

PENERAPAN ALGORITMA KLASIFIKASI UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA MENGGUNAKAN RAPPIDMINER

Yulifda Elin Yuspita¹, Riri Okra², Muhammad Rezeki³

1,2) Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi

3) Universitas Putra Indonesia "YPTK" Padang

Article Info	ABSTRACT
Article history: Received: 16 Maret, 2025 Revised: 29 Maret 2025 Accepted: 23 April 2025	Abstrak Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi tingkat kelulusan mahasiswa secara tepat waktu pada Program Studi PTIK di UIN Bukittinggi. Dalam melakukan prediksi tingkat kelulusan mahasiswa terdapat berbagai faktor yang mempengaruhi keberhasilan mahasiswa dalam menyelesaikan studi tepat waktu. Penelitian ini menerapkan metode klasifikasi guna memprediksi kelulusan mahasiswa dengan menggunakan alat bantu RapidMiner. Dataset yang digunakan adalah data mahasiswa, yang diambil dari sistem e-campus UIN Bukittinggi. Sebanyak 164 data mahasiswa angkatan 2019 digunakan sebagai sampel untuk membangun model prediksi. Data dianalisis dan diproses melalui tahapan integrasi, pembersihan data, penanganan missing values, serta normalisasi. Model klasifikasi Decision Tree dan Random Forest digunakan untuk memprediksi kelulusan mahasiswa. Hasil penelitian menunjukkan bahwa model dengan akurasi 92,92% mampu memprediksi status kelulusan mahasiswa secara efektif. Confusion Matrix menunjukkan model memiliki performa yang baik dalam memprediksi kelas "Lulus" dan "Tidak Lulus" dengan kesalahan minimum. Model ini diharapkan dapat digunakan sebagai alat bantu bagi pihak kampus dalam melakukan deteksi dini mahasiswa yang berpotensi tidak lulus tepat waktu sehingga dapat dilakukan intervensi yang sesuai. Kata Kunci: Klasifikasi, Decicion Tree, Random Forest Abstract <i>This study aims to identify factors that affect the timely graduation rate of students in the PTIK Study Program at UIN Bukittinggi. In predicting student graduation rates, there are various factors that influence the success of students in completing their studies on time. This research applies the classification method to predict student graduation using RapidMiner tools. The dataset used is student data, which is taken from the e-campus system of UIN Bukittinggi. A total of 164 2019 student data were used as samples to build a prediction model. The data is analyzed and processed through the stages of integration, data cleaning, handling missing values, and normalization. Decision Tree and Random Forest classification models are used to predict student graduation. The results showed that the model with an accuracy of 92.92% was able to predict student graduation status effectively. Confusion Matrix shows that the model has good performance in predicting the "Pass" and "Not Pass" classes with minimum error. This model is expected to be used as a tool for the campus in conducting early detection of students who have the potential not to graduate on time so that appropriate interventions can be made.</i> Keywords: Classification, Decision Tree, Random Forest Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercialL ShareAlike 4.0 (CC-BY-NC-SA).

Corresponding Author:

E-mail : yulifdaelinyuspita@uinbukittinggi.ac.id



1. PENDAHULUAN

Mahasiswa adalah elemen penting yang perlu diperhatikan dalam evaluasi program studi secara mendalam [1]. Salah satu indikator utama keberhasilan program studi adalah durasi yang dibutuhkan mahasiswa untuk menyelesaikan pendidikan mereka. Masa studi ini mencerminkan proses belajar mahasiswa, serta menunjukkan tingkat kinerja belajar mereka. Dari sudut pandang yang lebih luas, rata-rata durasi studi mahasiswa memberikan gambaran tentang kualitas program studi secara keseluruhan. Oleh sebab itu, lama waktu studi mahasiswa menjadi salah satu kriteria yang digunakan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) [2].

Menurut standar utama, kualitas Perguruan Tinggi, terutama Program Studi di Indonesia, dinilai melalui berbagai indikator mutu [3]. Salah satu poin penting dalam akreditasi perguruan tinggi adalah proporsi mahasiswa yang berhasil menyelesaikan studi mereka tepat waktu [4].

Perguruan Tinggi memerlukan proses akreditasi yang mempertimbangkan tingkat kelulusan mahasiswa. Ketepatan waktu kelulusan memainkan peran penting dalam penilaian akreditasi perguruan tinggi dan program studi [5]. Banyak institusi pendidikan tinggi dan program studi menekankan pentingnya faktor-faktor yang memengaruhi ketepatan waktu mahasiswa dalam menyelesaikan studi mereka [6]. Lama studi, atau durasi waktu yang dibutuhkan mahasiswa untuk menyelesaikan pendidikannya, menjadi salah satu aspek utama dalam evaluasi ini dalam melakukan penilaian akreditasi [7]. [8]

Dalam beberapa tahun terakhir, isu mengenai tingkat kelulusan mahasiswa telah menjadi perhatian yang signifikan di berbagai institusi pendidikan tinggi, khususnya terkait kelulusan mahasiswa tepat waktu [9]. Faktor-faktor yang memengaruhi kelulusan mahasiswa sangat beragam, termasuk kinerja akademik, variabel demografis, kondisi sosial ekonomi, faktor psikologis, dan struktur kurikulum. Beberapa program studi, termasuk yang ada di UIN Bukittinggi, mengalami peningkatan jumlah mahasiswa yang tidak menyelesaikan studi tepat waktu, bahkan sebagian di antaranya mengalami putus studi.

