

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU DENGAN METODE *RANDOM FOREST* BERDASARKAN KLASIFIKASI ALGORITMA *K-MEANS*

M. Hafizh Sytar¹, Ermatita²
Universitas Sriwijaya^{1,2}
hafizhsytar@gmail.com¹

ABSTRAK

Tujuan penelitian ini adalah untuk mengetahui prediksi tingkat kelulusan tepat waktu di Pascasarjana Universitas Sriwijaya menggunakan metode *Random Forest* dan algoritma *K-Means*. Metode penelitian yang digunakan adalah metode penelitian klasifikasi dengan lima langkah-langkah dalam kerangka penelitian yaitu 1) Pengumpulan Data; 2) Pengolahan Awal Data; 3) Klasifikasi; 4) Pengujian Model; 5) Evaluasi dan Validasi Hasil. Hasil penelitiannya menunjukkan penggunaan metode data mining *K-Means* dan *Random Forest* dapat efektif untuk memprediksi kelulusan mahasiswa tepat waktu pada program pascasarjana. Model yang dihasilkan memiliki tingkat akurasi yang tinggi dengan kemampuan mengenali pola kelulusan mahasiswa yang baik. Kombinasi kedua metode ini memberikan kontribusi signifikan dalam pengembangan model prediksi yang akurat, sehingga dapat membantu mahasiswa dan institusi dalam merencanakan dan mengambil keputusan terkait kelulusan tepat waktu. Simpulan, kombinasi metode data mining *K-Means* dan *Random Forest* efektif untuk memprediksi kelulusan mahasiswa tepat waktu pada program pascasarjana. Model ini memiliki tingkat akurasi yang tinggi dan dapat membantu mahasiswa dan institusi dalam merencanakan dan mengambil keputusan terkait kelulusan tepat waktu. Penelitian ini dapat menjadi acuan bagi institusi pendidikan untuk meningkatkan kualitas pendidikan dan membantu mahasiswa mencapai tujuan akademik mereka.

Kata Kunci: Prediksi Kelulusan Mahasiswa, Tepat Waktu, Metode *Random Forest*

ABSTRACT

The purpose of this study was to predict the level of on-time permits at the Postgraduate Program of Sriwijaya University using the Random Forest method and the K-Means algorithm. The research method used is the classification research method with five steps in the research framework, namely 1) Data Collection; 2) Initial Data Processing; 3) Classification; 4) Testing Model; 5) Evaluation and Validation of Results. The results of the study indicate that the use of the K-Means and Random Forest data mining methods can be effective in predicting on-time student admissions to the fill-in program. The resulting model has a high level of accuracy with the ability to recognize good student accuracy patterns. The combination of these two methods makes a significant contribution to the development of accurate prediction models, so that it can help students and institutions in planning and making decisions related to on-time permits. In conclusion, the combination of the K-Means and Random Forest data mining methods is effective in predicting on-time student graduation in the sustainability

program. This model has a high level of accuracy and can help students and institutions in planning and making decisions related to on-time permits. This study can be a reference for educational institutions to improve the quality of education and help students achieve their academic goals.

Keywords: *Student Graduation Prediction, On Time, Random Forest Method*

PENDAHULUAN

Pendidikan yang berkualitas pada suatu negara berawal dari kualitas pendidikan masing-masing daerah. Kualitas pendidikan masing-masing daerah di negara Indonesia belum cukup merata dan belum maksimal. Kualitas pendidikan menjadi kurang bagus bisa terjadi karena terletak di daerah yang jaraknya jauh dari pusat kota (Andrianof et al., 2024). Pendidikan tinggi memainkan peran penting dalam pengembangan sumber daya manusia yang berkualitas (Armansyah & Ramli, 2022). Tingkat kelulusan mahasiswa tepat waktu menjadi salah satu indikator keberhasilan institusi pendidikan tinggi. Namun, pada kenyataannya, banyak mahasiswa yang mengalami keterlambatan dalam menyelesaikan studi mereka. Di Pascasarjana Universitas Sriwijaya, masalah kelulusan tepat waktu ini juga menjadi perhatian utama.

Faktor-faktor yang mempengaruhi kelulusan mahasiswa mencakup berbagai aspek seperti latar belakang akademik, kondisi sosial ekonomi, keterlibatan dalam kegiatan kampus dan dukungan finansial. Dukungan finansial seperti beasiswa, bantuan biaya pendidikan, dan kemampuan ekonomi keluarga, merupakan faktor penting yang dapat mempengaruhi kelulusan tepat waktu (Yuda et al., 2022). Mahasiswa yang menghadapi kesulitan finansial seringkali harus bekerja paruh waktu

untuk membantu dalam mendanai pendidikan mereka yang tentunya dapat mengurangi waktu dan energi yang didedikasikan dalam menyelesaikan studi (Ameliatus & Fatah, 2024). Oleh karena itu, diperlukan metode yang efektif untuk memprediksi kelulusan mahasiswa tepat waktu guna mengidentifikasi dan mengatasi masalah-masalah yang dapat menghambat proses pendidikan mereka.

Metode prediksi yang akurat dapat membantu pihak universitas dalam mengambil langkah-langkah preventif dan korektif untuk meningkatkan tingkat kelulusan tepat waktu (Sudarmadji et al., 2023). Dalam penelitian ini berfokus pada penggunaan metode *Random Forest* untuk memprediksi kelulusan mahasiswa tepat waktu. *Random Forest* dikenal sebagai salah satu algoritma machine learning yang memiliki akurasi tinggi dan kemampuan generalisasi yang baik. Algoritma ini bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat dan stabil (Hanif & Setiadi, 2020).

Algoritma *Random Forest*, berbasis pohon keputusan, telah diaplikasikan dalam konteks pendidikan untuk memprediksi keberhasilan akademik (Diantika et al., 2024). *Random Forest* sebagai teknik yang mampu mengatasi overfitting dan menghasilkan prediksi yang lebih akurat pada data yang kompleks (Asro

et al., 2025). *Random Forest* sebagai model yang efektif dalam prediksi kelulusan mahasiswa karena kemampuannya dalam memproses berbagai faktor akademik yang saling terkait (Arrohman et al., 2025). Keunggulan lain dari algoritma ini adalah kemampuan dalam menangani dataset yang tidak seimbang dalam mengidentifikasi variabel kritis melalui fitur *feature importance* yang memungkinkan pihak universitas membuat kebijakan berbasis data yang lebih efektif (Maftucha et al., 2025). Penggunaan *Random Forest*, tidak hanya memberi prediksi yang akurat tetapi juga membantu memahami faktor yang mempengaruhi kelulusan mahasiswa tepat waktu sehingga memungkinkan intervensi dini dan perencanaan strategis yang lebih baik oleh pihak universitas.

Selain itu, penelitian ini juga menggunakan algoritma *K-Means*. Algoritma *K-Means* menggunakan metode klusterisasi untuk mengelompokkan data dengan cara mengidentifikasi titik pusat kluster (*centroid*) yang paling dekat dengan data (Mardzuki et al., 2024). Algoritma clustering, seperti *K-Means*, telah digunakan untuk mengelompokkan data mahasiswa dengan tujuan menemukan pola dalam data pendidikan. *K-Means* efektif untuk mengelompokkan data dalam bidang pendidikan karena kemampuannya dalam mengidentifikasi pola tersembunyi. Penelitian oleh (Andrianof et al., 2024). menjelaskan bahwa clustering membantu institusi pendidikan mengelompokkan mahasiswa berdasarkan atribut tertentu, seperti tingkat IPK dan partisipasi di kelas, yang berpengaruh pada performa akademik mahasiswa.

