

Analisis Kelayakan Penerimaan Bantuan Pendidikan Beasiswa Luwu Timur Tahun 2025 Menggunakan Metode Random Forest

Komang Rosmiani¹, Gusti Ayu KM Anggraeni^{1*}, Aprizal², Suci Rahma Dani Rachman²

¹Teknik Informatika, Universitas Dipa Makassar, Makassar, Indonesia

²Universitas Dipa Makassar, Makassar, Indonesia

e-mail: ¹amirosmiani2304@gmail.com, ^{2*}ayu.anggi2004@gmail.com, ³aprizal@undipa.ac.id, ⁴sucirachman@undipa.ac.id

Abstrak

Proses seleksi penerima Beasiswa Luwu Timur masih didominasi oleh pemeriksaan dokumen secara manual, sehingga berpotensi menimbulkan inkonsistensi penilaian dan keterbatasan transparansi dalam pengambilan keputusan. Penelitian ini bertujuan menganalisis kelayakan penerimaan Beasiswa Luwu Timur Tahun 2025 menggunakan algoritma Random Forest serta menentukan variabel yang paling berpengaruh melalui analisis feature importance. Data yang digunakan berupa data sekunder historis penerima beasiswa yang diperoleh dari Bagian Kesejahteraan Rakyat Kabupaten Luwu Timur sebanyak 8.789 data. Tahapan penelitian meliputi pra-pemrosesan data, pembentukan fitur validasi persyaratan beasiswa, pelatihan model Random Forest, serta evaluasi kinerja model menggunakan metode confusion matrix dan pembagian data latih 80% dan data uji 20%. Hasil evaluasi data uji menunjukkan akurasi 72,41%, presisi 86,97%, recall 79,03%, F1-score 82,81%, dan ROC-AUC 62,13%. Analisis feature importance menunjukkan IPK sebagai variabel paling berpengaruh, diikuti umur dan semester. Hasil penelitian menunjukkan bahwa Random Forest dapat melakukan klasifikasi dengan baik dan transparan sebagai pendukung dalam pengambilan keputusan seleksi.

Kata kunci—Beasiswa, Random Forest, Machine Learning.

Abstract

The selection process for Luwu Timur Scholarship recipients is still dominated by manual document verification, which may lead to inconsistent assessments and limited transparency in decision-making. This study aims to analyze the eligibility of Luwu Timur Scholarship recipients for 2025 using the Random Forest algorithm and to identify the most influential variables through feature importance analysis. The data used consists of 8,789 historical secondary records of scholarship recipients obtained from the Social Welfare Office of Luwu Timur Regency. The research stages include data preprocessing, construction of scholarship requirement validation features, Random Forest model training, and performance evaluation using a confusion matrix with an 80% training and 20% testing data split. The evaluation results on the test data show an accuracy of 72.41%, precision of 86.97%, recall of 79.03%, F1-score of 82.81%, and ROC-AUC of 62.13%. Feature importance analysis indicates that GPA is the most influential variable, followed by age and semester. The findings demonstrate that Random Forest can provide reliable and transparent classification to support scholarship selection decisions.

Keywords—Scholarship, Random Forest, Machine Learning.

1. PENDAHULUAN

Pendidikan merupakan salah satu faktor penting dalam membentuk sumber daya manusia yang berkualitas. Pemerintah pusat dan daerah memiliki tanggung jawab dalam menyediakan akses pendidikan yang merata dan terjangkau bagi seluruh masyarakat. Salah satu upaya diberikan melalui bantuan pendidikan atau beasiswa yang ditujukan bagi peserta didik berprestasi maupun yang memiliki keterbatasan ekonomi. Melalui program beasiswa, diharapkan mahasiswa dapat melanjutkan pendidikan tinggi tanpa terkendala oleh biaya, serta mampu berkontribusi dalam meningkatkan kualitas SDM (Sumber Daya Manusia) daerah [1], [2].

Di Luwu Timur, Beasiswa Luwu Timur dirancang untuk meringankan beban biaya kuliah serta mendorong peningkatan kualitas SDM lokal. Proses pendaftaran saat ini telah dilakukan melalui laman resmi (<https://beasiswa.luwutimurkab.go.id/>) mahasiswa mengisi data, mengunggah berkas pendukung, lalu berkas diverifikasi oleh petugas di tingkat kecamatan. Mekanisme ini mempermudah alur administrasi, namun penilaian kelayakan masih dominan berupa pemeriksaan dokumen sehingga rentan pada keterbatasan konsistensi penilaian antar verifikator dan belum tersedianya analitik faktor yang menjelaskan variabel mana paling mempengaruhi kelayakan penerima.

Kondisi tersebut menimbulkan sejumlah konsekuensi kebijakan. Pertama, risiko salah sasaran meningkat ketika evaluasi hanya bertumpu pada cek administratif tanpa model analitik yang memadai. Kedua, akuntabilitas dan transparansi keputusan menjadi lebih sulit dijelaskan kepada publik jika tidak ada ukuran kinerja misalnya akurasi, presisi, recall atau Area Under Curve (AUC) yang terstruktur. Ketiga, tanpa analisis kepentingan variabel yang sistematis, pemerintah daerah kehilangan peluang untuk memperbaiki desain kriteria beasiswa pada periode berikutnya.

Pada konteks ini, pendekatan Machine Learning berpotensi meningkatkan objektivitas dan akurasi penentuan kelayakan. Random Forest adalah algoritma ansambel yang dibangun dari sejumlah pohon keputusan dan efektif digunakan pada data tabular dengan berbagai tipe variabel, seperti alamat sesuai KTP, IPK, semester, dan umur sesuai persyaratan beasiswa [3]. Algoritma ini relatif tahan terhadap noise dan interaksi non-linier, serta menyediakan ukuran kepentingan fitur (feature importance) yang dapat ditindaklanjuti sebagai dasar kebijakan [4]. Temuan terkini guna menunjukkan Random Forest mampu memprediksi status penerima bantuan pendidikan dengan performa baik dan menjelaskan faktor dominan pada data nyata [5], [6].

