




bima Amikom tes

-  Norhikmah
-  S1-Sistem Informasi
-  Universitas Amikom

Document Details

Submission ID

trn:oid::1:3254905715

Submission Date

May 20, 2025, 5:54 AM GMT+7

Download Date

May 20, 2025, 5:57 AM GMT+7

File Name

Paper_ke_Jurnal_SISTEMASI_1.docx

File Size

354.2 KB

10 Pages




3,926 Words

24,867 Characters

14% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Top Sources

- 0%  Internet sources
- 14%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 0% Internet sources
- 14% Publications
- 0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

- 1

Publication

Angelos Charitopoulos, Maria Rangoussi, Dimitrios Koulouriotis. "Blending E-Lear... <1%
- 2

Publication

Arif Mudi Priyatno, Yunia Ningsih, Arnes Yuli Vandika, Muhammadong Muhamm... <1%
- 3

Publication

Syahrul Sobari, Ade Irma Purnamasari, Agus Bahtiar, Kaslani Kaslani. "MENINGK... <1%
- 4

Publication

Zhafira Zafitri, Muhammad Ihsan Jambak. "Karakteristik Pembatalan Reservasi K... <1%
- 5

Publication

Chao Ma. "Improving the Prediction of Student Performance by Integrating a Ran... <1%
- 6

Publication

Abdellatif HARIF, Moulay Abdellah KASSIMI. "Predictive Modeling of Student Perf... <1%
- 7

Publication

Xuehui Zhang, Lin Yang. "A Convolutional Neural Network-Based Predictive Mode... <1%
- 8

Publication

Elfira Aulia Septrian, Erna Zuni Astuti. "Application of Random Forest Method for ... <1%
- 9

Publication

Mengyao Chen, Zhengqi Liu. "Predicting performance of students by optimizing t... <1%
- 10

Publication

A. Reza Baehaqa Jamroni Jamroni, Wahyu Hadikristanto, Muhamad Fatchan. "An... <1%
- 11

Publication

Tai Le Quy, Gunnar Friege, Eirini Ntoutsis. "Chapter 2 A Review of Clustering Model... <1%

12	Publication	Achmad Ridwan, Uly Muzakir, Safitri Nurhidayati. "Optimizing E-commerce Inven...	<1%
13	Publication	Ali Dongan Harahap, Didi Juardi, Agung Susilo Yuda Irawan. "RANCANG BANGUN ...	<1%
14	Publication	Anggi Putri Meriani, Alam Rahmatulloh. "PERBANDINGAN GATED RECURRENT UN...	<1%
15	Publication	Dewi Wardani, Chairul Achmad. "SSTI: Semantic Similarity to detect Novelty of Th...	<1%
16	Publication	Yasmeen Aldossary, Majeed Ebrahim, Salman Alhaddad, Ahmed M. Zeki. "Predicti...	<1%
17	Publication	Agustia Kuspita Aryanti, Rudy Herteno, Fatma Indriani, Radityo Adi Nugroho, Mul...	<1%
18	Publication	Yonal Supit. "Penerapan Algoritma Support Vector Machine (SVM) untuk Deteksi ...	<1%
19	Publication	Putra, Rudy Syah. "Development of electrokinetic remediation for caesium: A fea...	<1%
20	Publication	Seligmann, Judy. "Rural Teachers' Learning of English in a Distance Education Pro...	<1%
21	Publication	Zalabila Rani, Bain Khusnul Khotimah. "ANALISIS SENTIMEN TERHADAP KARAPA...	<1%
22	Publication	Fitrah Amelia Ramelan, Lukman Hakim. "Perbandingan Prediksi terhadap Pening...	<1%
23	Publication	Rahmad Firdaus, Husnul Habibie, Yoze Rizki. "Implementasi Algoritma Random F...	<1%
24	Publication	Wulan Patria Saroinsong, Muhamad Nurul Ashar, Irena Y. Maureen, Lina Purwani...	<1%
25	Publication	Hastangka Hastangka, Yayuk Hidayah. "KEBIJAKAN DAN MANAJEMEN PENDIDIKA...	<1%

26	Publication	Mohammad Farid Naufal, Selvia Ferdiana Kusuma, Kevin Christian Tanus, Raynal...	<1%
27	Publication	Yoga Nanda Khoiril Umat, Dhiaz Rusyda Nafsyi, Diana Kusumaningsih, Lukman H...	<1%
28	Publication	Defri Yona, Dhira Kurniawan, Ledhyane Ika Harlyan, Adrian Sakti Gangsar Pinilih,...	<1%
29	Publication	Warner, Elise. "Will These Feelings Last Forever? Examining the Relationship Betw...	<1%
30	Publication	Xue Han, Nan Hu. "Prediction of one- and three-months yoga practices effect on c...	<1%
31	Publication	Barnes, Emily J.. "From Data to Decisions: Machine Learning in Higher Education",...	<1%
32	Publication	De Kock, Martha Georgina. "Inligtingontsluiting in 'n Geintegreerde Biblioteekre...	<1%
33	Publication	Gibran Nasrizal Masacgi, Muhammad Syaifur Rohman. "Optimasi Model Algoritm...	<1%
34	Publication	Patria Gita Laksamana, Willy Prihartono, Fathurrohman .. "IMPLEMENTASI NAÏVE ...	<1%
35	Publication	Hamidah Lutfiyanti Maharani, Syahiduz Zaman. "Implementasi Metode Random ...	<1%

Klasifikasi Performance Siswa untuk Potensi Kelulusan Menggunakan Algoritma Random Forest

Student Performance Prediction for Graduation Potential Using the Random Forest Algorithm

¹Puguh Hasta Gunawan*, ²Penulis2, ^NPenulisN

^{1,2,3}Magister Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia

Jl. Kaliurang km 14.5, Sleman, Yogyakarta 55584

*e-mail: 23917022@students.uii.ac.id

(received: ?, revised: ?, accepted: ? diisi oleh editor)