Untuk mengatasi tantangan ini, deteksi dini dan intervensi bagi mahasiswa yang berisiko gagal akademis menjadi sangat penting. Ilmu data menawarkan pendekatan yang menjanjikan untuk memprediksi tingkat kelulusan mahasiswa dengan menganalisis berbagai faktor yang berpengaruh. Penelitian ini memanfaatkan data yang bersumber dari sistem e-campus UIN Bukittinggi, mencakup catatan akademik, latar belakang keluarga, dan informasi demografis mahasiswa angkatan 2019.

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi menggunakan teknik klasifikasi, khususnya Decision Tree dan Random Forest, guna mengidentifikasi kemungkinan kelulusan mahasiswa. Klasifikasi merupakan salah satu metode dalam pembelajaran mesin (machine learning) yang bertujuan untuk

mengelompokkan data ke dalam kategori tertentu berdasarkan pola-pola yang ditemukan dalam data historis. Dalam konteks ini, algoritma klasifikasi digunakan untuk memprediksi apakah seorang mahasiswa akan lulus tepat waktu atau tidak, berdasarkan berbagai variabel seperti IPK, jumlah SKS per semester, latar belakang keluarga, dan data demografis lainnya.

Sebagai studi kasus, penelitian ini memanfaatkan data nyata dari sistem e-campus UIN Bukittinggi, khususnya mahasiswa angkatan 2019. Data yang digunakan mencakup catatan akademik mahasiswa, latar belakang sosial ekonomi, serta informasi demografis lainnya. Model prediksi yang dikembangkan kemudian diuji dan dievaluasi berdasarkan akurasi dan ketepatan klasifikasinya dalam memprediksi kelulusan mahasiswa. Dengan studi kasus ini, diharapkan hasil penelitian tidak hanya bersifat teoretis, tetapi juga memiliki kontribusi praktis bagi institusi dalam merumuskan strategi intervensi yang lebih efektif untuk meningkatkan ketepatan waktu kelulusan mahasiswa.

Algoritma Decision Tree

Adalah salah satu metode yang digunakan dalam klasifikasi, Metode ini mengungkapkan model klasifikasi dalam bentuk pohon, di mana data disajikan dan dianalisis melalui struktur node. Untuk mengkategorikan sekumpulan data tertentu, setiap atribut harus melewati node yang berbeda pada pohon keputusan. Ketika atribut tersebut mencapai node daun, data tersebut akan diklasifikasikan ke dalam kategori atau kelas yang sesuai [10]. Menurut [11] pada pohon keputusan ada 3 jenis node yang berbeda yaitu:

Root Node

Root Node adalah simpul yang berada di puncak pohon keputusan, atau akar dari seluruh struktur. Node ini tidak menerima input apapun, menunjukkan bahwa tidak ada data yang masuk ke simpul ini dari node lain [12].

Internal Node

Internal Node merupakan simpul percabangan yang menerima satu input dan memiliki satu atau lebih output [13]. Node ini berfungsi untuk mengarahkan data ke cabang berikutnya dalam pohon berdasarkan kondisi atau atribut tertentu.

Leaf Node

Leaf Node, atau node daun, adalah simpul terakhir dalam pohon keputusan. Node ini hanya memiliki satu input dan tidak memiliki output, berfungsi untuk mengklasifikasikan data ke dalam kategori atau kelas yang telah ditentukan.

Algoritma Random Forest

Algoritma Random Forest adalah metode pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi [14]. Algoritma ini merupakan pengembangan dari pohon keputusan, yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting. Dalam Random Forest, setiap pohon keputusan dibangun dari subset data acak yang diambil dengan metode bootstrap, serta memilih subset acak dari fitur yang tersedia pada setiap split [15]. Dengan cara ini, Random Forest mampu menangkap keragaman dalam data dan menghasilkan model yang lebih robust [16]

Salah satu keunggulan Random Forest adalah kemampuannya untuk mengatasi masalah data yang besar dan kompleks dengan baik. Algoritma ini tidak hanya memberikan hasil klasifikasi yang akurat, tetapi juga dapat memberikan estimasi pentingnya fitur, yang membantu dalam memahami faktor-faktor yang paling berpengaruh dalam prediksi. Random Forest juga cenderung memiliki performa yang stabil dan tidak terlalu sensitif terhadap parameter yang diatur, sehingga menjadikannya pilihan yang populer dalam berbagai aplikasi, seperti analisis data kesehatan, pengenalan pola, dan prediksi pasar [17]

Cross-validation

Cross-validation adalah teknik evaluasi model yang digunakan dalam statistik dan pembelajaran mesin untuk menilai seberapa baik model dapat melakukan generalisasi terhadap data yang tidak terlihat sebelumnya [18]. Tujuannya adalah untuk menghindari masalah overfitting, di mana model bekerja dengan baik pada data pelatihan tetapi gagal saat dihadapkan pada data baru [19].

Menurut [20] ada beberapa cara Kerja *Cross Validation*

Cara Kerja Cross-Validation

Proses cross-validation umumnya melibatkan langkah-langkah berikut:

Pembagian Data: Data dibagi menjadi dua atau lebih subset (atau "fold"). Salah satu subset digunakan untuk melatih model, sedangkan subset lainnya digunakan untuk menguji model [21].

Pelatihan dan Pengujian: Model dilatih menggunakan subset pelatihan dan diuji pada subset pengujian. Proses ini diulang beberapa kali, dengan subset yang berbeda digunakan sebagai data pelatihan dan pengujian setiap kali.

Evaluasi Kinerja: Hasil dari setiap iterasi evaluasi dicatat, dan metrik kinerja seperti akurasi, presisi, recall, atau F1-score dihitung. Akhirnya, metrik kinerja ini dirata-ratakan untuk mendapatkan gambaran keseluruhan tentang bagaimana model berkinerja pada data yang tidak terlihat.