Penelitian lebih lanjut oleh Ubaidilah, (2023) juga menunjukkan bahwa faktor-faktor akademik dan demografi memberikan kontribusi signifikan pada prediksi kelulusan mahasiswa. Penggabungan metode *Random Forest* dan *K-Means* ini memungkinkan analisis yang lebih mendalam dan pembuatan model prediksi yang lebih kuat sehingga dapat membantu pihak universitas dalam merancang program dan kebijakan yang lebih tepat sasaran. Penelitian (Budi Adnyana, 2023) algoritma *Random Forest* memiliki akurasi tertinggi dalam memprediksi lama studi mahasiswa, sehingga dapat menjadi pilihan yang tepat untuk digunakan dalam pengembangan sistem prediksi. Dengan demikian, dapat membantu meningkatkan kualitas pendidikan dan membantu mahasiswa mencapai tujuan akademik mereka.

Penggunaan algoritma machine learning seperti *Random Forest* dan *K-Means* dalam konteks ini merupakan langkah inovatif yang dapat membuka peluang untuk penelitian lebih lanjut dan pengembangan aplikasi yang lebih luas (Puspa et al., 2025). Hasil penelitian ini juga diharapkan dapat diterapkan di universitas lain yang menghadapi masalah serupa, sehingga memberikan dampak positif yang lebih luas.

Penelitian ini penting karena prediksi kelulusan mahasiswa tepat waktu merupakan salah satu aspek penting dalam penilaian kualitas perguruan tinggi. Dengan menggunakan metode prediksi yang lebih canggih, universitas dapat mengidentifikasi masalah-masalah yang menghambat kelulusan mahasiswa dan merancang strategi yang lebih baik untuk memastikan bahwa lebih banyak

mahasiswa dapat lulus tepat waktu. Penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan metode prediksi yang lebih canggih di bidang pendidikan tinggi dan dapat diterapkan di universitas lain yang menghadapi masalah serupa.

Tujuan dari penelitian ini adalah untuk memprediksi kelulusan mahasiswa tepat waktu menggunakan metode *Random Forest* dan klasifikasi algoritma *K-Means*, serta mengidentifikasi variabel-variabel yang paling berpengaruh terhadap kelulusan tersebut. Maksud dari penelitian ini adalah untuk memberikan kontribusi dalam pengembangan metode prediksi yang lebih canggih di bidang pendidikan tinggi dan membantu universitas dalam meningkatkan kualitas pendidikan dan layanan akademik. Dengan demikian, penelitian ini diharapkan dapat memberikan dampak positif yang lebih luas dan membantu meningkatkan kualitas pendidikan tinggi di Indonesia.

METODE PENELITIAN

Penelitian ini menggunakan metode penelitian klasifikasi dengan lima langkah yang sistematis, yaitu pengumpulan data sekunder dari dataset yang sudah ada, pengolahan awal data yang meliputi pembersihan, pengelompokan, dan pemecahan data menjadi data testing dan training, klasifikasi menggunakan algoritma *K-Means*, Naive Bayes, dan *Random Forest* untuk mengklasifikasikan data, pengujian model dengan menggunakan Rapidminer dan pengulangan pengujian untuk mendapatkan hasil terbaik, serta evaluasi dan validasi hasil klasifikasi dan pengujian model untuk menarik

kesimpulan yang akurat. Data sekunder yang digunakan dalam penelitian ini diperoleh dari bagian kemahasiswaan pascasarjana Universitas Sriwijaya, sehingga memungkinkan analisis yang lebih mendalam tentang faktor-faktor yang mempengaruhi kelulusan mahasiswa. Dengan menggunakan metode ini, penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa dan mengidentifikasi variabel-variabel yang paling berpengaruh, sehingga dapat memberikan kontribusi dalam pengembangan metode prediksi yang lebih canggih di bidang pendidikan tinggi.

HASIL PENELITIAN

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data akademik mahasiswa Universitas Sriwijaya berdasarkan ketepatan waktu lulus studi. Ketepatan waktu lulus studi dibagi menjadi 2 yaitu tepat dan terlambat. Mahasiswa dikatakan tepat waktu jika dapat menyelesaikan masa studi kurang dari 4 semester atau kurang dari 2 tahun, dan terlambat jika masa studi lebih dari 4 semester atau lebih dari 2 tahun.

Pengumpulan data untuk prediksi kelulusan mahasiswa pascasarjana di Universitas Sriwijaya memerlukan pendekatan yang sistematis. Data yang dikumpulkan harus mencakup faktor-faktor yang relevan dengan keberhasilan akademik dan kelulusan mahasiswa. Data tersebut meliputi atribut seperti umur, IPK, semester aktif, status pekerjaan, dan status kelulusan. Berikut adalah dataset mahasiswa beserta jenis data yang diperlukan:

Tabel 1.
Dataset mahasiswa

No	Nama	Umur	IPK	SKS	Semester	Pekerjaan	Kelulusan Tepat Waktu
1	Ali	23	3,4	40	5	Tidak	Ya
2	Budi	25	2,8	35	4	Ya	Tidak
3	Citra	22	3,5	42	5	Tidak	Ya
4	Dian	26	2,9	30	5	Ya	Tidak
5	Edo	24	3,2	38	5	Tidak	Ya
6	Fira	27	3,0	32	6	Ya	Tidak
7	Gina	23	3,6	44	4	Tidak	Ya
8	Hasan	28	2,7	34	5	Ya	Tidak
9	Intan	22	3,7	46	4	Tidak	Ya
10	Joko	26	2,6	30	6	Ya	Tidak

Pengolahan Awal Data

Implementasi algoritma *K-Means* pada data mahasiswa pascasarjana Universitas Sriwijaya dilakukan untuk mengelompokkan mahasiswa berdasarkan faktor-faktor yang memengaruhi kelulusan tepat waktu. Data yang digunakan mencakup atribut seperti umur, IPK, jumlah SKS, semester, status pekerjaan, dan status kelulusan tepat waktu. Sebelum pengelompokan, dilakukan prapemrosesan data untuk memastikan kualitas dan konsistensi informasi. Salah satu langkah prapemrosesan adalah memperbaiki nilai yang tidak valid, seperti nilai IPK pada data Gina yang menunjukkan angka tidak realistis (36). Selain itu, data dinormalisasi agar atribut dengan skala berbeda, seperti umur dan IPK, memiliki bobot yang seimbang dalam proses pengelompokan. Atribut kategori seperti status pekerjaan dan status kelulusan juga dikonversi menjadi nilai numerik untuk memungkinkan analisis algoritmik.