Abstrak

Prediksi potensi kelulusan mahasiswa melibatkan evaluasi berbagai faktor akademik dan non-akademik yang mempengaruhi performa mahasiswa. Data yang digunakan dalam penelitian ini mencakup informasi seperti nilai akademik, kehadiran, waktu belajar, serta faktor demografis dan sosial yang relevan dalam mempengaruhi capaian studi mahasiswa. Proses penelitian diawali dengan tahap pra-pemrosesan data, yang meliputi penghapusan data yang tidak relevan, pengkodean variabel kategorikal menjadi nilai numerik. Algoritma *Random Forest* digunakan sebagai model utama dalam penelitian ini, dengan konfigurasi parameter terbaik yang diperoleh melalui proses *tuning*, yaitu *max_depth* sebesar 15 dan *random_state* sebesar 42 untuk memastikan kestabilan hasil. Hasil evaluasi model yang dilakukan menggunakan confusion matrix menunjukkan bahwa model *Random Forest* yang diterapkan mampu mencapai akurasi keseluruhan sebesar 89%. Model menunjukkan kinerja yang sangat baik dalam memprediksi mahasiswa yang lulus (*Pass*), dengan *recall* sebesar 87% dan *f1-score* sebesar 91%, yang menunjukkan keseimbangan yang baik antara kemampuan model mengidentifikasi mahasiswa yang benar-benar lulus dan tingkat presisinya. Sementara itu, dalam memprediksi mahasiswa yang tidak lulus (*Fail*), model juga menunjukkan performa yang cukup baik, dengan *recall* sebesar 92% dan *f1-score* sebesar 84%. Pada Penelitian selanjutnya disarankan melakukan pelatihan model menggunakan *random forest* lebih dari 2 parameter.

Kata kunci: *Random Forest*, Kinerja Mahasiswa, Klasifikasi, Analisis Data Pendidikan, Prediksi Kelulusan.

Abstract

Predicting student graduation potential involves evaluating various academic and non-academic factors that influence student performance. The data used in this study includes information such as academic grades, attendance, study time, as well as demographic and social factors that are relevant to students' academic achievements. The research process begins with a data preprocessing stage, which includes removing irrelevant data and encoding categorical variables into numerical values. The Random Forest algorithm is used as the main model in this study, with the best parameter configuration obtained through a tuning process, specifically max_depth of 15 and random_state of 42, to ensure result stability. The model evaluation, conducted using a confusion matrix, shows that the applied Random Forest model achieved an overall accuracy of 89%. The model demonstrates excellent performance in predicting students who graduate (Pass), with a recall of 87% and an f1-score of 91%, indicating a good balance between the model's ability to identify actual graduates and its precision level. Meanwhile, in predicting students who do not graduate (Fail), the model also shows fairly good performance, with a recall of 92% and an f1-score of 84%. For future research, it is recommended to train the model using Random Forest with more than two parameters to further optimize its performance.

Keywords: *Random Forest*, Student Performance, Classification, Educational Data Mining, Graduation Prediction.

<http://sistemasi.ftik.unisi.ac.id>

1 Pendahuluan (or Introduction)

Kinerja akademik siswa merupakan indikator utama dalam menilai keberhasilan suatu sistem. Kinerja akademik mahasiswa merupakan indikator strategis dalam menilai efektivitas sistem pendidikan tinggi, tidak hanya bagi mahasiswa secara individu, namun juga bagi institusi seperti perguruan tinggi. Dalam beberapa tahun terakhir, penerapan pendekatan berbasis pembelajaran mesin telah banyak dilakukan untuk menganalisis data akademik mahasiswa guna memprediksi potensi kelulusan mereka. Hal ini menjadi penting karena Kinerja siswa dipengaruhi oleh berbagai faktor, termasuk nilai akademik, kehadiran, waktu belajar, dukungan dari keluarga, dan akses ke fasilitas pendidikan. Oleh karena itu, penting untuk memahami pola-pola yang mempengaruhi performa siswa agar dapat diidentifikasi potensi masalah sejak dini, sehingga intervensi yang tepat dapat dilakukan. Evaluasi ini tidak hanya melibatkan hasil akademik berupa nilai, tetapi juga mencakup faktor-faktor lain seperti keterlibatan dalam kegiatan akademik, keterampilan riset, aktivitas ekstrakurikuler, serta faktor-faktor non-akademik seperti latar belakang ekonomi dan sosial mahasiswa [1]. Memahami pola-pola tersebut memungkinkan institusi pendidikan untuk mengidentifikasi potensi keterlambatan studi sejak dini dan melakukan intervensi yang tepat sasaran. Dari sisi institusi, urgensi prediksi kelulusan mahasiswa ini terkait erat dengan pencapaian Indikator Kinerja Utama (IKU) perguruan tinggi, khususnya IKU 1 (Lulusan mendapatkan pekerjaan yang layak) dan IKU 2 (Mahasiswa mendapatkan pengalaman di luar kampus), di mana capaian IKU tersebut sangat bergantung pada kelulusan mahasiswa yang tepat waktu. Apabila mahasiswa mengalami keterlambatan atau kemoloran studi, mereka tidak hanya menghadapi risiko kerugian secara finansial karena bertambahnya beban biaya studi, tetapi juga kehilangan momentum karier, kepercayaan diri, dan peluang berdaya saing di dunia kerja yang dinamis. Selain itu, kemoloran studi juga berkontribusi negatif terhadap citra dan kinerja universitas, baik dari sisi akreditasi maupun peringkat institusi. Kelulusan penuh pada setiap angkatan siswa menjadi salah satu tujuan utama institusi pendidikan, namun berbagai faktor dapat mempengaruhi kelulusan, yang membuat prediksi terhadap potensi hambatan menjadi tantangan tersendiri [2],[3].