Cross-validation membantu memastikan bahwa model yang Anda buat tidak hanya akurat tetapi juga generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Teknik ini penting untuk menghindari overfitting dan memberikan pemahaman yang lebih baik tentang kinerja model di dunia nyata [22].

Rapid Miner

Rapid Miner adalah platform perangkat lunak yang digunakan untuk analisis data, pembelajaran mesin, dan pengolahan data [23]. Ini dirancang untuk memungkinkan pengguna, baik yang berpengalaman maupun pemula, untuk melakukan eksplorasi data, membangun model analitis, dan mengimplementasikan algoritma pembelajaran mesin dengan cara yang intuitif dan efisien. Rapid Miner menyediakan antarmuka grafis yang memudahkan pengguna untuk mengatur alur kerja analisis tanpa perlu menulis banyak kode [5]

Menurut [24] beberapa Fitur dalam rapidminer

Antarmuka Visual: RapidMiner menggunakan antarmuka berbasis drag-and-drop, sehingga pengguna dapat dengan mudah menyusun langkah-langkah analisis dengan menghubungkan operator yang berbeda.

Beragam Operator: Platform ini menyediakan banyak operator untuk berbagai tahap analisis data, termasuk pembersihan data, transformasi, pemodelan, dan evaluasi. Pengguna dapat mengakses algoritma pembelajaran mesin, statistik, dan teknik pengolahan teks.

Integrasi Data: RapidMiner dapat menghubungkan dan mengambil data dari berbagai sumber, seperti basis data, file CSV, dan layanan web. Ini memudahkan pengguna untuk mengintegrasikan data dari berbagai sumber untuk analisis lebih lanjut.

Kustomisasi dan Ekstensi: Pengguna dapat menyesuaikan operator dan membuat skrip menggunakan bahasa pemrograman seperti R dan Python. RapidMiner juga memiliki marketplace yang memungkinkan pengguna untuk menambahkan ekstensi dan plugin untuk meningkatkan fungsionalitas.

Kemampuan Kolaborasi: RapidMiner mendukung kolaborasi tim dengan menyediakan fitur berbagi proyek dan pengelolaan versi, sehingga beberapa pengguna dapat bekerja sama dalam proyek analisis yang sama.

Analisis Prediktif: Dengan alat dan algoritma yang kuat, RapidMiner memungkinkan pengguna untuk membangun model prediktif dan melakukan analisis yang mendalam untuk mendapatkan wawasan dari data.

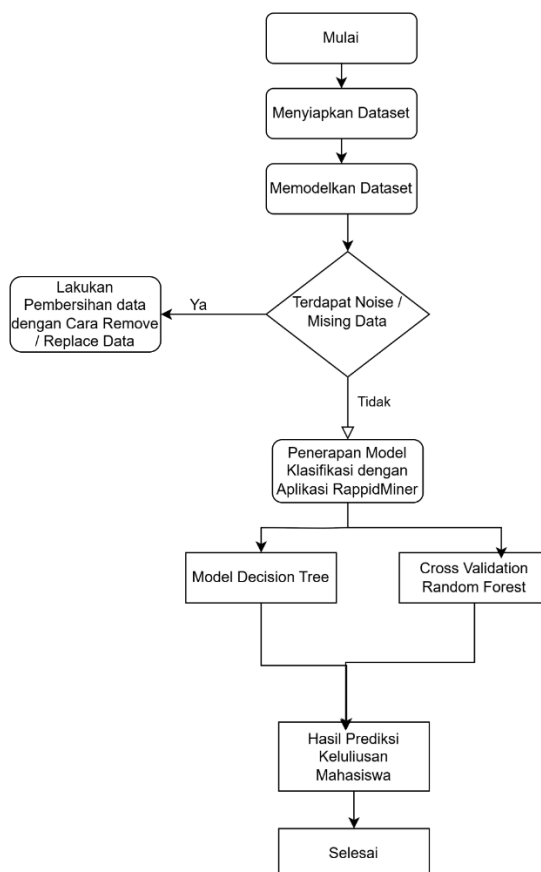
2. METODOLOGI PENELITIAN

Tahapan awal dalam penelitian ini adalah pengumpulan data yang diperoleh dari sistem e-Campus UIN Bukittinggi. Data tersebut digunakan dalam proses klasifikasi untuk memprediksi tingkat kelulusan mahasiswa. Desain penelitian ini mencakup beberapa tahapan utama, yaitu: pemilihan data, pra-pemrosesan data, pembangunan model klasifikasi, dan evaluasi model.

Penelitian ini menggunakan pendekatan pembelajaran mesin (machine learning) dengan menerapkan algoritma Decision Tree dan Random Forest untuk membangun model prediksi kelulusan mahasiswa tepat waktu. Kedua algoritma tersebut dipilih karena kemampuannya dalam menangani data dengan variabel kompleks serta menghasilkan model yang bersifat interpretatif. Untuk menghindari overfitting dan meningkatkan keandalan model, penelitian ini menggunakan teknik cross-validation (khususnya k-fold cross-validation) dalam proses pelatihan, yang secara acak membagi data ke dalam beberapa subset untuk memastikan bahwa model divalidasi pada bagian data yang berbeda secara berulang. Teknik ini sangat efektif untuk mengevaluasi stabilitas dan generalisasi model, khususnya dalam penerapan algoritma Random Forest yang juga memiliki mekanisme internal untuk mengurangi overfitting melalui proses pembentukan banyak pohon keputusan dan agregasi hasil prediksinya.