Klasifikasi

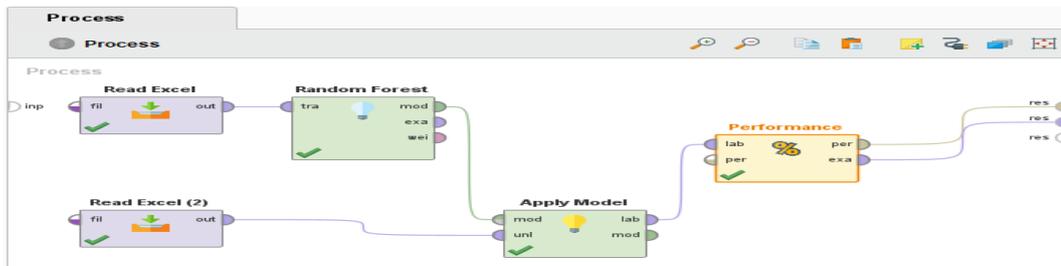
Algoritma *K-Means* digunakan untuk mengelompokkan mahasiswa menjadi dua cluster: yang cenderung

lulus tepat waktu dan yang berisiko tidak lulus tepat waktu. Hasil pengelompokan menunjukkan bahwa IPK, jumlah SKS, dan status pekerjaan adalah faktor dominan yang memengaruhi kelulusan tepat waktu. Universitas dapat meningkatkan tingkat kelulusan tepat waktu dengan memberikan perhatian khusus kepada mahasiswa berisiko dan memfasilitasi mahasiswa dengan potensi lulus tepat waktu. Implementasi *K-Means* dan *Random Forest* dapat membantu universitas dalam memprediksi kelulusan mahasiswa dan mengidentifikasi variabel yang berpengaruh, sehingga dapat membuat keputusan yang berbasis bukti untuk meningkatkan kinerja akademik mahasiswa.

Setelah evaluasi, tuning parameter dilakukan untuk mengoptimalkan model melalui metode seperti *grid search* atau *random search*. Model yang telah dioptimasi dapat digunakan untuk memprediksi status kelulusan mahasiswa baru dan menganalisis faktor yang paling memengaruhi kelulusan dengan melihat *feature importance*. Hasil implementasi ini tidak hanya menghasilkan prediksi yang akurat tetapi juga memberikan wawasan

tentang variabel yang berpengaruh, sehingga lembaga pendidikan dapat merancang intervensi yang tepat untuk meningkatkan tingkat kelulusan mahasiswa. *Random Forest* dipilih

karena kemampuannya yang robust dalam menangani data kompleks, risiko *overfitting* yang rendah, dan kemudahan interpretasi hasilnya.



Gambar 1.
Model *Random Forest*

Diagram pada Gambar 1 menunjukkan implementasi proses prediksi menggunakan algoritma *Random Forest* dengan alur kerja yang

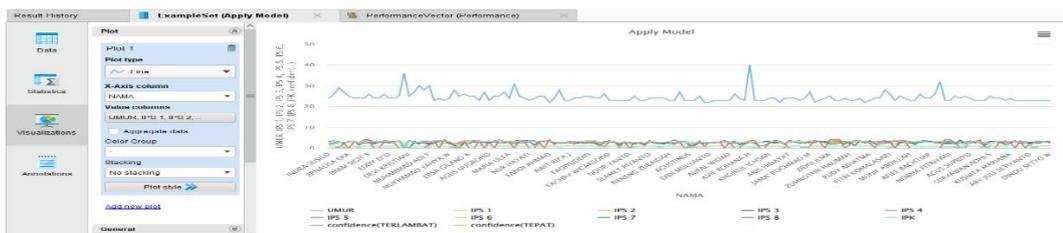
dirancang dalam suatu platform analisis data (seperti RapidMiner atau alat serupa).

Row No.	NAMA	STATUS KEL...	prediction%	confidencel...	JENIS KELA...	STATUS MA...	UMUR	STATUS NIK...	IPS 1	IPS 2	IPS 3	
1	INDRA SU...	TEPAT	TEPAT	0.052	0.948	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	3.170	2.700	3.230
2	LEVLA MEL...	TEPAT	TEPAT	0.063	0.937	PEREMPUNIAN	MAHASISWA	26	BELUM MENI...	3.000	3.000	3.420
3	AKBAR SO...	TEPAT	TEPAT	0.085	0.915	LAKI - LAKI	MAHASISWA	29	BELUM MENI...	2.670	2.660	2.930
4	ADHYA MA...	TEPAT	TEPAT	0.177	0.823	LAKI - LAKI	MAHASISWA	27	BELUM MENI...	2.480	2.980	2.090
5	ERNALISA ...	TEPAT	TEPAT	0.047	0.953	PEREMPUNIAN	MAHASISWA	25	BELUM MENI...	3.190	3.080	3.310
6	FARIDZ NU...	TEPAT	TEPAT	0.027	0.973	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	3.100	2.980	3.170
7	DARFIN HA...	TEPAT	TEPAT	0.179	0.821	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	2.980	2.980	2.230
8	WASKU FIT...	TEPAT	TEPAT	0.014	0.986	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	3.450	3.150	3.540
9	IMAM SIGIT S	TEPAT	TEPAT	0.048	0.952	LAKI - LAKI	MAHASISWA	26	BELUM MENI...	3.310	3.020	3.480
10	AHMAD SO...	TEPAT	TEPAT	0.035	0.965	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	2.620	2.730	2.110
11	HERI SISMI...	TEPAT	TEPAT	0.057	0.943	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	3.240	3.060	2.790
12	MULIYA AMB...	TEPAT	TEPAT	0.045	0.955	LAKI - LAKI	MAHASISWA	25	BELUM MENI...	2.830	3.020	3.020
13	EDDY EKO	TEPAT	TEPAT	0.030	0.970	LAKI - LAKI	MAHASISWA	24	BELUM MENI...	2.830	3	2.640

Gambar 2.
Hasil apply model dari metode *Random Forest*

Gambar 2 menunjukkan hasil penerapan metode *Random Forest* pada dataset. Proses ini melibatkan penerapan model yang sudah dilatih menggunakan data training untuk memprediksi status

kelulusan mahasiswa. Hasil yang diperoleh mencakup prediksi kelulusan berdasarkan variabel input yang telah ditentukan.



Gambar 3.

Plot apply model

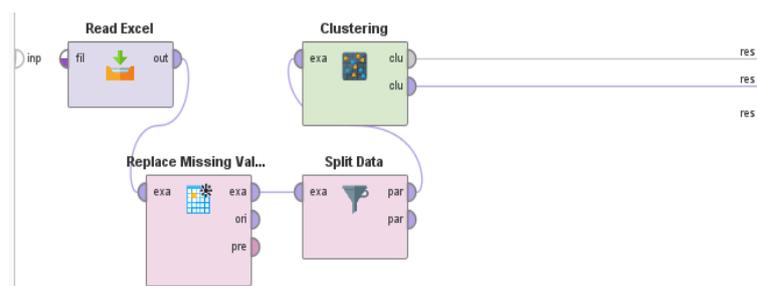
Gambar 3 menunjukkan visualisasi hasil penerapan model *Random Forest* dalam bentuk plot. Plot ini menggambarkan distribusi mahasiswa berdasarkan kategori kelulusan, antara mahasiswa tepat waktu dan terlambat. Visualisasi ini membantu melihat sebaran data secara lebih jelas.

Implementasi algoritma *K-Means* untuk prediksi kelulusan mahasiswa pascasarjana Universitas Sriwijaya dilakukan melalui beberapa tahap utama. Langkah pertama adalah pengumpulan data yang relevan, seperti data demografi mahasiswa (usia, jenis kelamin, status pernikahan), data akademik (nilai Indeks Prestasi Semester/IPS, Indeks Prestasi Kumulatif/IPK), tingkat kehadiran, dan riwayat pengambilan mata kuliah. Data ini diperoleh dari sistem informasi akademik universitas atau sumber terkait lainnya. Setelah data terkumpul, dilakukan prapemrosesan, termasuk penanganan data yang hilang, normalisasi nilai untuk menyamakan skala antar variabel, dan transformasi data kategori menjadi numerik.