Berdasarkan kajian literatur yang telah dilakukan, mayoritas penelitian terkait prediksi kelulusan mahasiswa masih didominasi oleh pendekatan yang berfokus pada data akademik semata, seperti yang dilakukan oleh Andrianof et al. (2025) yang memanfaatkan data IPK, kehadiran, dan nilai mata kuliah sebagai prediktor utama [4], serta yang menggunakan faktor hafalan, kehadiran, dan nilai ujian dalam memprediksi kelulusan santri Tahfidz [5]. Meskipun pendekatan ini telah menunjukkan performa yang cukup baik dari sisi akurasi prediksi, penelitian-penelitian tersebut masih memiliki keterbatasan dalam mengintegrasikan faktor-faktor non-akademik yang juga berpengaruh signifikan terhadap kelulusan mahasiswa. Faktor-faktor seperti kondisi sosial ekonomi, motivasi belajar, keterlibatan dalam organisasi atau kegiatan kampus, hingga kondisi psikologis mahasiswa masih jarang diadopsi sebagai variabel pendukung dalam model prediksi yang ada.

Algoritma machine learning seperti *Decision Tree*, *k-Nearest Neighbors* (k-NN), *Support Vector Machine* (SVM) berfungsi untuk klasifikasi data, namun selain keempat algoritma tersebut, algoritma yang dapat digunakan adalah *Random Forest*, yang dikenal memiliki keunggulan dalam menangani data yang beragam serta mencegah *overfitting* [6]. Dengan kemampuan analisisnya, algoritma ini memberikan peluang besar untuk meningkatkan akurasi prediksi, yang dapat mendukung pengambilan keputusan berbasis data dalam dunia pendidikan. *Random Forest* unggul dalam tugas klasifikasi, khususnya dalam memprediksi performa akademik siswa. Sebagai contoh, algoritma ini menunjukkan hasil yang lebih baik dibandingkan metode lain dalam mengidentifikasi siswa yang berpotensi tidak lulus berdasarkan pola dalam data akademik [7]. Integrasi pemodelan prediktif, khususnya dengan menggunakan *Random Forest Classifier* (RFC) [8]. Pendekatan hibrida semacam ini menjadi solusi untuk memperbaiki kelemahan masing-masing algoritma.

Pendekatan pembelajaran mesin, baik terawasi (*Random Forest*) maupun tidak terawasi (*Clustering*), diterapkan untuk menganalisis hubungan tersebut. Hasilnya mengungkap pola menarik terkait partisipasi siswa dan menunjukkan bahwa partisipasi serta skor evaluasi merupakan indikator yang efektif dalam memprediksi kinerja akademik siswa [9]. Kerangka *Random Forest* memberikan potensi kemajuan besar dalam analitik prediktif bagi institusi pendidikan. Hasil penelitian ini menyoroti keefektifan metode yang diusulkan sekaligus membuka peluang baru untuk meningkatkan presisi dan

<http://sistemasi.ftik.unisi.ac.id>

keandalan dalam memprediksi performa akademik siswa[10]. Algoritma Random Forest lebih unggul karena kemampuannya dalam menangani data dengan dimensi yang besar dan menghasilkan prediksi yang lebih akurat [3]. Algoritma *Random Forest* dapat digunakan untuk memprediksi performa akademik mahasiswa Program Studi Informatika Universitas Baturaja. Data yang digunakan mencakup variabel akademik (nilai mata kuliah semester 1 dan 2) dan demografis (jenis kelamin, kelas, angkatan), dengan target prediksi berupa Indeks Prestasi (IP) semester 3[11].

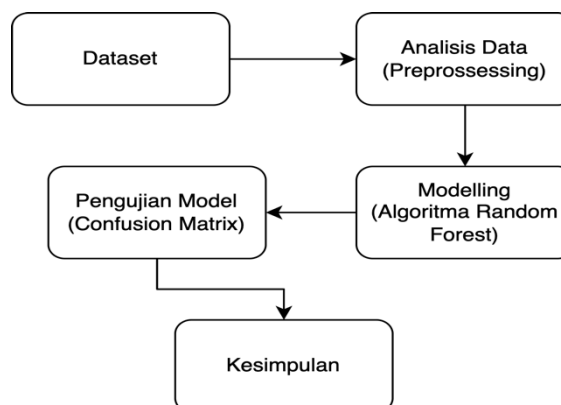
Algoritma ini juga dapat memproses berbagai variabel input secara paralel dan memberikan hasil yang lebih stabil dibandingkan algoritma lain seperti Decision Tree tunggal atau K-NN [2]. Selain itu, teknik seperti *Correlation-Based Feature Selection* (CFS) dapat digunakan untuk meningkatkan akurasi prediksi dengan memilih fitur-fitur yang paling relevan [12].

