JPIM: Jurnal Penelitian Ilmiah Multidisipliner
Vol. 03, No. 01, Tahun 2026, Hal. 1478-1492, ISSN: 3089-0128 (Online)

Sistem Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma *Random Forest* di Universitas Muhammadiyah Bangka Belitung

Anggun Apriliana[1], Yudistira Bagus Pratama[2], Arvi Pramudyantoro[3]
[1,2,3] Universitas Muhammadiyah Bangka Belitung
Email: anggunapriliana0@gmail.com

Article Info

Article history:

Received April 23, 2026
Revised Mei 04, 2026
Accepted Mei 16, 2026

Keywords:

Student Graduation, Random Forest, Machine Learning, Prediction, Streamlit

ABSTRACT

Student graduation delays are currently still a big problem in various universities in Indonesia because they are influenced by various academic and non-academic factors. The process to identify students who are at risk of graduating late is still carried out manually and is even considered less effective in processing large and complex data. This research aims to develop a web-based predictive system for timely student graduation using the Random Forest algorithm. The method used in this study is quantitative, with stages in the form of data preprocessing, modeling with Random Forest Regressor, and model evaluation using R-squared (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE) metrics. Data was obtained through an online questionnaire on alumni of the University of Muhammadiyah Bangka Belitung batch 2020–2021, which included academic and non-academic variables. The results showed that this model had a satisfactory performance with an MAE value of 0.0119, MSE of 0.0235, and R^2 of 0.8418. The model is then implemented in a web-based system using Streamlit which is able to process student data and generate predictions and visualizations automatically. This system can be used to support academic monitoring and more objective and efficient decision-making.

Article Info

Article history:

Received April 23, 2026
Revised Mei 04, 2026
Accepted Mei 16, 2026

Keywords:

Kelulusan Mahasiswa, *Random Forest*, *Machine Learning*, Prediksi, *Streamlit*

ABSTRAK

Keterlambatan kelulusan mahasiswa saat ini masih menjadi permasalahan besar di berbagai perguruan tinggi di Indonesia karena, dipengaruhi oleh berbagai faktor akademik maupun non-akademik. Proses untuk mengidentifikasi mahasiswa yang berisiko terlambat lulus masih dilakukan secara manual bahkan dinilai kurang efektif dalam mengolah data yang besar dan kompleks. Penelitian ini memiliki tujuan untuk mengembangkan sistem prediksi kelulusan mahasiswa tepat waktu yang berbasis web dengan menggunakan algoritma *Random Forest*. Metode yang digunakan dalam penelitian ini adalah kuantitatif, dengan tahapan berupa *preprocessing* data, pemodelan dengan *Random Forest Regressor*, dan evaluasi model menggunakan metrik *R-squared* (R^2), *Mean Absolute Error* (MAE), dan *Mean Squared Error* (MSE). Data diperoleh melalui kuesioner online pada alumni Universitas Muhammadiyah Bangka Belitung angkatan 2020–2021, yang mencakup variabel akademik dan non-akademik. Hasil penelitian menunjukkan bahwa model ini memiliki kinerja yang memuaskan dengan nilai MAE sebesar 0.0119, MSE sebesar 0.0235, dan R^2 sebesar 0.8418. Model tersebut kemudian diimplementasikan dalam sistem berbasis web menggunakan *Streamlit* yang mampu memproses data mahasiswa dan menghasilkan prediksi serta visualisasi secara otomatis. Sistem ini dapat digunakan untuk

mendukung pemantauan akademik dan pengambilan keputusan yang lebih objektif dan efisien.

Corresponding Author:

Anggun Apriliana
Universitas Muhammadiyah Bangka Belitung
Email: anggunapriliana0@gmail.com

PENDAHULUAN

Perguruan tinggi memiliki peran penting dalam menghasilkan sumber daya manusia yang berkualitas dan mampu bersaing di era global. Salah satu indikator keberhasilan institusi pendidikan tinggi adalah tingkat kelulusan mahasiswa dalam jangka waktu yang ditentukan. Tingginya persentase kelulusan tepat waktu tidak hanya menunjukkan seberapa efektif proses pembelajaran berlangsung, tetapi juga berpengaruh terhadap reputasi institusi, akreditasi program studi, serta kepercayaan masyarakat terhadap kualitas perguruan tinggi (Wahyudi & Wibowo, 2023). Namun, pada kenyataannya, masih banyak mahasiswa yang terlambat menyelesaikan pendidikan mereka, yang disebabkan oleh berbagai faktor baik yang berkaitan dengan akademik maupun non-akademik.

Faktor akademik yang berpengaruh pada kelulusan mahasiswa meliputi Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS), jumlah Satuan Kredit Semester (SKS), serta jumlah mata kuliah yang diulang (Yazid, 2024). Sementara itu, faktor non-akademik seperti motivasi belajar, tingkat stres, dukungan keluarga, kondisi sosial-ekonomi, serta keaktifan dalam organisasi juga memiliki peran penting dalam menentukan keberhasilan studi mahasiswa (Sytar & Ermatita, 2025). Di Universitas Muhammadiyah Bangka Belitung, proses identifikasi mahasiswa yang berisiko terlambat lulus masih dilakukan secara manual, seperti dengan melihat indeks prestasi semester serta tingkat kehadiran di kelas. Metode ini dinilai kurang efektif dalam mengolah data yang besar dan kompleks serta sulit dalam mengidentifikasi pola hubungan antar variabel secara menyeluruh.

Seiring dengan perkembangan teknologi, penerapan *machine learning* menjadi salah satu alternatif untuk menyelesaikan masalah ini (Amri et al., 2023). Berbagai penelitian sebelumnya menunjukkan bahwa algoritma seperti *Random Forest*, *Naïve Bayes*, *Decision Tree*, *K-Nearest Neighbor*, dan *Support Vector Machine* dapat digunakan untuk memprediksi kelulusan mahasiswa (Zeniarja et al., 2022). Di antara algoritma tersebut, *Random Forest* terbukti menunjukkan kinerja yang lebih unggul dalam menangani data dengan variabel yang kompleks serta mampu mengurangi risiko *overfitting* melalui pendekatan *ensemble learning* (Hidayat et al., 2025). Namun, sebagian besar penelitian sebelumnya masih berfokus pada perbandingan kinerja algoritma dan belum mengembangkan sistem prediksi yang dapat diimplementasikan secara langsung dalam lingkungan perguruan tinggi (Junaidi et al., 2024).