Tahap berikutnya adalah proses penerapan algoritma *K-Means*. Algoritma ini bekerja dengan mengelompokkan mahasiswa berdasarkan kesamaan karakteristik dalam dataset yang telah diproses. Dalam implementasi ini, jumlah cluster

(k) ditentukan terlebih dahulu, misalnya untuk membagi mahasiswa ke dalam beberapa kelompok, seperti mahasiswa dengan kemungkinan tinggi untuk lulus tepat waktu, mahasiswa yang mungkin terlambat lulus, atau mahasiswa berisiko tidak lulus. Penentuan jumlah cluster dapat dilakukan melalui metode *elbow method* untuk menemukan nilai k yang optimal. Setelah jumlah cluster ditentukan, algoritma *K-Means* akan menginisialisasi titik pusat (*centroids*) secara acak dan kemudian memulai iterasi untuk mengelompokkan data berdasarkan jarak terpendek ke titik pusat tersebut. Proses ini berulang hingga mencapai konvergensi, yaitu ketika posisi titik pusat tidak berubah secara signifikan.

Implementasi *K-Means* membagi mahasiswa ke dalam cluster berdasarkan karakteristik mereka, membantu memahami pola dan faktor yang memengaruhi kelulusan. Cluster dengan IPK tinggi dan kehadiran baik cenderung memiliki tingkat kelulusan tinggi, sedangkan cluster dengan IPK rendah atau ketidakhadiran signifikan berisiko keterlambatan. Hasil ini dapat digunakan universitas untuk melakukan intervensi yang tepat. Penelitian ini juga menggunakan metode *K-Means* dengan perancangan proses pada *RapidMiner* sebagai berikut:



Gambar 4

Proses model *K-Means*

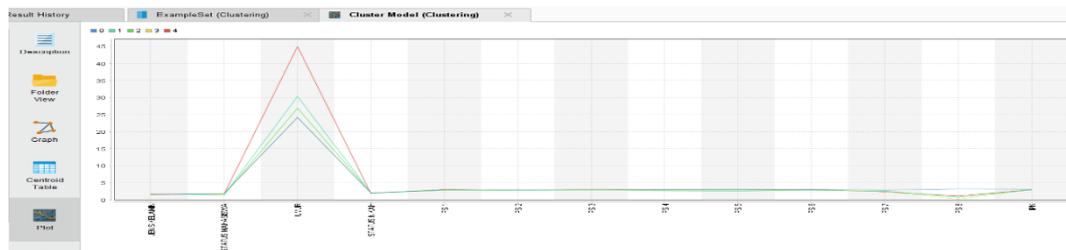
Diagram pada gambar menunjukkan proses analisis data dikelompokkan berdasarkan pola yang serupa. Dengan total 265 item, model ini berhasil membagi data ke dalam lima cluster dengan berbagai ukuran. Pada cluster 0 terdapat 64 items, cluster 1 36 items, cluster 2 71 items, cluster 3 88 items, dan cluster 4 6 items

Dalam model *K-Means* terdapat 5 cluster dimana data dalam dataset

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
JENIS KELAMIN	1.609	1.722	1.652	1.568	1.500
STATUS MAHASISWA	1.859	1.417	1.437	1.727	1.667
UMUR	24.156	30.333	26.887	24.148	45
STATUS NIKAH	1.984	1.844	1.985	1.989	1.833
IPS 1	2.905	2.691	2.807	2.977	3.047
IPS 2	2.946	2.739	2.771	2.820	2.765
IPS 3	2.911	2.726	2.913	3.052	3.038
IPS 4	3.043	2.879	2.740	2.839	2.938
IPS 5	3.016	2.617	2.671	2.885	2.947
IPS 6	3.049	2.751	2.854	2.849	3.063
IPS 7	2.816	2.486	2.548	2.347	2.282
IPS 8	3.147	0.741	0.851	0.218	1.157
IPK	3.080	2.827	2.844	2.927	2.962

Gambar 5

Centroid table



Gambar 6

Hasil plot cluster model

Gambar 5 menunjukkan centroid tabel dari model *K-Means*. Sedangkan Gambar 6 menunjukkan hasil plot cluster model tersebut.

Implementasi dengan menggunakan Python, pada penelitian ini dilakukan juga implementasi metode

dengan menggunakan google colab sebagai alat untuk perhitungan analisa. Berikut langkah yang dilakukan dalam pengimplementasian:

Melakukan import library yang dibutuhkan

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.cluster import KMeans
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification_report, accuracy_score
```

Gambar 7

Import library

Gambar 7 menunjukkan untuk pemrosesan data dan penerapan mengimpor library yang diperlukan algoritma machine learning, seperti

Pandas untuk mengelola data, NumPy untuk operasi matematika, Scikit-learn untuk algoritma *Random Forest* dan *K-Means*, serta LabelEncoder untuk mengubah data kategorikal. Selain itu, juga diimpor fungsi untuk membagi

dataset dan mengevaluasi kinerja model, seperti `train_test_split`, `classification_report`, dan `accuracy_score`. Load dataset

```
# Load the dataset
from google.colab import drive
drive.mount('/content/drive')
file_path = '/content/drive/MyDrive/data.xlsx' # Update the file path as needed
try:
    data = pd.read_excel(file_path, sheet_name='Sheet1')
    print("File loaded successfully.")
except FileNotFoundError:
    print("Error: File not found. Please check the file path.")
```

Gambar 8.

Load dataset

Gambar 8 menunjukkan kode yang digunakan untuk memuat dataset dari Google Drive ke dalam lingkungan Google Colab. Pertama, Google Drive di-mount menggunakan perintah `drive.mount('/content/drive')`, yang memungkinkan akses file dari Drive melalui jalur `/content/drive`. Kemudian, jalur file Excel yang akan dibaca

didefinisikan dalam variabel `file_path`, dan pengguna dapat memperbarui jalur tersebut sesuai dengan lokasi file yang diinginkan. Fungsi `pd.read_excel` dari pustaka Pandas digunakan untuk membaca file Excel dengan menentukan jalur file dan nama sheet, yaitu 'Sheet1'.

 Mounted at /content/drive
File loaded successfully.

Gambar 9

Hasil load dataset

Jika file berhasil dimuat, program akan mencetak pesan "File loaded successfully.". Namun, jika file tidak ditemukan pada jalur yang diberikan, blok exception akan menangkap kesalahan `FileNotFoundError` dan mencetak pesan "Error: File not found.

Please check the file path." untuk memberi tahu pengguna bahwa jalur file perlu diperiksa kembali. Kode ini memastikan bahwa file dapat dimuat dengan aman dan memberikan pesan kesalahan yang jelas jika terjadi masalah.