Dalam tahap evaluasi, penelitian ini menggunakan confusion matrix sebagai alat utama untuk mengukur kinerja model klasifikasi. Confusion matrix memungkinkan analisis terhadap jumlah prediksi yang benar dan salah, baik untuk kelas positif maupun negatif. Selain itu, metrik evaluasi seperti akurasi, presisi, recall, dan F1-score digunakan

untuk memberikan gambaran yang lebih menyeluruh tentang performa model. Interpretasi dari metrik-metrik ini memberikan wawasan penting mengenai kekuatan dan kelemahan model dalam mengklasifikasikan mahasiswa yang lulus tepat waktu dan yang tidak, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan strategis oleh pihak akademik. Agar lebih jelas desain penelitian yang digunakan dapat dilihat pada tahapan penelitian pada gambar sebagai berikut :



Gambar 1 Tahapan Penelitian

Pengumpulan Data

Populasi penelitian terdiri dari mahasiswa UIN Bukittinggi angkatan 2019, dengan sampel sebanyak 164 data mahasiswa yang dipilih secara purposif berdasarkan kelengkapan informasi yang relevan. Data yang digunakan mencakup informasi akademik seperti Indeks Prestasi Kumulatif (IPK) dan jalur masuk, data demografis seperti usia, jenis kelamin, dan tempat tinggal, serta data sosial ekonomi yang meliputi penghasilan dan pendidikan orang tua. Data ini diambil dari database e-campus, kemudian diintegrasikan menjadi satu dataset untuk keperluan analisis.

Contoh data yang digunakan dapat dilihat pada tabel berikut.

Nim	Nama	Kelamin	JenisSeleksi	Jenis_Tinggi	Ratapengl	Ip1	Ip2	Ip3	Ip4	Ip5	Ip6	Ip7	Ip8	Status
N001	DYMAS DETA PRASTYA	Laki-laki	4-SPAN-PTKIN	Asrama	5000000	0	3	3,6	0	3,057	3,015	3,278	2,864	Tidak Lulus
N002	NITA LIYANTIKA SARI	Perempuan	4-SPAN-PTKIN	Kost	1000000	2,85	2,85	3,045	2,85	3,739	3,425	3,365	3,045	Lulus
N003	FARIS ISWANDI	Laki-laki	4-SPAN-PTKIN	Asrama	2000000	3,6	3,6	3,545	3,6	3,092	3,299	2,339	3,545	Lulus
N004	DIYAN RIANTI	Perempuan	4-SPAN-PTKIN	Asrama	20000000	3,045	3,045	3,318	3,045	3,289	3,739	3,009	3,318	Lulus
N005	RIYAN AKBAR	Laki-laki	4-SPAN-PTKIN	Asrama	1000000	3,545	3,545	3,136	3,545	3,233	0	0	0	Tidak Lulus
N006	NURSALIMA	Perempuan	4-SPAN-PTKIN	Asrama	500000	3,318	3,318	3,318	3,318	2,548	2,94	3,494	3,318	Lulus
N007	SUCI FITRIANA	Perempuan	4-SPAN-PTKIN	Asrama	5000000	3,136	3,136	3,136	3,136	2,968	2,894	3,574	3,909	Lulus
N008	RINA SARI	Perempuan	4-SPAN-PTKIN	Asrama	500000	3,318	3,318	3,682	3,318	2,871	3,525	2,955	3,227	Lulus
N009	ROKHAYAH	Perempuan	4-SPAN-PTKIN	Asrama	0	3,136	3,136	4	3,136	2,991	3,327	3,074	3,227	Lulus
N010	FUDYA RIZKY ANGGARI	Laki-laki	4-SPAN-PTKIN	Asrama	500000	3,682	3,682	3,227	3,682	2,755	3,037	3,5	3,409	Lulus
N011	APRIAN TERMULO	Laki-laki	4-SPAN-PTKIN	Asrama	500000	4	4	2,955	4	3,179	3,289	3,227	3,227	Lulus
N012	M. ZAINI RANGKUTI	Laki-laki	4-SPAN-PTKIN	Asrama	1000000	3,227	3,227	3,364	0	0	0	0	0	Tidak Lulus
N013	SITI NURHALIZA JASER	Perempuan	4-SPAN-PTKIN	Asrama	2000000	2,955	2,955	2,864	2,955	3,425	3,428	3,587	3,5	Lulus
N014	NURUL AZIMAH	Perempuan	4-SPAN-PTKIN	Asrama	500000	3,364	3,364	2,864	3,364	3,31	2,984	2,214	2,5	Tidak Lulus
N015	RESTIKA DWI ALDA	Perempuan	4-SPAN-PTKIN	Asrama	500000	2,864	2,864	3,045	2,864	2,99	3,397	3,682	2,955	Lulus

Gambar 2 Tabel Data Mahasiswa**Alat dan bahan**

Alat dan bahan yang digunakan dalam penelitian ini antara lain adalah:

Ms. Excel digunakan untuk pengolahan data awal

RapidMiner merupakan tools yang digunakan untuk mengimplementasikan serta menggambarkan Model data yang akan di Gunakan [25].

Decision Tree adalah Algoritma yang digunakan dalam memodelkan data dalam bentuk pohon keputusan dengan cara mengelompokkan data berdasarkan klasifikasinya [26].

Random Forest adalah algoritma yang digunakan dalam memodelkan data untuk meningkatkan akurasi dan hasil perediksi [27].