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pembangunan sistem prediksi kelulusan mahasiswa tepat waktu berbasis web menggunakan algoritma *Random Forest*. Pendekatan yang digunakan adalah *Random Forest Regressor* untuk memprediksi estimasi masa studi mahasiswa, yang kemudian diklasifikasikan menjadi dua kategori, yaitu lulus tepat waktu dan terlambat. Sistem dikembangkan menggunakan *framework Streamlit* sehingga memungkinkan pengguna untuk menginput data mahasiswa dan memperoleh hasil prediksi secara interaktif.

Kebaruan (*novelty*) dalam penelitian ini terletak pada integrasi antara data akademik dan non-akademik dalam satu model prediksi serta implementasi model ke dalam sistem berbasis web yang dapat digunakan secara langsung oleh pihak akademik. Dengan adanya sistem ini, diharapkan pihak universitas dapat melakukan pemantauan terhadap mahasiswa secara lebih dini, mengidentifikasi mahasiswa yang berisiko terlambat lulus, serta mendukung pengambilan keputusan akademik yang lebih objektif dan berbasis data.

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan memanfaatkan algoritma *Random Forest* untuk membangun sistem yang dapat memprediksi kelulusan mahasiswa tepat waktu. Metode kuantitatif ini dirancang untuk mempelajari data akademik dan non-akademik mahasiswa yang berbentuk angka untuk menemukan pola serta keterkaitan antar variabel yang berpengaruh terhadap kelulusan. Hasil penelitian ini diharapkan mampu menghasilkan model prediksi yang dapat dimanfaatkan sebagai alat bantu dalam pemantauan akademik dan pengambilan keputusan di Universitas Muhammadiyah Bangka Belitung. Penelitian ini menggunakan metode *waterfall*. Metode waterfall dipilih karena pendekatan ini dilakukan dengan cara yang terstruktur dan mengikuti urutan dalam pengembangan sebuah sistem. Metode waterfall dipilih karena metode ini dilakukan secara berurutan. Sistem yang dihasilkan akan memiliki kualitas yang baik, sebab pelaksanaannya dilakukan secara bertahap sehingga tidak terpusat pada satu tahapan tertentu (Renyut et al., 2022). Pada tahapannya metode *waterfall* memiliki 5 tahapan, yaitu kebutuhan, desain, implementasi, verifikasi dan pemeliharaan.

HASIL DAN PEMBAHASAN

Penentuan Variabel Akademik dan Non-Akademik

Tahap awal penelitian ini dilakukan melalui penentuan variabel yang akan dijadikan dasar dalam mengembangkan model prediksi kelulusan mahasiswa secara tepat waktu. Penentuan variabel ini dilakukan agar model yang dihasilkan dapat mencerminkan keadaan mahasiswa secara menyeluruh, baik dari aspek akademik maupun non-akademik. Data penelitian diperoleh dari 278 responden alumni Universitas Muhammadiyah Bangka Belitung angkatan 2020–2021 yang berasal dari Fakultas Keguruan dan Ilmu Pendidikan serta Fakultas Teknik dan Sains. Pengumpulan data dilakukan dengan menggunakan kuesioner yang dibuat melalui *Google Form* dan didukung oleh data akademik mahasiswa.

Variabel akademik yang digunakan terdiri atas Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS) semester 1 sampai semester 11, jumlah Satuan Kredit Semester (SKS) yang ditempuh, dan jumlah mata kuliah yang diulang. Variabel tersebut dipilih karena mampu menggambarkan performa akademik mahasiswa selama masa studi. IPK digunakan sebagai indikator capaian akademik secara keseluruhan, IPS menggambarkan perkembangan prestasi setiap semester, jumlah SKS menunjukkan progres studi, sedangkan jumlah mata kuliah ulang menunjukkan hambatan akademik mahasiswa.

Selain variabel akademik, penelitian ini juga menggunakan variabel non-akademik yang terdiri dari motivasi belajar, dukungan keluarga, tingkat stres, kondisi sosial ekonomi, pekerjaan paruh waktu, dan keaktifan organisasi. Variabel non-akademik digunakan karena keberhasilan studi mahasiswa tidak hanya dipengaruhi oleh kemampuan akademik, tetapi juga faktor psikologis dan lingkungan sosial. Motivasi belajar mencerminkan semangat mahasiswa dalam menyelesaikan studi, dukungan keluarga menunjukkan peran lingkungan terdekat, sedangkan tingkat stres berhubungan dengan tekanan selama perkuliahan.

Data non-akademik yang berbentuk kategori dan skala Likert kemudian ditransformasikan ke bentuk numerik agar dapat diproses oleh algoritma machine learning. Pada variabel motivasi belajar, dukungan keluarga, dan tingkat stres dilakukan proses agregasi nilai rata-rata untuk menghasilkan satu skor representatif pada masing-masing variabel. Langkah ini dilakukan agar data menjadi lebih sederhana, efisien, dan tetap mampu merepresentasikan kondisi responden secara utuh.

Berdasarkan hasil seleksi variabel, diperoleh kombinasi variabel akademik dan non-akademik yang dinilai relevan dalam memprediksi kelulusan mahasiswa tepat waktu. Penggunaan dua kelompok variabel ini diharapkan mampu meningkatkan kemampuan model dalam mengenali pola keterlambatan maupun keberhasilan studi mahasiswa secara lebih akurat dibanding hanya menggunakan data akademik semata.