```
# Preprocessing
# Mengubah nama kolom agar lebih mudah diakses
data.columns = [
    "Nama", "Jenis_Kelamin", "Status_Mahasiswa", "Umur", "Status_Nikah",
    "IPS1", "IPS2", "IPS3", "IPS4", "IPS5", "IPS6", "IPS7", "IPS8", "IPK", "Status_Kelulusan"
]

# Menghapus kolom yang tidak diperlukan untuk prediksi
X = data.drop(columns=["Nama", "Status_Kelulusan"])
y = data["Status_Kelulusan"]

# Encode data kategorikal
encoder = LabelEncoder()
for col in ["Jenis_Kelamin", "Status_Mahasiswa", "Status_Nikah"]:
    X[col] = encoder.fit_transform(X[col])

# Mengganti nilai NaN dengan 0 (jika ada)
X = X.fillna(0)
```

Gambar 10
Preprocessing data

Kode tersebut melakukan tahap preprocessing data, yang merupakan langkah penting sebelum menerapkan algoritma machine learning. Kode melakukan preprocessing data yang meliputi mengubah nama kolom, memisahkan fitur dan target, mengubah data kategorikal menjadi numerik menggunakan LabelEncoder, dan menangani nilai kosong dengan mengisi nilai 0. Proses ini bertujuan untuk mempersiapkan data agar dapat diproses oleh algoritma machine learning dengan baik dan akurat. Clustering dengan menggunakan *K-Means*

```
# Clustering menggunakan KMeans
kmeans = KMeans(n_clusters=5, random_state=42)
data['Cluster'] = kmeans.fit_predict(X)

# Menampilkan hasil KMeans
print("Hasil Clustering KMeans:\n", data[['Nama', 'Cluster']])

# Menambahkan hasil clustering ke fitur
X['Cluster'] = data['Cluster']
```

Gambar 11.
Clustering dengan *K-Means*

Gambar 11 menunjukkan kode untuk proses clustering menggunakan algoritma KMeans pada dataset. Pertama, objek KMeans diinisialisasi dengan parameter `n_clusters=5` yang menentukan jumlah cluster yang diinginkan, dan `random_state=42` yang memastikan bahwa pembagian cluster akan konsisten setiap kali kode dijalankan. Selanjutnya, metode `fit_predict(X)` digunakan untuk melatih model KMeans pada data X, yang berisi fitur-fitur yang digunakan untuk clustering, serta memprediksi label cluster untuk setiap data. Hasil clustering ini kemudian disimpan dalam kolom baru pada DataFrame data dengan nama 'Cluster'. Setelah itu, hasil clustering ini ditampilkan dengan mencetak nama produk/data dan cluster yang sesuai.

```
Hasil Clustering KMeans:
  Nama Cluster
0 ANIK WIDAYANTI 2
1 DWI HESTYNA P 3
2 MURYA ARIEF B 3
3 NANIK SUSANTI 2
4 RIFKA ISTIQFARINA 3
.. ...
374 ARY JULI SETIYANTO 4
375 RINA ZAHROTUL U 4
376 TULISA WAHYUHADI K 4
377 NI'MATUL JANNAH 4
378 DINDU SETYO W 4

[379 rows x 2 columns]
```

Gambar 12.

Hasil cluster *K-Means*

Melakukan metode *Random Forest*

```
# Split data untuk training dan testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Model Random Forest
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Prediksi
predictions = model.predict(X_test)

# Evaluasi
print("Accuracy Random Forest:", accuracy_score(y_test, predictions))
print("Classification Report Random Forest:\n", classification_report(y_test, predictions))

# Menampilkan hasil prediksi Random Forest
X_test['Predicted_kelulusan'] = predictions
print("Hasil Prediksi Random Forest:\n", X_test)
```

Gambar 1.

Prediksi *Random Forest*

Pertama, kode ini membagi dataset menjadi dua bagian: data untuk pelatihan (training) dan data untuk pengujian (testing) dengan menggunakan fungsi `train_test_split`. Fungsi ini memisahkan variabel fitur (X) dan target (y), di mana 80% data digunakan untuk pelatihan dan 20% sisanya untuk pengujian. Parameter `random_state=42` memastikan bahwa pembagian data akan konsisten setiap kali kode dijalankan, sehingga hasilnya dapat dipertanggungjawabkan dan direproduksi. Variabel `X_train` dan `y_train` menyimpan data yang digunakan untuk melatih model, sementara `X_test` dan `y_test` digunakan untuk menguji model yang telah dilatih.

Selanjutnya, model klasifikasi *Random Forest* diinisialisasi dengan menggunakan `RandomForestClassifier(random_state=42)`. Parameter `random_state=42` bertujuan untuk memastikan bahwa inisialisasi model dilakukan dengan cara yang sama setiap kali kode dijalankan. Model ini kemudian dilatih dengan data pelatihan menggunakan metode `fit(X_train, y_train)`, yang memungkinkan model untuk

mempelajari hubungan antara fitur-fitur dalam `X_train` dan target dalam `y_train`.

Setelah model selesai dilatih, proses prediksi dilakukan dengan menggunakan metode `predict(X_test)`, yang menghasilkan prediksi berdasarkan data uji (`X_test`). Prediksi ini kemudian disimpan dalam variabel `predictions`. Hasil prediksi ini akan dibandingkan dengan label sebenarnya pada data uji (`y_test`) untuk mengevaluasi seberapa baik model bekerja. Untuk mengevaluasi kinerja model, kode ini mencetak dua metrik utama: akurasi dan laporan klasifikasi. Akurasi dihitung dengan membandingkan prediksi dengan label sebenarnya menggunakan fungsi `accuracy_score(y_test, predictions)`. Akurasi memberikan gambaran seberapa sering model berhasil memprediksi dengan benar. Selain itu, laporan klasifikasi yang lebih rinci dicetak dengan `classification_report(y_test, predictions)`, yang mencakup metrik lain seperti `precision`, `recall`, `f1-score`, dan `support` untuk setiap kelas, memberikan gambaran lebih dalam mengenai performa model dalam mengklasifikasikan setiap kategori.

Berdasarkan hasil ini, model memprediksi bahwa sebagian besar mahasiswa dalam dataset ini akan lulus tepat waktu (kategori "TEPAT"), meskipun ada beberapa yang diprediksi terlambat (kategori "TERLAMBAT"). Misalnya, mahasiswa dengan indeks prestasi yang lebih rendah atau lebih tidak stabil seperti pada baris ke-55 dan ke-182 diprediksi mengalami keterlambatan dalam kelulusan, sementara mahasiswa dengan IPK tinggi dan nilai IPS yang stabil lebih cenderung untuk lulus tepat waktu,

seperti yang terlihat pada mahasiswa pada baris ke-288, ke-283, dan ke-327.

Secara keseluruhan, hasil ini menunjukkan bahwa model *Random Forest* dapat memanfaatkan berbagai faktor seperti IPK, umur, jenis kelamin, dan performa akademis untuk memprediksi kelulusan tepat waktu atau terlambat bagi mahasiswa. Cluster yang ditambahkan juga membantu mengelompokkan mahasiswa dengan karakteristik serupa, yang mungkin terkait dengan pola kelulusan mereka. Menyimpan hasil clustering dan prediksi data ke drive