2 Tinjauan Literatur (or Literature Review)[11]

Berbagai studi menunjukkan bahwa algoritma *Random Forest* telah diterapkan secara luas dalam berbagai bidang prediksi dengan hasil yang menjanjikan. *Random Forest* untuk memprediksi waktu tunggu pasien rawat jalan di RSJ Dr. Soeharto Heerdjan dan mencapai akurasi 97,6%, menunjukkan kemampuan model dalam mengelola data pelayanan kesehatan yang kompleks [13]. Mengaplikasikan algoritma yang sama untuk memprediksi kelulusan santri Tahfidz di Pondok Pesantren Al-Kautsar, dengan akurasi sangat tinggi yaitu 99,64%, mengungkap bahwa faktor non-akademik seperti hafalan dan kehadiran berpengaruh signifikan [5]. *Random Forest* dalam prediksi tarif penerbangan berbasis *AutoML* dan mendapatkan R2 sebesar 85,87%, lebih unggul dibandingkan *Logistic Regression* dan *Gradient Boosting* [14]. Di bidang pendidikan tinggi, menerapkan *Random Forest* untuk prediksi kelulusan mahasiswa di Indonesia, dengan akurasi 87,5%, yang menegaskan bahwa IPK dan kehadiran menjadi faktor kunci dalam ketepatan prediksi [4]. Sementara itu, Swarga & Asyhari (2025) mengembangkan sistem klasifikasi kualitas air berbasis web dengan *Random Forest* dan berhasil mencapai akurasi 78%, memberikan solusi praktis untuk monitoring kualitas air secara otomatis [15]. Di sektor properti, membuktikan bahwa *Random Forest* efektif memprediksi harga properti di Indonesia dengan MAE sebesar Rp2,48 miliar dan RMSE Rp2,89 miliar, di mana faktor jarak ke pusat kota menjadi determinan utama [16]. Keseluruhan studi ini menegaskan keunggulan *Random Forest* dalam memproses data yang kompleks, menangani variabel numerik dan kategorikal secara simultan, serta memberikan prediksi yang lebih stabil dan akurat dibandingkan metode tradisional, sekaligus membuka peluang pengembangan model lebih lanjut dengan integrasi data real-time maupun *AutoML*, dengan menggunakan 2 parameter yaitu $max_depth=15$, $random_state=42$

3 Metode Penelitian (or Research Method)

Metodologi penelitian yang digunakan dalam penelitian ini sebagai berikut:



Gambar 1. Alur Penelitian

3.1 Penelitian tahap identifikasi dan pengumpulan data

- a. Menggunakan pendekatan kuantitatif dengan metode survei dan analisis data sekunder. Desain penelitian yang dipilih adalah desain eksplanatori karena bertujuan untuk mengidentifikasi dan menjelaskan faktor-faktor yang mempengaruhi kelulusan mahasiswa.
- b. dataset didapatkan melalui *uci edu student performance* “<https://archive.ics.uci.edu/dataset/320/student+performance>”.

3.2 Analisis data (*Preprocessing*)

Pada tahapan selanjutnya dilakukan *Preprocessing* pada dataset seperti *delete, encode, split data*.

3.3 *Modelling*

Menggunakan Algoritma *Random Forest* untuk memprediksi kelulusan mahasiswa berdasarkan berbagai faktor yang mempengaruhi hasil dari data akademik mereka. *Random Forest* adalah metode *ensemble learning* yang menggabungkan beberapa *decision tree* untuk menghasilkan prediksi yang lebih akurat dan kuat terhadap *overfitting*. Setiap *decision tree* dalam *random forest* dilatih menggunakan subset acak dari data dan fitur, dan hasil dari semua tree digabungkan menggunakan teknik voting (untuk klasifikasi) atau averaging (untuk regresi).

3.4 Pengujian model menggunakan *confusion matrix*

Confusion matrix memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan data ke dalam kategori yang benar, dengan membandingkan hasil prediksi model dengan data aktual, untuk melihat tingkat akurasi dan ketepatan dalam prediksi kelulusan mahasiswa. *Confusion matrix* adalah tabel 2x2 (untuk klasifikasi biner) yang terdiri dari empat elemen utama:

1. *True Positive (TP)*: Jumlah mahasiswa yang sebenarnya lulus dan diprediksi lulus oleh model.
2. *True Negative (TN)*: Jumlah mahasiswa yang sebenarnya tidak lulus dan diprediksi tidak lulus oleh model.
3. *False Positive (FP)*: Jumlah mahasiswa yang sebenarnya tidak lulus, tetapi diprediksi lulus oleh model (*error tipe I*).
4. *False Negative (FN)*: Jumlah mahasiswa yang sebenarnya lulus, tetapi diprediksi tidak lulus oleh model (*error tipe II*).

4 Hasil dan Pembahasan (or Results and Analysis)

Pada bagian ini, akan dipaparkan hasil dari proses klasifikasi *performance* siswa menggunakan algoritma *Random Forest*, yang bertujuan untuk memprediksi potensi kelulusan berdasarkan data akademik dan faktor pendukung lainnya. Proses analisis dimulai dengan tahapan *preprocessing* data, pembagian data menjadi subset pelatihan model dan pengujian, hingga evaluasi performa model menggunakan metrik akurasi, *precision, recall, and f1-score*.

3.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah sebanyak 495 data dengan jumlah data gender untuk perempuan 187 dan untuk laki-laki 208, Berikut hasil contoh dataset yang digunakan

Tabel 1. Contoh Dataset

SE	FA	PS	ME	FE	TT	ST	FAI	SC	HI	IN	RO	FE	HE	AB	G1	G2	G3
F	GT3	A	4	4	2	2	0	yes	yes	no	no	3	3	6	5	6	6
F	GT3	T	1	1	1	2	0	no	yes	yes	no	3	3	4	5	5	6
F	LE3	T	1	1	1	2	3	yes	yes	yes	no	3	3	10	7	8	10

<http://sistemasi.ftik.unisi.ac.id>

SE	FA	PS	ME	FE	TT	ST	FAI	SC	HI	IN	RO	FE	HE	AB	G1	G2	G3
F	GT3	T	4	2	1	3	0	no	yes	yes	yes	2	5	2	15	14	15
F	GT3	T	3	3	1	2	0	no	yes	no	no	3	5	4	6	10	10

Berikut adalah penjelasan pada tabel 1:

- (SE) Sex : Jenis kelamin siswa: F untuk perempuan, M untuk laki-laki.
- (FA) famsize: Ukuran keluarga: LE3 untuk keluarga kecil (3 anggota atau kurang), GT3 untuk keluarga besar (lebih dari 3 anggota).
- (PS) Pstatus: Status tempat tinggal orang tua: T untuk tinggal bersama, A untuk tinggal terpisah (misalnya, karena perceraian atau pekerjaan).
- (ME)Medu: Tingkat pendidikan ibu: Skala 0-4 (0: tidak berpendidikan, 4: pendidikan tinggi seperti universitas).
- (FE) Fedu: Tingkat pendidikan ayah: Skala 0-4 (0: tidak berpendidikan, 4: pendidikan tinggi seperti universitas).
- (TT) *traveltime*: Waktu perjalanan ke sekolah: Skala 1-4 (1: kurang dari 15 menit, 4: lebih dari 1 jam).
- (ST) *studytime*: Waktu belajar mingguan di luar sekolah: Skala 1-4 (1: kurang dari 2 jam, 4: lebih dari 10 jam).
- (FAI) *failures*: Jumlah kegagalan akademik sebelumnya: Total jumlah mata pelajaran yang gagal (misalnya, 0 untuk tidak ada kegagalan).
- (SC) *schoolsup*: Apakah siswa menerima dukungan tambahan dari sekolah: yes untuk menerima, no untuk tidak.
- (HI) *higher*: Apakah siswa berencana melanjutkan pendidikan ke jenjang lebih tinggi (universitas): yes untuk iya, no untuk tidak.
- (IN) *internet*: Apakah siswa memiliki akses internet di rumah: yes untuk ya, no untuk tidak.
- (RO) *romantic*: Apakah siswa memiliki hubungan romantis: yes untuk ya, no untuk tidak.
- (FE) *freetime*: Waktu luang siswa setelah sekolah: Skala 1-5 (1: sangat sedikit waktu luang, 5: banyak waktu luang).
- (HE) *health*: Status kesehatan siswa: Skala 1-5 (1: sangat buruk, 5: sangat baik).
- (AB) *absences*: Jumlah ketidakhadiran siswa di sekolah.
- G1: Nilai akademik pada evaluasi periode pertama (G1) dalam skala 0-20.
- G2: Nilai akademik pada evaluasi periode kedua (G2) dalam skala 0-20.
- G3: Nilai akhir (G3) pada akhir periode akademik dalam skala 0-20 (biasanya digunakan untuk evaluasi performa siswa secara keseluruhan).

3.2. Pre-processing

Pre-processing pada penelitian ini melalui beberapa tahapan sebagai berikut:

- Delete*
Menghapus kolom yang tidak relevan atau tidak diperlukan untuk klasifikasi.
- Encode Categorical Variables*
Mengubah kolom yang berisi data kategorikal menjadi data numerik menggunakan *LabelEncoder*.
- Split Data into Features and Target*
Memisahkan data menjadi fitur (*input*) dan target (*output*) untuk model.
- Split Data into Train and Test Sets*
Membagi data menjadi data pelatihan dan pengujian untuk memastikan model dapat diuji dengan data baru.

3.3. Algoritma Random Forest

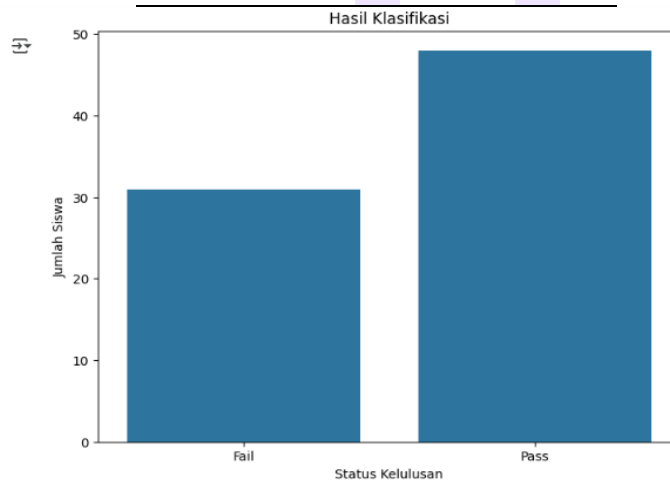
Pada tabel 2 data tersebut merupakan hasil evaluasi model *Random Forest* dalam memprediksi status kelulusan siswa berdasarkan data akademik. Model prediksi yang digunakan menunjukkan hasil yang sangat baik, di mana seluruh data mahasiswa yang diuji berhasil diprediksi sesuai dengan label aktualnya. Mahasiswa dengan Student ID 225 dan 307 yang memiliki status aktual *Fail*, diprediksi dengan benar sebagai *Fail*, sehingga model

<http://sistemasi.ftik.unisi.ac.id>

mampu mengenali kasus gagal dengan tepat. Sementara itu, mahasiswa dengan Student ID lainnya (305, 237, 318, 182, 94, 34) yang memiliki status aktual *Pass*, juga diprediksi dengan benar sebagai *Pass*.

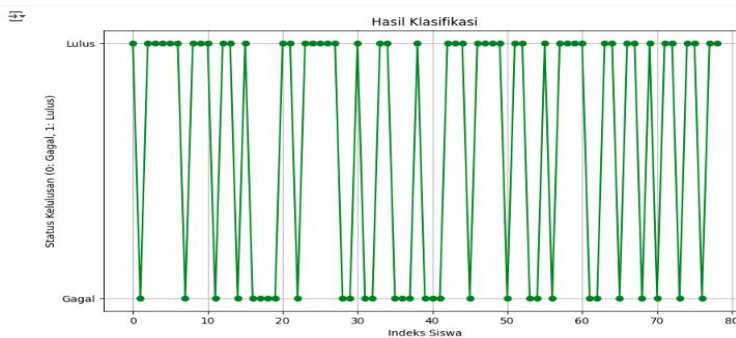
Tabel 2. Hasil Klasifikasi

Student ID	Actual	Predicted
305	<i>Pass</i>	<i>Pass</i>
225	<i>Fail</i>	<i>Fail</i>
237	<i>Pass</i>	<i>Pass</i>
318	<i>Pass</i>	<i>Pass</i>
182	<i>Pass</i>	<i>Pass</i>
94	<i>Pass</i>	<i>Pass</i>
34	<i>Pass</i>	<i>Pass</i>
307	<i>Fail</i>	<i>Fail</i>