Penerapan Algoritma Random Forest dalam Memprediksi Kelulusan Mahasiswa

***a) Preprocessing* Data**

Pada tahap ini, dilakukan *preprocessing* data pada *dataset* mahasiswa Universitas Muhammadiyah Bangka Belitung untuk mempersiapkan data sebelum tahap pemodelan. Tahapan ini meliputi pembersihan data, transformasi data kategorikal, penanganan *missing value*, normalisasi, dan seleksi fitur. Proses ini dilakukan agar data memiliki format yang konsisten dan siap digunakan dalam algoritma *Random Forest Regressor*, sehingga menghasilkan prediksi yang lebih optimal terhadap estimasi masa studi dan status kelulusan mahasiswa

1. Pembersihan dan Pemahaman Data

   Tahap awal yang dilakukan dalam *preprocessing* data adalah memahami struktur *dataset* serta melakukan pembersihan awal terhadap data yang digunakan. *Dataset* yang digunakan merupakan data mahasiswa Universitas Muhammadiyah Bangka Belitung yang terdiri dari variabel akademik dan non-akademik sebagai fitur dalam model *Random Forest Regressor*. Untuk memahami karakteristik awal *dataset*, dilakukan eksplorasi menggunakan fungsi dasar pandas yaitu df.head() untuk menampilkan beberapa data awal, df.info() untuk melihat struktur *dataset* seperti jumlah kolom, tipe data, dan data non-null,

serta df.isnull().sum() untuk mengecek *missing value* pada setiap kolom. Hasil dari proses ini menunjukkan bahwa terdapat beberapa nilai kosong terutama pada variabel IPS di beberapa semester sehingga diperlukan penanganan lebih lanjut pada tahap *preprocessing* berikutnya.

2. Seleksi Fitur

Seleksi fitur dilakukan untuk menentukan variabel-variabel yang memang signifikan dan memberikan kontribusi pada proses pemodelan. Tidak semua variabel di dalam *dataset* digunakan dalam model *Random Forest Regressor*, karena terdapat beberapa variabel yang tidak memiliki pengaruh langsung terhadap hasil prediksi estimasi masa studi mahasiswa. Proses seleksi fitur dilakukan menggunakan fungsi drop() untuk menghapus variabel yang tidak relevan dari *dataset*, yaitu Nama, Jenis Kelamin, Program Studi dan Tahun Lulus. Variabel tersebut dihapus karena tidak memiliki pengaruh langsung terhadap prediksi estimasi masa studi sehingga dapat meningkatkan efisiensi dan fokus model *Random Forest Regressor.*

3. Transformasi Data Kategorikal

Tahap ini diawali dengan pembuatan *dictionary mapping* untuk mengonversi variabel kategorikal menjadi nilai numerik, yaitu mapping_5 untuk skala Likert 1–5, mapping_stres untuk tingkat stres, mapping_ekonomi untuk kondisi ekonomi sosial, serta mapping_kerja dan mapping_org untuk variabel biner. Selanjutnya, kolom kategorikal dikelompokkan dalam cols_kategori dan distandarisasi dengan mengubah teks menjadi huruf kecil serta menghapus spasi berlebih menggunakan str.lower() dan str.strip() agar proses *mapping* tidak mengalami kesalahan. Setelah itu, setiap kolom dikonversi menjadi nilai numerik sesuai aturan *mapping* yang telah ditentukan sehingga seluruh variabel kategorikal berhasil diubah menjadi data numerik yang siap digunakan untuk pemodelan *Random Forest Regressor*.

Hasil dari proses transformasi menunjukkan bahwa seluruh variabel kategorikal berhasil dikonversi menjadi bentuk numerik sesuai dengan skala masing-masing. Variabel motivasi belajar dan dukungan keluarga menggunakan skala 1–5, tingkat stres menggunakan skala 1–3, kondisi sosial-ekonomi menggunakan skala 1–3, sedangkan pekerjaan paruh waktu dan keaktifan organisasi dikonversi menjadi nilai biner 0 dan 1.

4. Penanganan *Missing Value*

Setelah transformasi data kategorikal, ditemukan *missing value* pada variabel IPS 1 hingga IPS 11 akibat ketidaklengkapan pencatatan data. *Missing value* ditangani dengan pendekatan imputasi nilai rata-rata (mean imputation)**,** yaitu dengan mengisi nilai yang kosong menggunakan nilai rata-rata pada masing-masing variabel IPS. Pendekatan ini dilakukan untuk memastikan tidak ada nilai kosong pada data serta menjaga konsistensi distribusi data dalam *dataset*.

Pada tahap ini, dibuat daftar kolom menggunakan list comprehension untuk mengakses seluruh variabel IPS dari IPS 1 hingga IPS 11 sesuai dengan kolom yang tersedia dalam dataset. Selanjutnya, dilakukan perhitungan nilai rata-rata (*mean*) pada masing-masing kolom IPS. Nilai rata-rata tersebut digunakan untuk menggantikan nilai yang kosong (*missing value*) pada setiap variabel IPS menggunakan fungsi fillna().

Proses imputasi dilakukan secara kolom per kolom sehingga setiap variabel IPS memiliki nilai pengganti yang sesuai dengan karakteristik distribusi datanya. Pemilihan metode mean imputation bertujuan untuk menjaga representasi data asli serta mempertahankan pola distribusi tanpa menghilangkan variasi nilai dalam dataset.

5. Data *Cleaning* dan Normalisasi

Setelah penanganan *missing value*, dilakukan proses data *cleaning* dan normalisasi untuk memastikan nilai data berada dalam rentang yang wajar serta mengurangi pengaruh *outlier*. Tahap ini penting karena nilai ekstrem dapat mempengaruhi kinerja model *Random Forest Regressor*. Pada penelitian ini dilakukan pembatasan nilai (*clipping*) pada variabel IPK dan jumlah mata kuliah yang diulang sesuai batasan akademik yang berlaku.

Fungsi clip() digunakan untuk membatasi nilai agar tetap berada dalam rentang tertentu. Pada variabel IPK, nilai dibatasi antara 0 hingga 4 sesuai standar akademik perguruan tinggi. Sementara itu, jumlah mata kuliah yang diulang dibatasi antara 0 hingga 10 untuk menghindari nilai ekstrem yang tidak realistis dan menjaga konsistensi data.