```
# Menyimpan model dan hasil clustering
x_test.to_excel('/content/drive/MyDrive/predicted_kelulusan.xlsx', index=False)
data[['Nama', 'Cluster']].to_excel('/content/drive/MyDrive/clustering_results.xlsx', index=False)

print("Hasil prediksi disimpan ke file predicted_kelulusan.xlsx")
print("Hasil clustering disimpan ke file clustering_results.xlsx")
```

Gambar 16.

Menyimpan hasil clustering dan prediksi

Gambar 17 menunjukkan bahwa hasil clustering dan prediksi data akan disimpan dalam bentuk file xlsx di driver.

A	B
Nama	Cluster
1	2
2	2
3	2
4	2
5	2
6	2
7	2
8	3
9	2
10	4
11	2
12	2
13	2
14	2
15	2
16	3
17	3

Gambar 17

Hasil clustering data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
Jenis Kelamin	Umur	Jatus_Niloi	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPS9	IPS10	IPS11	IPS12	Kted_Kelulusan
0	1	24	0	2.33	2.25	2.3	2.39	2.68	2.33	3.28	1.5	2.62	0	TEPAT	
1	1	24	0	2.95	2.84	2.2	2.77	3.16	2.53	2.95	2.9	3.07	3	TEPAT	
1	1	25	0	2.9	2.89	3.25	3.13	3.4	3.28	3.25	0	3.96	4	TEPAT	
1	1	25	0	3.33	2.48	3.73	3.52	3.44	3.47	2.72	0	3.95	4	TEPAT	
0	0	28	0	2.14	1.55	2.03	2.27	1.7	2.7	1.67	1.75	2.31	2	TERLAMBAT	
1	1	32	0	3.24	3.83	3.65	3.61	3.69	3.73	3.7	0	3.63	2	TEPAT	
0	1	22	0	2.86	2.86	3.25	3.25	3.31	3.48	2.43	3	3.28	3	TEPAT	
0	1	31	0	2.76	1.09	2.39	0.98	1.93	2.38	1.11	0	1.71	2	TERLAMBAT	
0	1	25	0	2.86	3.14	3.27	3.38	3.21	2.81	3.05	0	0.92	4	TEPAT	
0	0	24	0	2.76	2.19	2.77	0.96	1.33	1.11	0	0.36	1.87	0	TERLAMBAT	
1	1	23	0	3.02	2.88	2.45	2.14	3.24	3.41	3.2	4	2.81	3	TEPAT	
1	1	24	0	3.38	3.08	3.52	3.11	2.77	3.47	3.07	0	3.22	4	TEPAT	
0	1	27	0	3.12	3.1	2.73	2.8	3.07	3.31	2.77	2.77	3.07	2	TEPAT	
1	0	26	0	2.88	2.7	2.89	2.73	2.41	3.39	2.07	1.21	2.92	2	TERLAMBAT	
0	1	24	0	3.07	3.21	2.94	3.36	3.15	2.77	3.47	0	3.15	4	TEPAT	
0	1	23	0	2.5	2.08	2.25	2.27	2.21	2.16	2.86	2.89	2.88	3	TEPAT	
0	1	25	0	2.97	3.02	3.13	3.25	3.29	3.38	1.91	4	3.24	3	TEPAT	
0	1	25	0	3	2.83	3.57	3.26	3.83	3.53	2.43	0	3.37	4	TEPAT	
0	1	23	0	2.55	2.75	2.64	2.59	2.8	2.75	1.34	0	2.46	0	TEPAT	
0	1	26	0	2.88	2.71	3.45	3.15	2.88	0	3.02	0	3.02	4	TEPAT	
0	1	27	0	2.86	2.86	2.45	1.86	3.19	3.14	2.98	2	3.04	2	TEPAT	

Gambar 18.

Hasil prediksi data

Pengujian Model, Evaluasi dan Validasi Hasil

Visualisasi hasil pengelompokan dan prediksi menggunakan *K-Means* dan *Random Forest* menunjukkan bahwa variabel akademik seperti IPK dan SKS memiliki pengaruh besar

terhadap kelulusan tepat waktu. Visualisasi ini membantu pihak universitas memahami sebaran data dan pola yang ada, sehingga dapat mengambil langkah strategis untuk meningkatkan tingkat kelulusan mahasiswa.

accuracy: 100.00%			
	true TEPAT	true TERLAMBAT	class precision
pred. TEPAT	143	0	100.00%
pred. TERLAMBAT	0	2	100.00%
class recall	100.00%	100.00%	

Gambar 19
Hasil accuracy model *Random Forest*

Gambar 20 menampilkan nilai akurasi model *Random Forest*. Akurasi merupakan ukuran seberapa sering prediksi model benar dibandingkan dengan data aktual. Hasil Accuracy dari model dalam penelitian ini adalah

100%. Nilai akurasi yang tinggi menunjukkan bahwa model memiliki kinerja yang baik dalam memprediksi ketepatan waktu kelulusan mahasiswa. Berikut perhitungan manual untuk accuracy:

$$Accuracy = \frac{\text{Jumlah prediksi benar}}{\text{Total prediksi}} \times 100\% = \frac{143+2}{143+0+0+2} \times 100\% = 100\%$$

precision: 100.00% (positive class: TEPAT)			
	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	2	0	100.00%
pred. TEPAT	0	143	100.00%
class recall	100.00%	100.00%	

Gambar 20.
Hasil precision *Random Forest*

Gambar 22 menunjukkan nilai precision dari model *Random Forest*. Precision menunjukkan seberapa akurat model dalam memprediksi mahasiswa yang tepat waktu dibandingkan dengan prediksi positif yang dibuat oleh model. Hasil precision dalam model ini adalah

100%. Nilai precision yang tinggi menandakan bahwa model jarang menghasilkan prediksi salah positif. Berikut perhitungan manual untuk precision di setiap kelas dengan data yang diberikan ditunjukkan pada Tabel 6:

Tabel 6.
Data yang diberi

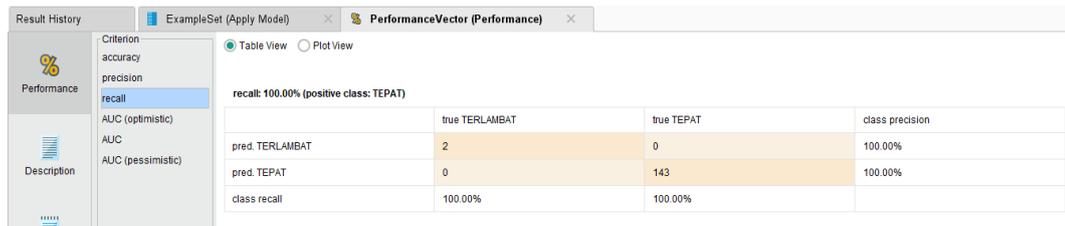
	True TEPAT	True TERLAMBAT
Pred TEPAT	143	0
Pred TERLAMBAT	0	2

Dari tabel tersebut diketahui bahwa:

1. True Positive (TP) = nilai diagonal di tabel (143 untuk TEPAT, 2 untk TERLAMBAT)
2. False Positive (FP) = nilai di baris prediksi kolom target lainnya
3. False Negative (FN) = nilai di kolom target baris prediksi lainnya

$$Precision_{TEPAT} = \frac{TP_{TEPAT}}{TP_{TEPAT} + FP_{TEPAT}} \times 100\% = \frac{143}{143 + 0} \times 100\% = 100\%$$

$$Precision_{TERLAMBAT} = \frac{TP_{TERLAMBAT}}{TP_{TERLAMBAT} + FP_{TERLAMBAT}} \times 100\% = \frac{2}{2+0} \times 100\% = 100\%$$



Gambar 21.
Hasil recall *Random Forest*

Gambar 23 menunjukkan hasil recall, yang mengukur sejauh mana model dapat menemukan semua mahasiswa yang tepat waktu dalam data aktual. Hasil recall model ini sebesar 100%.

Recall yang tinggi berarti model berhasil mengidentifikasi sebagian besar mahasiswa yang benar-benar lulus tepat waktu. Berikut perhitungan manual untuk recall:

$$Recall_{TEPAT} = \frac{TP_{TEPAT}}{TP_{TEPAT} + FN_{TEPAT}} \times 100\% = \frac{143}{143 + 0} \times 100\% = 100\%$$

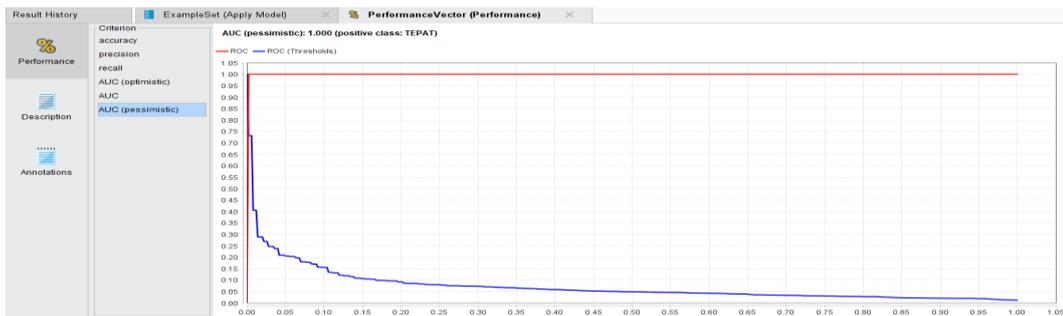
$$Recall_{TERLAMBAT} = \frac{TP_{TERLAMBAT}}{TP_{TERLAMBAT} + FN_{TERLAMBAT}} \times 100\% = \frac{2}{2+0} \times 100\% = 100\%$$