Gambar 2. Distribusi Klasifikasi

Gambar 2 di atas merupakan diagram batang yang menggambarkan distribusi status kelulusan siswa berdasarkan hasil klasifikasi. Perbandingan jumlah siswa berdasarkan hasil klasifikasi yang dilakukan. Terdapat dua kelompok, yaitu kelompok yang diprediksi *Fail* dan *Pass*. Dari grafik, dapat dilihat bahwa jumlah siswa yang diklasifikasikan sebagai *Pass* lebih banyak, yaitu mendekati 50 siswa, sedangkan jumlah siswa yang diklasifikasikan sebagai *Fail* berada di angka sekitar 30 siswa. Ini menunjukkan bahwa model prediksi memberikan hasil bahwa mayoritas siswa diperkirakan akan lulus (*Pass*), sementara sebagian lainnya diperkirakan tidak lulus (*Fail*). Pola ini menggambarkan bahwa dalam dataset yang digunakan, proporsi siswa yang diprediksi berhasil lebih dominan, yang bisa jadi mencerminkan kondisi data aktual atau memang kecenderungan model yang lebih optimis terhadap kemungkinan kelulusan siswa.



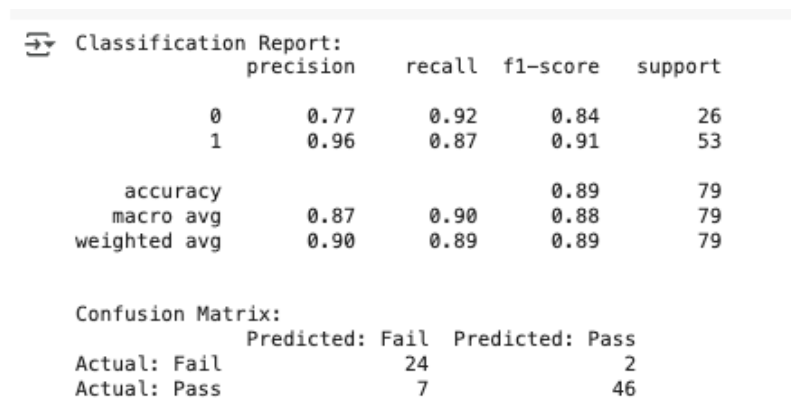
Gambar 3. Hasil Klasifikasi

Gambar 3 diatas merupakan diagram batang yang menggambarkan distribusi status kelulusan siswa berdasarkan hasil klasifikasi. Visualisasi ini menunjukkan hasil klasifikasi status kelulusan siswa berdasarkan indeks siswa. Sumbu Y menunjukkan dua kategori, yaitu Gagal (0) dan Lulus (1), sementara sumbu X menunjukkan urutan atau

<http://sistemasi.ftik.unisi.ac.id>

indeks siswa. Pola grafik menunjukkan bahwa prediksi model menghasilkan distribusi yang bervariasi di mana terdapat kombinasi antara siswa yang diprediksi Lulus dan Gagal di seluruh rentang indeks siswa. Meskipun sebagian besar siswa diklasifikasikan sebagai Lulus, terlihat adanya beberapa siswa yang diprediksi Gagal, yang tersebar di berbagai indeks. Ini menunjukkan bahwa model mampu membedakan antara siswa yang berpotensi lulus dan gagal, meskipun terdapat fluktuasi prediksi pada beberapa indeks yang menunjukkan adanya perbedaan karakteristik antar siswa yang mempengaruhi keputusan model. Pola ini mengindikasikan bahwa tidak terjadi prediksi satu arah saja, melainkan model secara dinamis memetakan kelulusan siswa sesuai data yang dimilikinya. Visualisasi ini juga membantu untuk mengidentifikasi siswa mana yang perlu diperhatikan lebih lanjut, terutama yang diprediksi Gagal, agar dapat diberikan intervensi lebih awal.