6. Penentuan Variabel X dan Y

Setelah proses *cleaning* dan normalisasi data selesai dilakukan, tahap selanjutnya adalah mengidentifikasi variabel independen (*X*) dan variabel dependen (*Y*) yang akan diterapkan dalam pemodelan *Random Forest Regressor*. Variabel *X* berfungsi sebagai fitur atau atribut yang digunakan untuk prediksi, yang meliputi data akademik dan non-akademik mahasiswa, sementara variabel *Y* digunakan sebagai target prediksi yang merupakan perkiraan waktu studi mahasiswa.

7. Split Dataset

Tahap selanjutnya adalah membagi dataset menjadi data latih dan data uji. Pembagian dataset ini penting untuk menilai seberapa baik model dapat membuat prediksi pada data baru yang tidak digunakan sebelumnya dalam proses pelatihan.

Untuk memisahkan dataset, kita menggunakan fungsi train_test_split() dari *library* sklearn.model_selection dengan proporsi 80% untuk data latih dan 20% untuk data uji (test_size=0.2). Data latih berfungsi untuk mengembangkan model, sementara data uji digunakan untuk mengukur kinerja model pada data yang baru. Parameter random_state=42 diterapkan agar hasil pemisahan data tetap seragam.

b) Modelling Menggunakan *Random Forest Regressor*

Pada tahap ini, pembuatan model prediksi dilakukan dengan menggunakan algoritma *Random Forest Regressor*. Model ini berfungsi untuk mengeksplorasi keterkaitan antara variabel akademik dan non-akademik dari mahasiswa terhadap estimasi masa studi. Sebelum proses pelatihan, *dataset* telah dibagi terlebih dahulu menjadi data latih dan data uji agar model dapat dinilai dengan cara yang objektif.

Proses pembangunan model dilakukan menggunakan RandomForestRegressor() dari *library* sklearn.ensemble dengan menetapkan n_estimators=500 untuk jumlah pohon keputusan, dan max_depth=7 untuk membatasi kedalaman pohon agar menghindari *overfitting*. Parameter random_state=42 juga diterapkan untuk menjaga hasil tetap konsisten.

Setelah model diterapkan, pelatihan dilakukan dengan menggunakan fungsi fit() dengan data latih (X_train dan y_train), di mana model mempelajari hubungan antara variabel input

seperti IPK, IPS, dan variabel non-akademik terhadap target yaitu estimasi masa studi mahasiswa dan status kelulusan mahasiswa. Hasil dari proses pelatihan ini digunakan sebagai dasar dalam melakukan prediksi pada data uji.

c) Evaluasi Model

Dalam penelitian ini, evaluasi model dilakukan menggunakan beberapa metrik pengukuran, yaitu *Mean Absolute* Error (MAE), *Mean Squared* Error (MSE), dan *R-squared* (R^2 Score). Metrik-metrik ini dipilih karena mampu memberikan gambaran yang komprehensif mengenai performa model dalam melakukan prediksi.

Setelah melaksanakan serangkaian uji coba, didapatkan penilaian evaluasi sebagai berikut:

Tabel 1. Hasil Evaluasi Model

MAE	0.11919341449156581
MSE	0.02356799443698902
R2 Score	0.8418546473640792

Berdasarkan hasil evaluasi model yang telah dilakukan, diperoleh nilai MAE sebesar 0.11919341449156581, MSE sebesar 0.02356799443698902, dan R^2 Score sebesar 0.8418546473640792. Nilai MAE yang relatif kecil menunjukkan bahwa rata-rata kesalahan prediksi model terhadap nilai aktual cukup rendah, sehingga model mampu menghasilkan prediksi yang mendekati nilai sebenarnya. Nilai MSE yang rendah menunjukkan bahwa model tidak mengalami kesalahan prediksi yang signifikan. Hal ini mengindikasikan bahwa model cukup stabil dalam mempelajari pola hubungan antar variabel dan tidak terlalu terpengaruh oleh error yang ekstrem. Sementara itu, nilai R^2 Score yang berada pada angka 0.8418546473640792 menunjukkan bahwa model dapat menjelaskan hampir 84% perbedaan data yang digunakan dalam penelitian. Nilai ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam menangkap hubungan antara variabel input dan variabel target.

d) Penyimpanan Model

Model *Random Forest Regressor* yang sudah melalui proses pelatihan dan penilaian selanjutnya disimpan dagar dapat dipakai lagi pada saat penerapan sistem tanpa harus melakukan pelatihan ulang. Pada tahap ini digunakan library *joblib*, yaitu pustaka *Python* yang digunakan untuk menyimpan dan memuat objek *machine learning* dalam bentuk *file*. Model disimpan dalam format .pkl dengan nama *model_random_forest.pkl* dan *fitur_model.pkl.*

```
import joblib
# menyimpan model
joblib.dump(model, 'model_random_forest.pkl')
# menyimpan daftar fitur
joblib.dump(X.columns.tolist(), 'fitur_model.pkl')
print("\n✅ Model & fitur berhasil disimpan!")
```

Kode pada gambar di atas menunjukkan proses penyimpanan model *Random Forest Regressor* beserta fitur yang digunakan dalam penelitian. Pada bagian awal, fungsi joblib.dump() dipakai untuk menyimpan model yang sudah dilatih ke dalam *file*

model_random_forest.pkl. Tujuan dari proses ini adalah supaya model tidak perlu dilatih kembali saat akan digunakan, sehingga dapat menghemat waktu dan pemanfaatan sumber daya komputasi. Model yang telah disimpan ini juga dapat digunakan lagi dalam tahap implementasi sistem berbasis web untuk melakukan prediksi kelulusan mahasiswa.

Selanjutnya, pada bagian kedua, daftar fitur yang digunakan dalam proses pelatihan juga disimpan menggunakan joblib.dump() ke dalam *file fitur_model.pkl*. Penyimpanan fitur ini dilakukan agar struktur input pada saat proses prediksi tetap konsisten dengan data yang digunakan saat pelatihan model. Dengan demikian, sistem dapat memastikan kesesuaian antara data input dan model, sehingga menghindari kesalahan dalam proses prediksi. Penyimpanan model dan fitur ini sangat penting dalam implementasi machine learning karena mendukung efisiensi, kemudahan *deployment*, serta memungkinkan model digunakan secara berulang tanpa perlu dilakukan training ulang.