Persiapan dan Pengolahan Data

Tahap persiapan dan pengolahan data yang mencakup proses integrasi, pembersihan, dan normalisasi. Data diperiksa untuk menghapus duplikasi serta menormalkan format yang berbeda agar konsisten, termasuk menangani missing values dengan cara menghapus atau mengisi data yang hilang sesuai dengan kriteria yang telah ditentukan. Selanjutnya, model klasifikasi dibangun dengan menggunakan algoritma *Decision Tree* dan *Random Forest*. Data dibagi menjadi dua bagian, yaitu 80% untuk pelatihan dan 20% untuk pengujian. Teknik cross-validation digunakan untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik dan tidak overfitting. Evaluasi model dilakukan dengan menghitung metrik performa seperti akurasi, presisi, recall, dan F1-score. Selain itu, Confusion Matrix digunakan untuk menganalisis prediksi model pada kelas “Lulus” dan “Tidak Lulus” guna memahami tingkat kesalahan prediksi. Dengan metode ini, penelitian diharapkan dapat memberikan hasil yang akurat dalam memprediksi kelulusan mahasiswa dan menjadi acuan dalam merumuskan kebijakan dan intervensi bagi mahasiswa yang berpotensi menghadapi kendala dalam menyelesaikan studi mereka.

Dataset Penelitian

Penelitian ini menggunakan data mahasiswa PTIK tahun Angkatan 2019 yang bersumber dari e-campus UIN Bukittinggi yang berjumlah 164 dataset. Atribut yang

digunakan yaitu : NIM, Nama, Jenis Kelamin, Tempat Tinggal, IPS1, IPS2, IPS3, IPS4, IPS5, IPS6, IPS7, IPS8, Jalur Masuk, Penghasilan Orang Tua, Pendidikan Orang Tua.

Setelah melakukan identifikasi dan perumusan masalah, penelitian melakukan pengumpulan dataset sebelum data ini diolah dengan menggunakan model *Decision Tree*, dengan menggunakan data Mahasiswa PTIK Angkatan 2019 untuk dijadikan landasan dalam pemberian label status kelulusan mahasiswa yang masih aktif saat ini.

Setelah dilakukan pra pemrosesan data, maka data set yang digunakan pada proses mining adalah 164 mahasiswa dengan 12 atribut yang dinormalisasikan dan *missing value* tidak terdapat pada dataset tersebut. Berikut adalah rincian atribut yang digunakan pada pemrosesan data.

NO	Nama Atribut	Jenis Data
1	Nim	Integer / id
2	Nama	polynomial
3	Kelamin	binominal
4	JenisSeleksi	binominal
5	Jenis_Tinggal_mhs	Integer
6	Ratapenghasilan	Integer
7	Ip4	real
8	Ip5	real
9	Ip6	real
10	Ip7	real
11	Ip8	real
12	Status	binominal / Label

Gambar 3 Atribut Data Setelah Pemrosesan Data

3. HASIL DAN PEMBAHASAN

Pengujian dengan Rapidminer

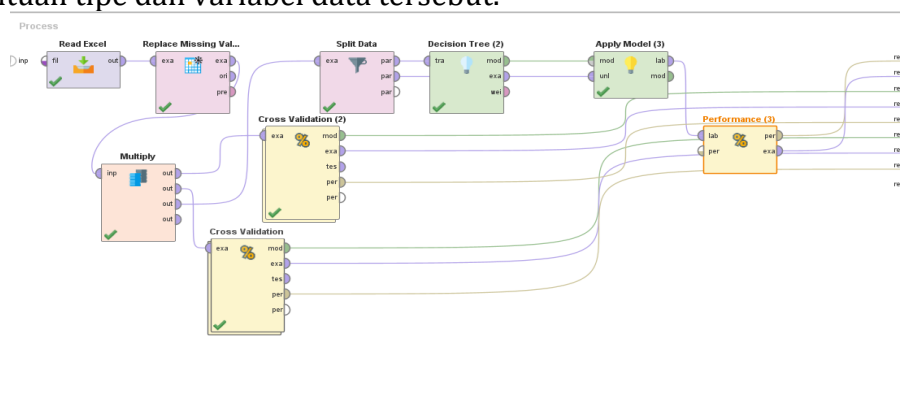
Pengujian dilakukan dengan menggunakan tools rapidminer untuk melihat akurasi dari algoritma *Decision Tree* dan Random Forest dalam melakukan klasifikasi data prediksi kelulusan mahasiswa.

Column	Type
nim	integer
nama	polynomial
kelamin	binomial
JenisSeleksi	binomial
Jenis_Tinggal_mhs	integer
RATAPEN...	integer
Ip4	real
Ip5	real
Ip6	real
Ip7	real
Ip8	real
Status	binomial / label

Gambar 4 Proses Data Preparation

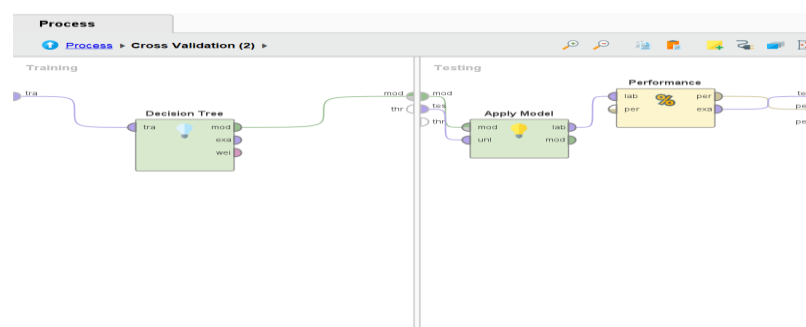
Import dataset ke RapidMiner dengan NIM sebagai Id gambar proses data preparation menunjukkan proses pengurutan data untuk persiapan analisis, seperti menghapus data

yang tidak relevan dan mengatasi missing value ataupun merubah format data sesuai dengan ketentuan tipe dan variabel data tersebut.