Gambar 22.
Hasil AUC (optimistic)



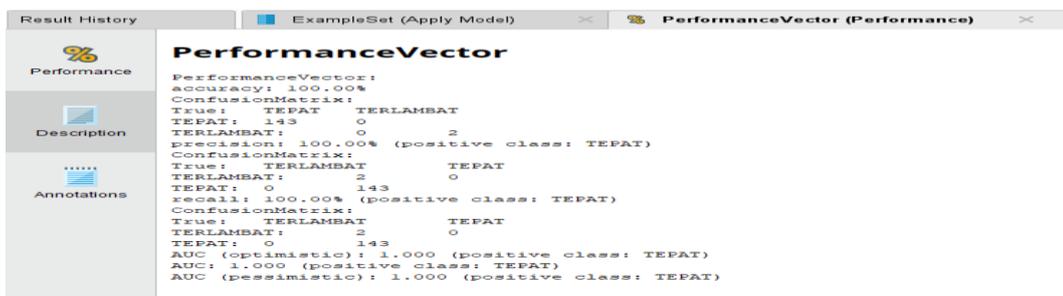
Gambar 24.
Hasil AUC



Gambar 253.
Hasil AUC (pessimistic)

Gambar di atas menampilkan nilai hasil AUC versi optimistic, versi umum, dan pesimistik. AUC mengukur performa model klasifikasi di berbagai tingkat threshold. Hasil AUC pada

gambar mendekati 1 yang berarti model *Random Forest* ini memiliki performa sangat baik dan dapat membedakan antara mahasiswa yang lulus tepat waktu dan yang terlambat.



Gambar 26.
Hasil performance vector

Hasil performance vector pada Gambar 27 mencakup berbagai matriks evaluasi dalam *Random Forest*. Di dalam gambar tersebut terlihat bahwa nilai akurasinya 100% dengan nilai presisi 100% (positive class: Tepat) dan recall 100%, serta nilai AUC 1.

Pemilihan Model Terbaik

Dalam penelitian ini, model terbaik dipilih berdasarkan evaluasi dan validasi terhadap hasil yang diperoleh dari kombinasi algoritma *K-Means* dan *Random Forest*. Kriteria evaluasi meliputi akurasi, presisi, recall, dan nilai AUC (Area Under Curve) untuk mengukur kinerja prediksi model.

Evaluasi Hasil *Random Forest*

Akurasi: Model *Random Forest* menunjukkan akurasi sebesar 100%, yang mengindikasikan bahwa model dapat memprediksi kelulusan mahasiswa secara tepat waktu dengan benar pada semua data uji. Precision dan Recall: Nilai precision dan recall masing-masing sebesar 100%, menunjukkan kemampuan model untuk meminimalkan kesalahan prediksi negatif dan mendeteksi semua mahasiswa yang lulus tepat waktu. AUC: Nilai AUC mendekati 1, menegaskan bahwa model memiliki performa sangat baik dalam membedakan antara kelas positif (tepat waktu) dan negatif (tidak tepat waktu).

Evaluasi Hasil *K-Means*:

Algoritma *K-Means* digunakan sebagai metode klusterisasi awal untuk mengelompokkan data berdasarkan kesamaan atribut. Pengelompokan ini mempermudah analisis pola data mahasiswa yang memengaruhi kelulusan. Hasil klusterisasi memberikan wawasan penting tentang faktor-faktor seperti IPK, jumlah SKS, dan status pekerjaan yang signifikan dalam menentukan kelulusan tepat waktu. Kombinasi *K-Means* dan *Random Forest*. Data hasil klusterisasi dengan *K-Means* digunakan sebagai input untuk model *Random Forest*, sehingga meningkatkan efisiensi dan akurasi prediksi. Pendekatan ini memastikan bahwa data dengan karakteristik serupa diproses lebih efektif dalam model klasifikasi. Fokus utama pada *Random Forest* adalah sebagai algoritma prediksi kelulusan, dengan *K-Means* berfungsi untuk mengelompokkan data di tahap awal analisis. Pemilihan Model Akhir. Berdasarkan hasil evaluasi, model

Random Forest dipilih sebagai model terbaik karena memberikan hasil akurasi tertinggi dan mampu menangani kompleksitas data dengan baik. Kombinasi dengan *K-Means* semakin memperkuat model dalam mengidentifikasi pola data secara lebih mendalam.

PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa Penggunaan metode data mining seperti *K-Means* dan *Random Forest* dapat efektif untuk memprediksi kelulusan mahasiswa tepat waktu pada program pascasarjana. Algoritma *K-Means* mampu mengelompokkan mahasiswa berdasarkan karakteristik yang relevan, seperti IPK dan jumlah kredit yang telah diselesaikan, sedangkan *Random Forest* digunakan untuk memberikan prediksi status kelulusan. Model yang dihasilkan mampu memberikan tingkat akurasi yang tinggi dengan precision dan recall yang baik, menunjukkan kemampuan model untuk mengenali pola kelulusan mahasiswa tepat waktu. Kombinasi kedua metode ini memberikan kontribusi yang signifikan dalam pengembangan model prediksi yang akurat, memberikan manfaat bagi mahasiswa dan institusi dalam merencanakan dan mengambil keputusan terkait kelulusan tepat waktu.

Hasil penelitian ini didukung oleh penelitian (Rizky et al., 2021) bahwa dari hasil perhitungan metode *K-Means* Clustering tersebut dapat menghasilkan prediksi kelulusan mahasiswa dengan lulus tepat waktu dan lulus terlambat di Universitas. Selain itu dapat mengetahui jumlah mahasiswa lulus berdasarkan tahun angkatan di Universitas. Penelitian ini juga didukung oleh (Hozairi et al., 2021) bahwa membantu manajemen universitas dalam mendeteksi kondisi mahasiswa sejak awal

untuk meningkatkan nilai akreditasi program studi. Dengan menggunakan model klasifikasi yang efektif, manajemen universitas dapat mendeteksi mahasiswa yang berisiko terlambat lulus atau tidak lulus sejak awal, sehingga dapat memberikan intervensi yang tepat waktu untuk membantu mahasiswa tersebut, yang pada akhirnya dapat meningkatkan kualitas pendidikan dan layanan kepada mahasiswa serta meningkatkan nilai akreditasi program studi.

SIMPULAN

Berdasarkan hasil penelitian maka dapat disimpulkan bahwa kombinasi metode data mining *K-Means* dan *Random Forest* efektif untuk memprediksi kelulusan mahasiswa tepat waktu pada program pascasarjana. Model ini memiliki tingkat akurasi yang tinggi dan dapat membantu mahasiswa dan institusi dalam merencanakan dan mengambil keputusan terkait kelulusan tepat waktu. Dengan demikian, penelitian ini dapat menjadi acuan bagi institusi pendidikan untuk meningkatkan kualitas pendidikan dan membantu mahasiswa mencapai tujuan akademik mereka.