Implementasi Sistem Prediksi Kelulusan Tepat Waktu Berbasis Web Menggunakan *Streamlit*

Implementasi sistem dalam penelitian ini bertujuan untuk mengintegrasikan model *Machine Learning* yang telah dibangun ke dalam sebuah aplikasi berbasis web sehingga dapat digunakan secara langsung oleh pengguna. Sistem ini dikembangkan untuk membantu pihak akademik, kemahasiswaan, fakultas, dan program studi dalam melakukan analisis serta pengambilan keputusan terkait estimasi masa studi mahasiswa serta statu kelulusan mahasiswa.

Pada penelitian ini, model yang digunakan yaitu *Random Forest Regressor* diimplementasikan ke dalam sistem berbasis web menggunakan *framework Streamlit*. Model ini telah melalui tahapd pelatihanengan memanfaatkan data mahasiswa yang meliputi variabel akademik dan non-akademik, serta telah dinilai menggunakan metrik MAE, MSE, dan R^2 Score yang menunjukkan bahwa model memiliki performa yang cukup memuaskan dalam memprediksi estimasi masa studi mahasiswa.

Integrasi model ke dalam sistem dilakukan dengan cara memuat kembali model yang telah disimpan dalam format .pkl menggunakan *library joblib*. Model yang telah dimuat kemudian digunakan untuk melakukan prediksi terhadap data mahasiswa aktif Universitas Muhammadiyah Bangka Belitung yang diunggah oleh pengguna melalui sistem. Selain model utama, daftar fitur yang digunakan pada saat proses pelatihan juga dimuat kembali untuk memastikan bahwa struktur input data pada tahap implementasi tetap konsisten dengan struktur data saat *training*.

Sistem ini dirancang agar dapat memberikan kemudahan bagi pengguna dalam melakukan proses prediksi tanpa perlu memahami proses teknis *machine learning* secara mendalam. Pengguna cukup mengunggah *file* data mahasiswa dalam format *Excel*, kemudian sistem akan secara otomatis memproses data tersebut mulai dari tahap pembacaan data, *preprocessing*, hingga menghasilkan output berupa estimasi masa studi dan status kelulusan mahasiswa. Hasil prediksi kelulusan mahasiswa divisualisasikan dalam bentuk tabel hasil prediksi, diagram pie, diagram batang, serta ringkasan statistik jumlah mahasiswa tepat waktu dan terlambat.

a) Tampilan Halaman Utama dan Petunjuk Penggunaan Sistem

Gambar 1. Tampilan Halaman Utama dan Petunjuk Penggunaan Sistem

Pada tahap awal implementasi, sistem menampilkan halaman utama (*dashboard*) yang berfungsi sebagai antarmuka utama bagi pengguna dalam mengakses sistem prediksi kelulusan mahasiswa. Halaman ini dirancang dengan tampilan yang informatif dan *user-friendly* agar dapat digunakan oleh berbagai pihak, seperti bidang akademik, bidang kemahasiswaan, fakultas, maupun program studi tanpa memerlukan pemahaman teknis yang mendalam terkait *machine learning*.

Pada bagian atas halaman utama, ditampilkan identitas sistem berupa judul aplikasi "Sistem Prediksi Kelulusan Mahasiswa" serta logo Universitas Muhammadiyah Bangka Belitung. Tampilan ini bertujuan untuk memberikan informasi awal kepada pengguna mengenai fungsi sistem serta memperkuat identitas institusi.

Selain itu, sistem juga dilengkapi dengan fitur petunjuk alur penggunaan yang disajikan dalam bentuk visual langkah demi langkah. Petunjuk ini bertujuan untuk mempermudah pengguna dalam memahami cara kerja sistem secara keseluruhan. Alur penggunaan sistem terdiri dari lima tahapan utama, yaitu: mengunduh template *Excel* kosong, mengisi data mahasiswa, mengunggah *file Excel*, melakukan proses prediksi dan mengunduh hasil prediksi

Dengan adanya petunjuk penggunaan ini, sistem menjadi lebih mudah dipahami dan digunakan, terutama bagi pengguna yang tidak memiliki latar belakang teknis di bidang teknologi informasi. Fitur ini juga berperan dalam meminimalkan kesalahan penggunaan sistem serta memastikan bahwa proses prediksi dapat berjalan dengan baik sesuai dengan alur yang telah dirancang.

b) Proses Unduh Template *Excel* Kosong dan Unggah Data Mahasiswa

Pada tahap ini, sistem menyediakan dua fitur utama yang berkaitan dengan proses input data, yaitu fitur unduh template *Excel* kosong dan fitur unggah data mahasiswa. Kedua fitur ini dirancang untuk memastikan bahwa data yang digunakan dalam proses prediksi memiliki format yang seragam, valid, dan sesuai dengan struktur fitur yang telah digunakan pada tahap pelatihan model *Random Forest Regressor.*

Fitur pertama adalah unduh template *Excel* kosong, yang berfungsi sebagai acuan bagi pengguna dalam menyiapkan data mahasiswa. Template ini dibuat secara otomatis oleh sistem dan berisi kolom-kolom yang telah disesuaikan dengan kebutuhan model, seperti Nama, Jenis Kelamin, Program Studi, tahun masuk, IPK, IPS, jumlah SKS, jumlah mata kuliah yang diulang, serta variabel non-akademik seperti motivasi belajar, dukungan keluarga, tingkat stres, sosial-ekonomi, pekerjaan paruh waktu, dan keaktifan dalam berorganisasi.