Gambar 5 Modeling Klasifikasi

Gambar diatas menggambarkan model klasifikasi, berikutnya akan digunakan *Cross Validation* membagi dataset ke dalam beberapa subset (folds). Model *Decision Tree* akan dilatih dan diuji di subset yang sama yang telah dibagi menggunakan *Split data* (80%, 20%). *Cross Validation* digunakan untuk memastikan agar model dapat digeneralisasi dengan baik



Gambar 6 Cross Validation Decision Tree

Dengan *Cross Validation*, kita dapatkan metrik performa yang lebih reliabel seperti akurasi, presisi, recall, dan F1-score [28]. Cross-validation pada model *Decision Tree* adalah metode yang digunakan untuk mengevaluasi dan memastikan keakuratan model dengan membagi data menjadi beberapa subset (lipatan), melatih model pada beberapa subset, dan menguji pada subset lainnya. Tujuannya adalah untuk meminimalkan risiko *overfitting* dan mendapatkan perkiraan performa model yang lebih andal [29].

Tree

```

Ip8 > 1.107
|
|   Ip8 > 2.267
|   |
|   |   Ip1 > 1.250
|   |   |
|   |   |   Ip7 > 2.267
|   |   |   |
|   |   |   |   Ip6 > 2.507
|   |   |   |   |
|   |   |   |   |   Ip1 > 2.675
|   |   |   |   |   |
|   |   |   |   |   |   Ip6 > 3.699
|   |   |   |   |   |   |
|   |   |   |   |   |   |   Ip2 > 3.196: Tidak Lulus (Tidak Lulus=1, Lulus=1)
|   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   Ip2 ≤ 3.196: Lulus (Tidak Lulus=0, Lulus=3)
|   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   Ip6 ≤ 3.699: Lulus (Tidak Lulus=4, Lulus=128)
|   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   Ip1 ≤ 2.675: Lulus (Tidak Lulus=1, Lulus=3)
|   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   Ip6 ≤ 2.507: Lulus (Tidak Lulus=1, Lulus=3)
|   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   Ip7 ≤ 2.267: Lulus (Tidak Lulus=1, Lulus=2)
|   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   Ip1 ≤ 1.250
|   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Ip2 > 3.257: Lulus (Tidak Lulus=0, Lulus=4)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Ip2 ≤ 3.257: Tidak Lulus (Tidak Lulus=2, Lulus=0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Ip8 ≤ 2.267: Tidak Lulus (Tidak Lulus=1, Lulus=1)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Ip8 ≤ 1.107: Tidak Lulus (Tidak Lulus=8, Lulus=0)

```

Gambar 7 Gambar Hasil Dari *Decision Tree*

Gambar diatas menunjukkan bagaimana *Decision Tree* bekerja dalam menentukan kelas atau hasil akhir berdasarkan serangkaian keputusan berjenjang.

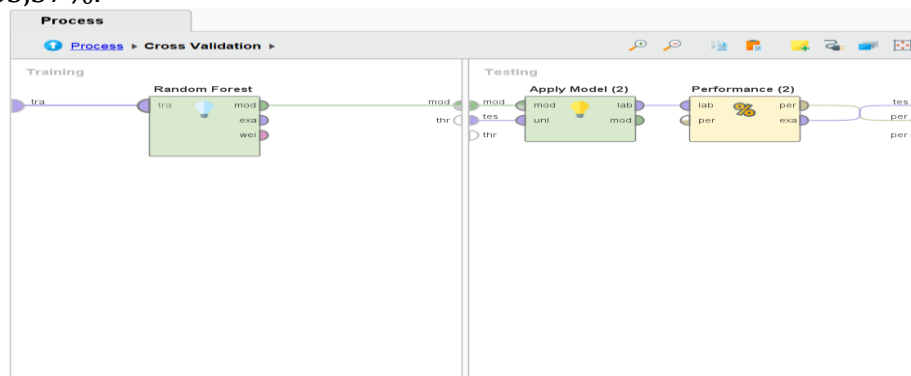
Table View Plot View

accuracy: 85.37% +/- 7.79% (micro average: 85.37%)

	true Tidak Lulus	true Lulus	class precision
pred. Tidak Lulus	8	13	38.10%
pred. Lulus	11	132	92.31%
class recall	42.11%	91.03%	

Gambar 8 Hasil Akurasi Menggunakan *Decision Tree*

Gambar ini menampilkan struktur model *Decision Tree* yang digunakan untuk melakukan klasifikasi data. Hasil dari model ini menunjukkan tingkat akurasi sebesar 85,37%, yang berarti model ini mampu memprediksi dengan benar 85,37% dari data uji. Berikut adalah penjelasan dari tiap elemen utama yang terlihat pada visualisasi diatas. Secara keseluruhan, Model ini memberikan gambaran yang komprehensif tentang cara kerja model *Decision Tree* dalam membagi data melalui serangkaian keputusan berjenjang, hingga mencapai hasil akhir klasifikasi yang akurat, dengan tingkat akurasi sebesar 85,37%.

**Gambar 9** Cross Validation Random Forest

Gambar diatas menunjukkan bagaimana algoritma Random Forest digunakan untuk mengklasifikasikan data sekaligus mengatasi masalah overfitting.

Random Forest merupakan algoritma pembelajaran mesin berbasis ansambel yang membangun banyak Decision Trees dan menggabungkan hasil dari masing-masing pohon untuk menghasilkan prediksi akhir [30]. Proses ini membantu meningkatkan akurasi klasifikasi sekaligus mengurangi risiko overfitting,

yaitu ketika model terlalu menyesuaikan diri dengan data pelatihan sehingga performanya menurun pada data baru [31].

