2, p. 332, Apr. 2024, doi: 10.31602/tji.v15i2.14381.
- [6] Z. Fatah and A. Ibrohim, “Klasifikasi Prestasi Akademik Siswa Menggunakan Algoritma Decision Tree dengan RapidMiner,” Situbondo, 2026.
- [7] A. Fatunnisa and H. Marcos, “Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest,” *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 1, pp. 101–111, Apr. 2024, doi: 10.34010/jamika.v14i1.12114.
- [8] S. Chayy Bilqisth and R. M. Ikhsanuddin, “ANALISIS PERBANDINGAN AKURASI KLASIFIKASI KEPUASAN SISWA TERHADAP KINERJA GURU MENGGUNAKAN ALGORITMA SVM, C4.5, DAN RANDOM FOREST,” *Jurnal Teknik Elektro dan Informatika*, vol. 20, pp. 37–44, 2025.
- Sofiyatus Zawiyah, Lailatul Qodriyah, and Moh. Badri Tamam, “Klasifikasi Prestasi Akademik Mahasiswa Menggunakan Metode Random Forest,” *Journal of Digital Business and Information Technology*, vol. 1, no. 2, pp. 61–71, Jan. 2025, doi: 10.23971/jobit.v1i2.317.
- H. Ghani, R. Santos, F. Supriadi, and D. Setiadi, “PREDIKSI JUMLAH PENGUNJUNG PERPUSTAKAAN MENGGUNAKAN ALGORITMA RANDOM FOREST,” Sumedang, Apr. 2026.
- M. A. Rachman, D. Ridho Maulana, B. Timothy, and R. Annisa, “JURNAL MEDIA INFORMATIKA [JUMIN] Aplikasi Prediksi Tingkat Kelulusan Mahasiswa Berdasarkan Data Akademik dan Demografi Menggunakan Algoritma Klasifikasi Random Forest,” Pontianak, Jan. 2026.
- J. Kuswanto, H. Lukmanul, A. Info, and K. Kunci, “Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa,” vol. 5, no. 1, pp. 262–270, 2025, doi: 10.51454/del.
- R. B. Hidayah, K. Latifah, and B. A. Herlambang, “PREDIKSI GAYA BELAJAR MENGGUNAKAN ALGORITMA RANDOM FOREST DENGAN OPTIMASI HYPERPARAMETER TUNING BERBASIS WEB,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 864–873, Jul. 2025, doi: 10.36341/rabit.v10i2.6393.
- Y. Caesarita, N. Wahyudi, and A. Kurniawan Faisol, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Perbandingan Algoritma Random Forest Dan XGBoost,” *INTER TECH*, vol. 4, no. 1, pp. 38–45, May 2026, doi: 10.54732/4nzaxe76.
- E. Novianto, S. Suhirman, and D. Prasetyo, “PERBANDINGAN METODE KLASIFIKASI RANDOM FOREST DAN SUPPORT VECTOR MACHINE DALAM MEMPREDIKSI CAPAIAN STUDI MAHASISWA,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 1821–1833, Nov. 2024, doi: 10.29100/jipi.v9i4.5423.