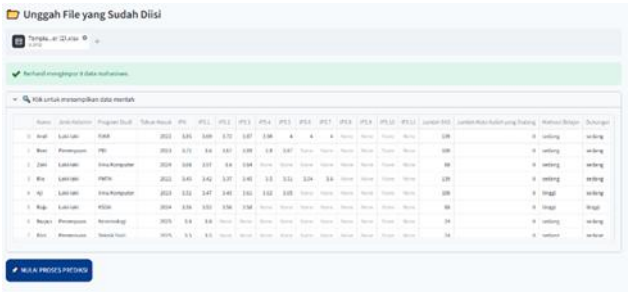
Dengan adanya template ini, pengguna tidak perlu membuat format data secara manual, sehingga dapat mengurangi risiko kesalahan input seperti ketidaksesuaian nama kolom, perbedaan penulisan, atau urutan atribut yang tidak sesuai dengan model. Selain itu, template ini juga memastikan bahwa data yang dimasukkan nantinya dapat langsung diproses oleh sistem tanpa perlu penyesuaian tambahan.

Untuk meningkatkan konsistensi pengisian data, pada beberapa variabel kategori telah diterapkan fitur *dropdown* (data validation) di dalam template *Excel*. Variabel seperti jenis kelamin, tingkat motivasi belajar, dukungan keluarga, tingkat stres, kondisi sosial-ekonomi, pekerjaan paruh waktu, dan keaktifan organisasi disediakan dalam bentuk pilihan yang telah ditentukan sebelumnya. Dengan adanya *dropdown* ini, pengguna hanya perlu memilih opsi yang tersedia tanpa harus mengetik secara manual.

Penerapan fitur *dropdown* ini sangat membantu dalam meminimalisir kesalahan input, seperti perbedaan penulisan (misalnya "Tinggi", "tinggi", atau "TINGGI"), kesalahan format, maupun nilai yang tidak sesuai dengan kategori yang telah ditetapkan oleh model. Selain itu, penggunaan *dropdown* juga memastikan bahwa data yang dikumpulkan lebih terstandarisasi, sehingga proses *preprocessing* dan prediksi oleh sistem dapat berjalan lebih optimal dan menghasilkan output yang lebih akurat.

Gambar 2. Tampilan Fitur Unduh Template *Excel* Kosong

Pada Tampilan Fitur Unduh Template Excel Kosong ditampilkan fitur tombol unduh template *Excel* kosong yang tersedia pada sistem. Fitur ini digunakan oleh pengguna untuk mengunduh format data yang telah disediakan sehingga dapat digunakan sebagai acuan dalam pengisian data mahasiswa. Setelah template diunduh dan diisi oleh pengguna, tahap berikutnya adalah fitur upload *file Excel*, di mana pengguna dapat mengunggah data mahasiswa yang telah dilengkapi sesuai dengan format yang telah disediakan. Sistem kemudian secara otomatis membaca *file* tersebut menggunakan *library Pandas* dan melakukan proses *parsing* data untuk memastikan seluruh kolom berhasil terbaca dengan benar.



Gambar 3. Tampilan Upload Data Mahasiswa dan Preview *Dataset*

Pada Tampilan Fitur Unduh Template Excel Kosong ditampilkan proses *upload* data mahasiswa ke dalam sistem serta tampilan *preview dataset*. Pada tahap ini, sistem menampilkan preview data dalam bentuk tabel interaktif, sehingga pengguna dapat melihat kembali isi data yang telah diunggah sebelum masuk ke tahap *pemrosesan* lebih lanjut. Fitur preview ini berfungsi sebagai bentuk validasi awal untuk memastikan tidak terdapat kesalahan pada data input, seperti kolom yang kosong, format yang tidak sesuai, atau kesalahan penulisan nilai.

Selain itu, sistem juga memberikan informasi jumlah data mahasiswa yang berhasil diimpor, sehingga pengguna dapat mengetahui apakah seluruh data telah terbaca dengan sempurna. Hal ini penting untuk memastikan bahwa proses prediksi yang dilakukan selanjutnya tidak mengalami *error* akibat data yang tidak lengkap atau tidak sesuai format.

Secara keseluruhan, fitur input data dan unduh template ini berperan penting dalam menjaga kualitas data yang masuk ke dalam sistem. Dengan adanya mekanisme template dan preview data, sistem dapat meminimalkan kesalahan input dari pengguna serta memastikan bahwa data yang digunakan dalam proses prediksi telah sesuai dengan standar yang ditetapkan oleh model *Machine Learning*.

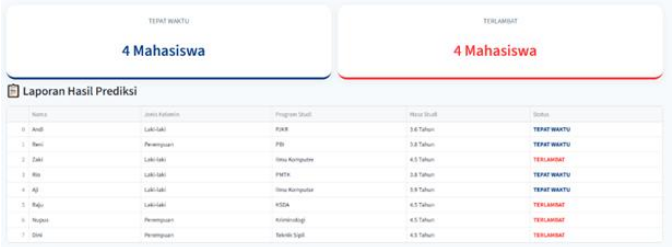
c) Proses Prediksi Menggunakan *Random Forest Regressor*

Pada tahap ini, dilakukan proses prediksi dengan menggunakan model *Random Forest Regressor* yang telah dilatih sebelumnya dan disimpan dalam format *file* .pkl. Model ini digunakan untuk menghasilkan dua *output* utama, yaitu estimasi masa studi mahasiswa dan status kelulusan (tepat waktu atau terlambat) berdasarkan data akademik dan non-akademik yang telah diunggah oleh pengguna melalui sistem berbasis web.

Sebelum proses prediksi dilakukan, sistem terlebih dahulu menyesuaikan struktur data *input* agar sesuai dengan fitur yang digunakan pada saat pelatihan model. Penyesuaian ini penting untuk memastikan tidak terjadi ketidaksesuaian jumlah maupun urutan variabel yang dapat mempengaruhi hasil prediksi. Setelah itu, data yang telah sesuai kemudian diproses menggunakan model untuk menghasilkan nilai prediksi.

```
X = df_proc.reindex(columns=fitur_sistem, fill_value=0)
y_pred = model.predict(X)
```

Pada kode tersebut, fungsi reindex() digunakan untuk menyamakan struktur data input dengan fitur model, sedangkan fungsi predict() digunakan untuk menghasilkan nilai estimasi masa studi mahasiswa dalam bentuk numerik (tahun).