3.4. Confusion Matrix

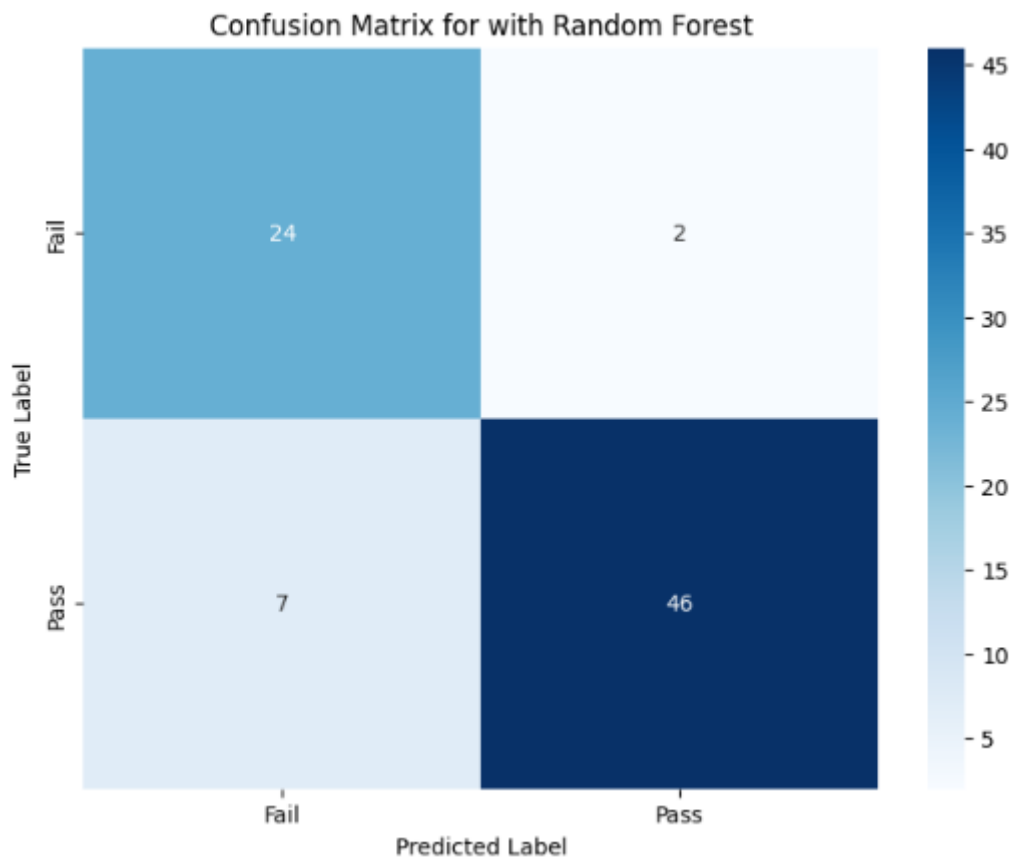


Gambar 4. Hasil Confusion Matrix

Pada gambar 4 menunjukkan performa model menunjukkan performa yang baik dengan menggunakan parameter $max_depth=15$, $random_state=42$ didapatkan tingkat akurasi keseluruhan sebesar 89%. Untuk kelas 0 (*Fail*), model memiliki *precision* sebesar 77%, artinya dari seluruh prediksi yang menyatakan *Fail*, 77% benar-benar *Fail*, dan *recall* sebesar 92%, yang berarti 92% dari data yang benar-benar *Fail* berhasil dikenali dengan benar. Ini menunjukkan model cukup baik dalam mendeteksi kelas *Fail*, meskipun masih ada kemungkinan *false positive*. Sedangkan untuk kelas 1 (*Pass*), *precision*-nya sangat tinggi yaitu 96%, menunjukkan bahwa hampir semua prediksi *Pass* benar-benar sesuai, sementara *recall*-nya sebesar 87%, artinya 87% dari data *Pass* yang sebenarnya berhasil terdeteksi dengan baik, dan sisanya 13% menjadi *false negative*. Dari *confusion matrix* diperoleh bahwa dari 26 data yang sebenarnya *Fail*, 24 berhasil diprediksi dengan benar sebagai *Fail*, dan hanya 2 salah diprediksi sebagai *Pass*. Sementara dari 53 data yang sebenarnya *Pass*, 46 berhasil diprediksi sebagai *Pass*, sedangkan 7 salah diklasifikasikan sebagai *Fail*. Secara keseluruhan, model menunjukkan kemampuan yang lebih kuat dalam mengenali kelas *Pass* dengan *precision* tinggi, namun *recall* kelas *Fail* lebih tinggi daripada kelas *Pass*, yang artinya model lebih cenderung hati-hati dan menghindari kesalahan pada kelas *Fail*. Nilai *macro average* dan *weighted average* yang mencapai 0.88 dan 0.89 juga memperlihatkan performa model yang konsisten dan seimbang dalam menangani kedua kelas. Namun demikian, masih terdapat ruang untuk meningkatkan *recall* pada kelas *Pass*, agar model mampu lebih baik lagi dalam mengurangi *false negative* yang bisa berdampak pada salah identifikasi kasus yang sebenarnya lolos.

35

2



Gambar 5. Diagram Hasil Confusion Matrix

Gambar 5 di atas menunjukkan hasil *confusion matrix*, model *Random Forest*, model menunjukkan kinerja yang baik dalam membedakan antara kelas *Fail* dan *Pass*. Dari 26 data aktual berlabel *Fail*, model berhasil memprediksi dengan benar 24 data sebagai *Fail*, dan hanya 2 data yang salah diprediksi sebagai *Pass*, yang menunjukkan *false negative* yang sangat rendah untuk kelas *Fail*. Sedangkan dari 53 data yang sebenarnya *Pass*, model berhasil memprediksi dengan benar 46 data, sementara 7 data lainnya salah diklasifikasikan sebagai *Fail*, yang menunjukkan adanya *false positive* pada kelas *Pass*. Performa ini mengindikasikan bahwa model *Random Forest* lebih akurat dalam mengidentifikasi data yang benar-benar *Fail*, dengan hanya sedikit kesalahan prediksi ke *Pass*, sementara pada data *Pass*, walaupun sebagian besar terdeteksi dengan benar, masih terdapat kesalahan prediksi ke *Fail* yang perlu diperhatikan. Pola ini menunjukkan bahwa model cenderung lebih konservatif atau hati-hati, dengan lebih banyak kesalahan prediksi yang mengategorikan *Pass* sebagai *Fail* dibandingkan sebaliknya. Hal ini bisa menjadi keuntungan jika dalam konteks bisnis lebih diutamakan untuk menghindari *false negative* pada kasus *Fail*, namun perlu diwaspadai potensi *false positive* yang bisa berdampak pada penilaian yang kurang akurat terhadap data yang sebenarnya *Pass*.