Random Forest adalah algoritma yang andal untuk tugas klasifikasi karena dapat menggabungkan hasil dari banyak model *Decision Trees*, meningkatkan akurasi sekaligus mengatasi masalah *overfitting* yang umum terjadi pada model yang hanya mengandalkan satu pohon Keputusan [32].

☒ Table View ☐ Plot View

accuracy: 90.92% +/- 4.08% (micro average: 90.85%)

	true Tidak Lulus	true Lulus	class precision
pred. Tidak Lulus	8	4	66.67%
pred. Lulus	11	141	92.76%
class recall	42.11%	97.24%	

Gambar 10 Hasil Akurasi Menggunakan Random Forest

Gambar diatas menunjukkan struktur dan performa dari model *Random Forest* yang digunakan untuk klasifikasi data, dengan tingkat akurasi mencapai 90,92%. Ini berarti model dapat mengklasifikasikan data dengan benar pada 90,92% dari data uji, menunjukkan kinerja yang sangat baik. Berikut rincian yang menjelaskan elemen-elemen utama dari model *Random Forest*.

cara kerja model *Random Forest* dalam mengombinasikan hasil dari banyak *Decision Trees* untuk mencapai akurasi yang optimal, sekaligus menunjukkan ketahanan model terhadap *overfitting*. Model ini ideal digunakan pada tugas-tugas klasifikasi yang memerlukan akurasi tinggi, seperti yang terlihat dari tingkat akurasi sebesar 90,92%

4. SIMPULAN

Penelitian ini meneliti penerapan algoritma klasifikasi *Decision Tree* dan *Random Forest* untuk memprediksi tingkat kelulusan mahasiswa di Program Studi PTIK UIN Bukittinggi. Penelitian ini penting mengingat ketepatan waktu kelulusan mahasiswa menjadi salah satu indikator kinerja utama yang mempengaruhi akreditasi dan evaluasi mutu pendidikan. Sebagai upaya untuk memprediksi kelulusan, data yang digunakan diambil dari sistem e-campus, mencakup berbagai variabel yang relevan, termasuk faktor akademik, demografis, dan sosial ekonomi, yang dikumpulkan dari 164 mahasiswa angkatan 2019.

Keunggulan algoritma *Random Forest* dalam penelitian ini terutama ditandai dengan kemampuannya dalam mengurangi risiko overfitting melalui metode bootstrap sampling dan pemilihan subset acak dari fitur pada setiap split. Sebaliknya, *Decision Tree* juga terbukti efektif dalam pemetaan data dengan akurasi tinggi. Evaluasi performa yang

melibatkan metrik seperti presisi, recall, dan F1-score juga memberikan validasi tambahan atas keakuratan model.

Penelitian ini menghasilkan model klasifikasi yang akurat dalam memprediksi tingkat kelulusan mahasiswa. Model yang dikembangkan dengan algoritma *Decision Tree* dan Random Forest telah dievaluasi menggunakan data uji sebesar 20% dari total dataset, dan menunjukkan hasil yang memuaskan. Hasil pengujian menunjukkan bahwa model *Decision Tree* memiliki akurasi sebesar 85,37%, dengan Random Forest menunjukkan hasil 90,92%. Metrik performa tambahan, seperti presisi, recall, dan F1-score, juga dihitung untuk memberikan evaluasi yang lebih mendalam terhadap kemampuan model dalam mengklasifikasikan mahasiswa yang "Lulus" dan "Tidak Lulus."

Hasil menunjukkan bahwa kombinasi faktor-faktor akademik, demografis, dan sosial ekonomi memiliki pengaruh signifikan terhadap prediksi kelulusan mahasiswa. Model ini tidak hanya mendukung hasil penelitian sebelumnya tetapi juga menunjukkan bahwa data akademik, seperti IPK, serta faktor demografis dan sosial ekonomi, adalah variabel yang relevan dalam memprediksi keberhasilan akademik.

Model Random Forest yang digunakan dalam penelitian ini, selain menunjukkan akurasi yang tinggi, juga memiliki keunggulan dalam mengurangi risiko overfitting, yang sering menjadi masalah pada model klasifikasi yang kompleks. Penggunaan cross-validation dalam penelitian ini juga memastikan bahwa model dapat digeneralisasi dengan baik ke data baru. Ke depan, penelitian dapat mengembangkan model ini dengan mempertimbangkan lebih banyak variabel, seperti faktor psikologis atau perilaku belajar mahasiswa, untuk meningkatkan akurasi prediksi. Hasil dari penelitian ini diharapkan dapat membantu institusi dalam merancang intervensi yang tepat untuk meningkatkan tingkat kelulusan mahasiswa tepat waktu.