Gambar 4. Laporan Hasil Prediksi

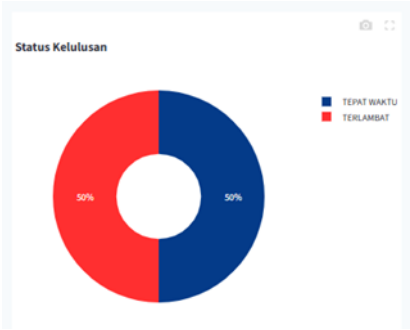
Hasil proses prediksi menunjukkan bahwa sistem mampu menghasilkan dua output utama, yaitu estimasi masa studi mahasiswa dalam bentuk nilai tahun dan status kelulusan mahasiswa yang dikategorikan menjadi "TEPAT WAKTU" dan "TERLAMBAT".

Estimasi masa studi diperoleh dari hasil perhitungan model *Random Forest Regressor* berdasarkan pola data yang telah dipelajari sebelumnya. Nilai ini kemudian dikonversi ke dalam bentuk yang lebih mudah dipahami, yaitu dalam satuan tahun.

Selanjutnya, sistem secara otomatis menentukan status kelulusan mahasiswa berdasarkan hasil prediksi tersebut. Mahasiswa dengan estimasi masa ≤ 4 tahun dikategorikan sebagai tepat waktu, sedangkan mahasiswa dengan nilai lebih dari > 4 tahun dikategorikan sebagai terlambat. Proses ini dilakukan untuk mempermudah interpretasi hasil oleh pihak akademik.

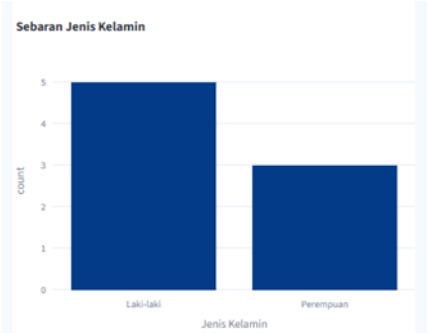
Dengan adanya dua output tersebut, sistem tidak hanya memberikan hasil berupa angka prediksi, tetapi juga memberikan informasi yang lebih informatif dalam bentuk keputusan kategorikal. Hal ini sangat membantu pihak akademik, kemahasiswaan, fakultas, dan program studi dalam melakukan analisis terhadap potensi keterlambatan studi mahasiswa.

Selain dalam bentuk tabel, hasil prediksi juga ditampilkan dalam bentuk visualisasi untuk memberikan gambaran distribusi data secara lebih jelas. Pada bagian visualisasi hasil prediksi, sistem menampilkan tiga jenis diagram yang disusun secara sejajar dalam satu tampilan *dashboard* untuk mempermudah analisis secara bersamaan.



Gambar 5. Visualisasi Distribusi Diagram *Pie Chart*

Pada gambar ini menampilkan diagram pertama yaitu diagram *pie chart* yang menunjukkan distribusi status kelulusan mahasiswa berdasarkan hasil prediksi model *Random Forest Regressor*. Pada diagram ini, hasil dibagi menjadi dua kategori, yaitu "TEPAT WAKTU" dan "TERLAMBAT". Visualisasi ini memberikan gambaran cepat mengenai proporsi mahasiswa yang diprediksi dapat menyelesaikan studi tepat waktu dibandingkan dengan mahasiswa yang berpotensi mengalami keterlambatan.



Gambar 6. Visualisasi Distribusi Diagram *Bar Chart*

Pada gambar ini menampilkan diagram kedua yaitu diagram *bar chart* yang menggambarkan distribusi mahasiswa berdasarkan jenis kelamin. Visualisasi ini digunakan untuk melihat sebaran data mahasiswa laki-laki dan perempuan dalam *dataset* yang dianalisis. Informasi ini tidak hanya berfungsi sebagai pelengkap data demografis, tetapi juga membantu pihak akademik dalam memahami komposisi mahasiswa yang menjadi objek prediksi dalam sistem.

Gambar 7. Visualisasi Distribusi Diagram *Bar Chart Horizontal*

Pada gambar ini menampilkan diagram ketiga yaitu diagram *bar chart horizontal* yang menampilkan distribusi mahasiswa berdasarkan program studi. Grafik ini menunjukkan jumlah mahasiswa pada setiap program studi yang terdapat dalam *dataset*. Dengan adanya visualisasi ini, pihak akademik dapat mengetahui sebaran data prediksi untuk setiap program studi, sehingga bisa dijadikan pertimbangan dalam evaluasi dan pengambilan keputusan akademik di tingkat fakultas atau program studi.

d) Fitur Unduh Hasil Prediksi

Pada tahap ini, sistem menyediakan fitur unduh hasil prediksi yang memungkinkan pengguna untuk menyimpan seluruh hasil analisis ke dalam bentuk *file Excel* (.xlsx). Fitur ini ditujukan untuk mempermudah pihak akademik, kemahasiswaan, fakultas, maupun program studi dalam mengelola dan mendokumentasikan hasil prediksi tanpa harus mengakses sistem secara langsung setiap saat.

Implementasi fitur ini dilakukan dengan memanfaatkan *library pandas* dan *xlsxwriter* untuk menghasilkan *file Excel* secara dinamis berdasarkan data hasil prediksi yang telah diproses oleh sistem. Data yang diunduh merupakan data mahasiswa yang telah melalui tahap *preprocessing* dan proses prediksi menggunakan model *Random Forest Regressor*, sehingga menghasilkan tambahan kolom berupa Estimasi Masa Studi dan Status Kelulusan.

Pada tampilan sistem, pengguna dapat menemukan tombol "*Download File Excel* (.xlsx)" yang berfungsi untuk mengunduh seluruh hasil prediksi. Tombol ini akan secara otomatis menghasilkan *file Excel* ketika ditekan, tanpa perlu proses tambahan dari pengguna. Hal ini menunjukkan bahwa sistem telah berhasil mengintegrasikan proses prediksi dengan fitur eksport data secara langsung dalam satu platform berbasis web.