Penelitian ini dapat dikembangkan lebih lanjut; Pertama, penyempurnaan algoritma dapat dilakukan dengan menggabungkan metode *K-Means* dan *Random Forest* dengan algoritma lain seperti Gradient Boosting (XGBoost atau LightGBM) untuk meningkatkan akurasi prediksi. Selain itu, penerapan teknik optimasi seperti *Grid Search* atau *Bayesian Optimization* dapat membantu menemukan parameter terbaik untuk model. Penambahan variabel baru, seperti tingkat kehadiran, keterlibatan dalam kegiatan akademik, serta faktor psikologis seperti motivasi dan stres, juga dapat memperkaya analisis.

Variabel eksternal, seperti sumber pendanaan mahasiswa, dapat dimasukkan untuk memberikan perspektif yang lebih komprehensif. Langkah selanjutnya adalah memastikan validasi model dengan menerapkan teknik seperti *k-fold cross-validation* untuk mengevaluasi performa model pada berbagai subset data. Pengujian pada dataset dari program studi atau universitas lain juga penting untuk mengukur sejauh mana model dapat digeneralisasi. Dari sisi visualisasi, pengembangan dashboard interaktif untuk menampilkan hasil prediksi dan pengelompokan mahasiswa dapat membantu pemangku kepentingan, seperti dosen atau pengelola program, dalam memantau perkembangan mahasiswa.

DAFTAR PUSTAKA

- Ameliatus, Q., & Fatah, Z. (2024). Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Menggunakan Aplikasi Rapidminer. *Jurnal Riset Teknik Komputer*, 1(1), 58–64. <https://doi.org/https://doi.org/10.37195/balok.v3i1.736>
- Andrianof, H., Gusman, A. P., & Putra, O. A. (2024). Implementasi Algoritma *Random Forest* untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia. *Jurnal Sains Informatika Terapan (JSIT) E-ISSN: 2828-1659*, 4(1), 24–28. <https://doi.org/https://doi.org/10.62357/jsit.v4i2.464>
- Armansyah, A., & Ramli, R. K. (2022). Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*,

- 6(1), 1–10.
<https://doi.org/10.29408/edumati.c.v6i1.4789>
- Arrohman, M. R., Khoiriyah, N., Fawaida, A., & Fithri, D. L. (2025). Analisis Klasifikasi Kelulusan Siswa Menggunakan Metode Knn (K-Nearest Neighbor) *Random Forest* Pada Dataset " Students Performance ". *JSMI*, 6(2), 299–305.
<https://doi.org/https://doi.org/10.24127/jmsi.v6i2.9102>
- Asro, A., Chaidir, J., Cahairuddin, C., & Friadi, J. (2025). Evaluasi Kinerja Algoritma Klasifikasi dalam Studi Kasus Prediksi Kelulusan di Universitas XYZ. *Zona Teknik: Jurnal Ilmiah*, 19(1), 15–22.
<https://doi.org/10.37776/zt.v19i1.1674>
- Budi Adnyana, I. M. (2023). Penerapan Teknik Klasifikasi untuk Prediksi Kelulusan Mahasiswa Berdasarkan Nilai Akademik. *Jurnal Teknologi Informasi Dan Komputer*, 7(3).
<https://doi.org/10.36002/jutik.v7i3.2428>
- Diantika, S., Nalatissifa, H., Maulidah, N., Supriyadi, R., & Fauzi, A. (2024). Penerapan Teknik Random Oversampling Untuk Memprediksi Ketepatan Waktu Lulus Menggunakan Algoritma *Random Forest*. *Computer Science (CO-SCIENCE)*, 4(1), 11–18.
<https://doi.org/10.31294/coscienc.e.v4i1.1996>
- Hanif, N., & Setiadi, T. (2020). Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung dan Pinjam Buku di Perpustakaan Menggunakan Metode C4.5. *JSTIE (Jurnal Sarjana Teknik Informatika) (E-Journal)*, 8(2), 66.
<https://doi.org/10.12928/jstie.v8i2.16950>
- Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining untuk Klasifikasi Kelulusan Mahasiswa dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2), 133.
<https://doi.org/10.21107/nero.v6i2.237>
- Maftucha, N., Salma, S., Rahmayuna, N., & Wakhidah, N. (2025). Perbandingan Algoritma Machine Learning dalam Memprediksi Kelulusan Siswa. *Jurnal Tekno Kompak*, 19(2), 116–128.
<https://doi.org/10.33365/teknokompak.v19i2.50>
- Mardzuki, T. H., Lubis, R., & Adiwijaya, F. F. (2024). Penerapan Algoritma *K-Means* Clustering pada Sistem Prediksi Kelulusan Tepat Waktu. *Komputika : Jurnal Sistem Komputer*, 13(2), 289–300.
<https://doi.org/10.34010/komputika.v13i2.14097>
- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, & Susanti. (2022). Penerapan Penerapan Data Mining untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Random Forest*. *SATIN - Sains Dan Teknologi Informasi*, 8(2), 122–131.
<https://doi.org/10.33372/stn.v8i2.885>
- Puspa, G., Rachmadhan Amri, M. F., & Nugraha, M. P. (2025). Pemanfaatan *Random Forest* untuk Prediksi Ketepatan Waktu

- Kelulusan Mahasiswa Studi Kasus: Institut Desain dan Bisnis Bali. *JIKO (Jurnal Informatika Dan Komputer)*, 9(2), 413. <https://doi.org/10.26798/jiko.v9i2.1886>
- Rizky, M., Fernanda, A., Sokibi, P., & Fahrudin, R. (2021). Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik dan Non Akademik Menggunakan Metode *K-Means*. *Jurnal Digit*, 11(1), 89–100. <https://doi.org/https://doi.org/10.51920/jd.v11i1.182>
- Sudarmadji, P. W., Fallo, N., & Peli, Y. S. (2023). Komparasi Algoritma Klasifikasi untuk Memprediksi Kelulusan Mahasiswa Program Studi Teknik Komputer Jaringan. *Jurnal Ilmiah Flash*, 8(2), 109. <https://doi.org/10.32511/flash.v8i2.998>
- Ubaidilah, R. M. (2023). Prediksi Kelulusan Mahasiswa Berdasarkan Data Kunjung dan Peminjaman Buku menggunakan Rapid Miner dengan Metode C.45 dan *Random Forest*. *International Research on Big-Data and Computer Technology: I-Robot*, 7(2), 14–20. <https://doi.org/10.53514/ir.v7i2.410>