REFERENCES

- [1] U. Al Faruq, M. Ainun Naja Fauzi, I. Fatayasya, E. Daniati, A. Ristyawan, and N. PGRI Kediri, "Prediksi Data Kelulusan Mahasiswa Dengan Metode Decision Tree menggunakan Rapidminer," *Agustus*, vol. 7, pp. 2549–7952, 2023.
- [2] E. Etriyanti, D. Syamsuar, and N. Kunang, "Implementasi data mining menggunakan algoritme naive bayes classifier dan c4. 5 untuk memprediksi kelulusan mahasiswa," *Telematika*, vol. 13, no. 1, pp. 56–67, 2020.
- [3] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, pp. 35–43, 2020.
- [4] C. O. F. N. BAYES, "Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, 2020.
- [5] S. Sumarlin and D. Anggraini, "Implementasi K-nearest neighbord pada rapidminer untuk prediksi kelulusan mahasiswa," *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.*, vol. 10, no. 1, pp. 35–41, 2019.
- [6] U. I. N. Ar-raniry, F. Tarbiyah, D. A. N. Keguruan, P. Studi, and P. Teknologi, "PENGUNAAN METODE NAIVES BAYES UNTUK SKRIPSI Diajukan Oleh," 2023.
- [7] E. F. Wati and B. Rudianto, "Penerapan Algoritma KNN, Naive Bayes Dan C4. 5 Dalam Memprediksi Kelulusan Mahasiswa," *J. Format*, vol. 11, no. 2, pp. 168–175, 2022.
- [8] U. A. Wiza, Y. E. Yuspita, and W. D. Rahayu, "Decision-Making System for KIP IAIN Bukittinggi Scholarship Recipients Using the SAW and TOPSIS Methods."
- [9] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan algoritma decision tree dalam klasifikasi data prediksi kelulusan mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023.
- [10] E. Novianto, A. Hermawan, and D. Avianto, "Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 2, pp. 146–154, 2023.
- [11] I. Iskandar, L. Hiryanto, and J. Hendryli, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma

-
- Decision Tree C4. 5 Dengan Teknik Pruning,” *J. Ilmu Komput. Dan Sist. Inf.*, vol. 6, no. 1, pp. 64–68, 2018.
- [12] E. Ismanto and M. Novalia, “Komparasi Kinerja Algoritma C4. 5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas,” *Techno. Com*, vol. 20, no. 3, 2021.
- [13] A. Andri, “Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang,” *Implementasi Tek. Data Min. Untuk Memprediksi Tingkat Kelulusan Mhs. Pada Univ. Bina Darma Palembang*, 2022.
- [14] Y. Yunus and G. W. Nurcahyo, “Perbandingan algoritma c4. 5 dan naive bayes dalam prediksi kelulusan mahasiswa,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 1, pp. 193–199, 2023.
- [15] S. Defiyanti, “Analisis dan Prediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining,” *Syntax J. Inform.*, vol. 2, no. 1, 2013.
- [16] I. Irawan, M. R. Qisthiano, M. Syahril, and P. M. Jakak, “Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO,” *J. Pengemb. Sist. Inf. Dan Inform.*, vol. 4, no. 4, pp. 26–36, 2023.
- [17] A. Darmawan, I. Yudhisari, A. Anwari, and M. Makruf, “Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest,” *J. Minfo Polgan*, vol. 12, no. 1, pp. 387–400, 2023.
- [18] A. Trimanto, F. Faqih, I. M. Irfani, and S. Timur, “Penerapan Data Mining Untuk Evaluasi Status Kelulusan Mahasiswa Fakultas Teknologi Pertanian Tahun 2015 Menggunakan Algoritma Naïve Bayes Classifier,” 2015.
- [19] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
- [20] M. Kamil and W. Cholil, “Analisis Perbandingan Algoritma C4. 5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang,” *J. Inform.*, vol. 7, no. 2, pp. 97–106, 2020.
- [21] A. Rohman and A. Rufiyanto, “Implementasi data mining dengan algoritma decision tree c4. 5 untuk prediksi kelulusan mahasiswa di universitas pandanaran,” 2019.
- [22] F. Sugandi, “Prediksi Kelulusan Mahasiswa Stmik Dharmawacana Menggunakan Algoritma K-Nearest Neighbors,” *Sienna*, vol. 4, no. 1, pp. 20–26, 2023, doi: 10.47637/sienna.v4i1.704.
- [23] U. Suriani, “Penerapan data mining untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma decision tree C4. 5,” *J. Comput. Inf. Syst. Ampera*, vol. 4, no. 2, pp. 55–65, 2023.
- [24] U. Al Faruq, M. A. N. Fauzi, I. Fatayasya, E. Daniati, and A. Ristyawan, “Prediksi Data Kelulusan Mahasiswa Dengan Metode Decision Tree menggunakan Rapidminer,” in *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2024, vol. 8, no. 1, pp. 131–138.
- [25] Z. Nurfadilla, “Implementasi Data Mining untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Random Forest,” *AGENTS J. Artif. Intell. Data Sci.*, vol. 2, no. 2, pp. 35–42, 2022.
- [26] E. S. Susanto, K. Kusri, and H. Al Fatta, “Prediksi Kelulusan Mahasiswa Magister Teknik Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor,” *Respati*, vol. 13, no. 2, 2018.
- [27] I. M. B. Adnyana, “PENERAPAN TEKNIK KLASIFIKASI UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN NILAI AKADEMIK,” *J. Teknol. Inf. Dan Komput.*, vol. 7, no. 3, 2021.
- [28] W. A. Firmansyach, U. Hayati, and Y. A. Wijaya, “Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 262–269, 2023.
- [29] A. Saifudin, “Metode data mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di Universitas Pamulang,” *J. Teknol.*, vol. 10, no. 1, pp. 25–36, 2018.
- [30] N. N. Amani, M. Martanto, and U. Hayati, “PENGUNAAN ALGORITMA DECISION TREE UNTUK PREDIKSI PRESTASI SISWA DI SEKOLAH DASAR NEGERI 3 BAYALANGU KIDUL,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 473–479, 2024.
- [31] B. Yusuf, M. Qalbi, B. Basrul, I. Dwitawati, M. Malahayati, and M. Ellyadi, “Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh,” *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 4, no. 1, pp. 50–58, 2020.
- [32] P. Oktaviani, I. Asror, and M. A. Bijaksana, “Analisis Implementasi Sistem OLAP dan Klasifikasi Ketepatan Waktu Lulus dan Undur Diri Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Random forest,” *eProceedings Eng.*, vol. 5, no. 2, 2018.