5 Kesimpulan (or Conclusion)

Model memiliki performa algoritma random forest untuk sangat baik menggunakan 2 parameter yaitu $max_depth=15$, $random_state=42$, mendapatkan nilai akurasi keseluruhan mencapai 89%. Model menunjukkan kemampuan yang sangat baik dalam mendeteksi kelas *Fail*, dengan *recall* mencapai 92%, yang berarti sebagian besar kasus *Fail* berhasil diidentifikasi dengan benar, meskipun *precision*-nya berada di angka 77%, menandakan masih terdapat beberapa prediksi *Fail* yang sebenarnya merupakan *Pass*. Untuk kelas *Pass*, *precision model* sangat tinggi yaitu 96%, menunjukkan bahwa hampir semua prediksi *Pass* benar-benar akurat, walaupun *recall*-nya sedikit lebih rendah di angka 87%, yang menunjukkan masih ada sebagian kecil data *Pass* yang salah diklasifikasikan sebagai *Fail*. Hal ini juga didukung oleh *confusion matrix*, di mana dari 26 data *Fail*, hanya 2 yang salah diprediksi sebagai *Pass*, sedangkan dari 53 data *Pass*, terdapat 7 yang salah diprediksi sebagai *Fail*. Secara keseluruhan, model *Random Forest* lebih dominan aman dalam mendeteksi kasus *Fail*, namun tetap mempertahankan *presisi* yang tinggi pada kasus *Pass*, menjadikannya cocok digunakan dalam situasi di mana kesalahan

<http://sistemasi.ftik.unisi.ac.id>

dalam mendeteksi kasus Fail lebih dihindari (lebih sensitif terhadap *False Negative* pada kelas *Fail*). Namun demikian, perlu perhatian khusus untuk mengurangi kesalahan prediksi pada kelas *Pass* agar meningkatkan sensitivitas model secara keseluruhan.

5. Referensi (Reference)

- [1] Supriyanto, "Strategi Membangun Budaya Akademik Mahasiswa," *Ilmu Pendidikan: Jurnal Kajian Teori dan Praktik Kependidikan*, vol. 6, no. 1, pp. 11–21, 2021.
- [2] M. H. bin Roslan and C. J. Chen, "Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021)," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 5, pp. 147–179, 2022, doi: 10.3991/ijet.v17i05.27685.
- [3] S. M. F. D. Syed Mustapha, "Predictive Analysis of Students' Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods," *Applied System Innovation*, vol. 6, no. 5, pp. 2–24, Oct. 2023, doi: 10.3390/asi6050086.
- [4] H. Andrianof, A. P. Gusman, and O. A. Putra, "Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia," *Jurnal Sains Informatika Terapan (JSIT) E-ISSN*, vol. 4, no. 1, pp. 24–28, 2025, Accessed: May 16, 2025. [Online]. Available: <https://rcf-indonesia.org/home/>
- [5] S. Sobari, A. I. Purnamasari, A. Bahtiar, and K. Kaslani, "MENINGKATKAN MODEL PREDIKSI KELULUSAN SANTRI TAHFIDZ DI PONDOK PESANTREN AL-KAUTSAR MENGGUNAKAN ALGORITMA RANDOM FOREST," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5704.
- [6] L. Breiman, "Random Forests," vol. 45, Kluwer Academic, 2001, pp. 5–32.
- [7] S. Kumar, F. Janan, and S. Kumar Ghosh, "Prediction of Student's Performance Using Random Forest Classifier," in *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management Singapore*, singapore, Mar. 2021, pp. 7089–7100. [Online]. Available: <https://www.researchgate.net/publication/354925634>
- [8] C. Ma, "Improving the Prediction of Student Performance by Integrating a Random Forest Classifier with Meta-Heuristic Optimization Algorithms," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 15, no. 6, pp. 1032–1044, 2024, [Online]. Available: www.ijacsa.thesai.org
- [9] F. Orji and J. Vassileva, "Using Machine Learning to Explore the Relation between Student Engagement and Student Performance," in *Proceedings of the International Conference on Information Visualisation*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 480–485. doi: 10.1109/IV51561.2020.00083.
- [10] Y. Chen and K. Jin, "Educational Performance Prediction with Random Forest and Innovative Optimizers: A Data Mining Approach," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, 2024, [Online]. Available: www.ijacsa.thesai.org
- [11] J. Kuswanto, H. Lukmanul, A. Info, and K. Kunci, "Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa," *Decode (Jurnal Pendidikan Teknologi Informasi)*, vol. 5, no. 1, pp. 262–270, 2025, doi: <http://dx.doi.org/10.51454/decode.v5i1.110311>.
- [12] Y. Priantama, T. Azhima, and Y. Siswa, "OPTIMASI CORRELATION-BASED FEATURE SELECTION UNTUK PERBAIKAN AKURASI RANDOM FOREST CLASSIFIER DALAM PREDIKSI PERFORMA AKADEMIK MAHASISWA," *Jurnal Informatika dan Komputer*, vol. 6, no. 2, pp. 251–260, 2022.
- [13] R. P. Munggaran, M. Nurmalasari, H. Hosizah, and D. Krismawati, "Prediksi Waktu Tunggu Pelayanan Pasien Rawat Jalan dengan Algoritma Random Forest," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 35–40, Nov. 2024, doi: 10.57152/malcom.v5i1.1529.
- [14] R. Herdiana, "PREDIKSI PENETAPAN TARIF PENERBANGAN MENGGUNAKAN AUTO-ML DENGAN ALGORITMA RANDOM FOREST," *Jurnal Ilmu Komputer Ruru*, vol. 2, no. 1, pp. 17–23, 2025, doi: 10.69688/jikr.v2i1.10.

- [15] R. Swarga Reza and M. Asyhari Yusuf, "Penerapan Algoritma Random Forest Untuk Klasifikasi Kualitas Air Berbasis Web," *Jurnal Ilmu Komputer Dan Informatika*, vol. 1, no. 3, pp. 79–88, Jan. 2025, Accessed: May 16, 2025. [Online]. Available: <https://jurnal.globalscients.com/index.php/jiki>
- [16] A. Fauzi, N. Maulidah, R. Supriyadi, H. Nalatissifa, and S. Diantika, "Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest," *Journal of Artificial Intelligence and Digital Business (RIGGS)*, vol. 4, no. 1, pp. 43–49, 2025, doi: [10.31004/riggs.v4i1.367](https://doi.org/10.31004/riggs.v4i1.367).