Setelah *file* berhasil diunduh, pengguna akan mendapatkan *dataset* yang berisi seluruh data mahasiswa beserta hasil prediksi. *File* tersebut dapat dibuka menggunakan *Microsoft Excel* atau aplikasi *spreadsheet* lainnya. Di dalam *file* tersebut, terdapat kolom tambahan hasil

prediksi yang mencakup estimasi lama masa studi mahasiswa dalam satuan tahun serta klasifikasi status kelulusan, yaitu "TEPAT WAKTU" dan "TERLAMBAT".

Nama	Jenis Kelamin	Program Studi	Masa Studi	Status
Andi	Laki-laki	PJKR	3.8 Tahun	TEPAT WAKTU
Reni	Perempuan	PBI	3.7 Tahun	TEPAT WAKTU
Zaki	Laki-laki	Ilmu Komputer	4.5 Tahun	TERLAMBAT
Rio	Laki-laki	PMTK	3.8 Tahun	TEPAT WAKTU
Aji	Laki-laki	Ilmu Komputer	3.6 Tahun	TEPAT WAKTU
Raju	Laki-laki	KSDA	4.5 Tahun	TERLAMBAT
Nupus	Perempuan	Kriminologi	4.5 Tahun	TERLAMBAT
Dini	Perempuan	Teknik Sipil	4.5 Tahun	TERLAMBAT

Gambar 8. Hasil Unduhan *Excel* Data Prediksi Kelulusan Mahasiswa

Secara keseluruhan, fitur unduh hasil prediksi ini memberikan nilai tambah pada sistem karena memungkinkan hasil analisis digunakan kembali untuk berbagai kebutuhan institusional. Data yang telah *diekspor* dapat dimanfaatkan untuk penyusunan laporan akademik, evaluasi kinerja mahasiswa, serta sebagai bahan pertimbangan dalam pengambilan kebijakan di tingkat program studi maupun fakultas. Dengan demikian, fitur ini bukan hanya sekadar pelengkap bagi sistem, tetapi juga menjadi bagian penting dalam membantu proses pengambilan keputusan yang didasarkan pada data.

KESIMPULAN

Penelitian ini berhasil mengembangkan sebuah sistem yang dapat memprediksi kelulusan mahasiswa tepat waktu dengan menggunakan platform web dan memanfaatkan algoritma *Random Forest Regressor*. Berdasarkan hasil penelitian, variabel akademik dan non-akademik berpengaruh terhadap kelulusan mahasiswa tepat waktu. Algoritma *Random Forest* mampu digunakan untuk memprediksi masa studi serta mengklasifikasikan mahasiswa ke dalam kategori lulus tepat waktu dan terlambat. Model yang telah dikembangkan juga diterapkan dalam sistem web menggunakan *Streamlit* sebagai alat bantu pengambilan keputusan akademik. Secara keseluruhan, sistem ini memiliki potensi untuk menjadi sarana yang efektif dalam membantu pemantauan dan dpengambilan keputusan akademikengan cara yang lebih objektif dan efisien.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, ada beberapa rekomendasi untuk penelitian selanjutnya. Pertama, penambahan jumlah *dataset* dari berbagai angkatan, program studi, dan fakultas, penambahan jumlah *dataset* ini diperlukan agar model dapat memiliki kemampuan generalisasi yang lebih baik. Kedua, penambahan variabel yang lebih beragam seperti kehadiran, aktivitas akademik, dan faktor psikologis lainnya yang diharapkan dapat meningkatkan akurasi prediksi. Ketiga, penelitian selanjutnya disarankan untuk membandingkan atau mengombinasikan metode lain seperti *Gradien Boosting* dan *XGBoost* agar dapat memperoleh model yang lebih baik. Keempat, penggunaan teknik evaluasi yang lebih komprehensif seperti *cross-validation* perlu diterapkan agar hasil model lebih stabil dan tidak *overfitting*. Terakhir, dari sisi implementasi, sistem dapat dikembangkan lebih lanjut

menggunakan *framework* seperti *Flask* atau *FastAPI* agar lebih skalabel serta mudah diintegrasikan dengan sistem informasi akademik yang lebih luas.

DAFTAR PUSTAKA

Amri, Z., Kusrini, & Kusnawi. (2023). Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM. *Jurnal Pendidikan Informatika*, *7*(2), 187–196. https://doi.org/10.29408/edumatic.v7i2.18620

Hidayat, R., Saputra, H. T., Husnah, M., Bintang, M., Nazhmi, M. N., Azra, J., & Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produk di Supermarket. *Jurnal Sistem Informasi Dan Sistem Komputer*, *10*(1), 101–109. https://doi.org/10.51717/simkom.v10i1.703

Junaidi, S., Anggela, R. V., & Kariman, D. (2024). Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes , Random Forest , Support Vector Machine (SVM) dan Artificial Neural Nerwork (ANN). *Jurnal Of Applied Computer Science and Technology*, *5*(1), 109–119. https://doi.org/10.52158/jacost.v5i1.489

Renyut, D. H., Yuyun, & Ferdinand. (2022). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C . 45 (Studi Kasus , Sekolah Tinggi Ilmu Administrasi Trinitas Ambon). *Jurnal Sistem Informasi Dan Teknik Komputer*, *7*(2).

Sytar, M. H., & Ermatita. (2025). *Prediksi Kelulusan Mahasiswa Tepat Waktu Dengan Metode Random Forest Berdasarkan Klasifikasi Algoritma K-Means*. *8*, 391–410.

Wahyudi, A., & Wibowo, F. W. (2023). *Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Naive Bayes*. *14*(November), 132–139.

Yazid, A. S. (2024). Eksplorasi Data Akademik untuk Memprediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, *11*(4), 558–568.

Zeniarja, J., Salam, A., Ma, A., Informatika, T., Komputer, F. I., Nuswantoro, U. D., Imam, J., Kidul, P., Tengah, K. S., & Semarang, K. (2022). *Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa*. *18*(2), 102–108. https://doi.org/10.17529/jre.v18i2.24047