Di sisi lain, studi penentuan kelayakan beasiswa sering memakai metode alternatif seperti SMART (pembobotan manual) dan Naive Bayes (klasifikasi probabilistik). SMART membantu seleksi berbasis bobot kriteria, namun kinerja sangat bergantung pada penetapan bobot awal dan tidak otomatis belajar dari data historis [7], [8]. Naive Bayes bisa akurat pada skenario tertentu, tetapi asumsi independensi antar faktor sering tidak terpenuhi pada data sosial - ekonomi dan administratif penerima beasiswa [9]. Selain itu, algoritma XGBoost juga sering digunakan dalam klasifikasi kelayakan, yang mengombinasikan banyak pohon keputusan dalam proses boosting untuk meningkatkan ketepatan/akurasi model. XGBoost sangat efektif dalam menangani data besar dengan kemampuan optimasi dan regularisasi yang lebih baik, namun memerlukan waktu tuning hyperparameter yang lebih lama dan lebih kompleks [10].

Berdasarkan celah tersebut, diperlukan pendekatan yang belajar dari data, tahan terhadap variabel beragam, serta transparan dalam menjalankan faktor penentu keputusan. Random Forest memenuhi kebutuhan ini karena memiliki kemampuan prediksi yang kuat pada data dengan berbagai jenis variabel, tahan terhadap gangguan data (noise) dan hubungan antar variabel yang kompleks, serta dapat memberikan informasi mengenai fitur mana yang paling penting sebagai dasar kebijakan [11]. Selain itu, Random Forest juga lebih mudah digunakan, memberikan hasil yang stabil tanpa perlu pengaturan parameter yang rumit, dan lebih transparan, sehingga cocok digunakan untuk analisis kelayakan penerima beasiswa [12]. Dengan metode ini, proses seleksi bisa menjadi lebih adil, jelas, dan akurat, serta membantu memperbaiki kriteria dan prosedur verifikasi ke depannya [13].

Hasil yang diharapkan adalah model klasifikasi yang teruji pada data Beasiswa Luwu Timur, dilengkapi dengan metrik kinerja dan analisis tingkat kepentingan fitur sehingga memberikan gambaran objektif mengenai faktor yang memengaruhi kelayakan penerima beasiswa. Hasil penelitian ini diharapkan dapat menjadi rujukan akademik dan kontribusi ilmiah dalam pengembangan metode analisis data beasiswa, sekaligus mendukung transparansi dan akuntabilitas dalam evaluasi kelayakan berbasis data.

2. METODE PENELITIAN

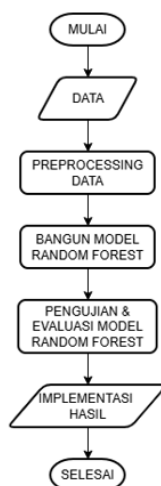
2.2 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dan deskriptif. Secara kuantitatif, penelitian ini menganalisis data historis penerima Beasiswa Luwu Timur yang diperoleh dari Bagian Kesejahteraan Rakyat (Kesra) Sekretariat Daerah Kabupaten Luwu Timur. Data diolah menggunakan metode Random Forest, yaitu algoritma machine learning berbasis ansambel pohon keputusan yang berfungsi untuk mengklasifikasikan kelayakan penerima secara objektif dan terukur.

Secara deskriptif, penelitian ini bertujuan untuk memberikan gambaran menyeluruh mengenai pola dan faktor-faktor yang paling berpengaruh dalam penentuan kelayakan penerima beasiswa berdasarkan hasil analisis komputasional. Pendekatan ini berfokus pada penerapan dan evaluasi model klasifikasi Random Forest untuk meningkatkan efisiensi, akurasi, dan transparansi proses seleksi beasiswa di Kabupaten Luwu Timur.

2.2 Tahapan Penelitian

Tahapan Penelitian ini memberikan pemahaman yang jelas mengenai langkah-langkah dalam pelaksanaan penelitian. Adapun prosedur penelitian yang dijalankan tercantum pada gambar 1.



Gambar 1 Tahapan Penelitian

Berikut merupakan penjelasan dari setiap tahapan pada penelitian ini sebagai berikut:

- a. **Pengumpulan Data**
Data yang digunakan merupakan data sekunder yang diperoleh dari Bagian Kesejahteraan Rakyat (Kesra) Sekretariat Daerah Kabupaten Luwu Timur. Variabel yang dikumpulkan meliputi alamat sesuai KTP, IPK, semester, serta umur. Data ini diperoleh dalam format digital yang akan digunakan untuk membangun model klasifikasi kelayakan penerima beasiswa.
- b. **Preprocessing Data**
Setelah data terkumpul, dilakukan pra-pemrosesan data untuk memastikan kualitas data yang akan digunakan. Tahapan ini meliputi:

- 1) Pembersihan data (data cleaning) untuk menghapus data yang tidak relevan, duplikat, atau salah.
 - 2) Penanganan data yang hilang (missing data), dengan menggunakan teknik imputasi atau penghapusan data yang tidak lengkap.
 - 3) Transformasi data untuk memastikan semua data dalam format yang dapat digunakan oleh algoritma Random Forest, termasuk normalisasi dan encoding variabel kategorikal jika diperlukan.
 - 4) Pembagian data menjadi data latih (training data) dan data uji (testing data), dengan proporsi umum 80% untuk data latih dan 20% untuk data uji.
- c. **Pembangunan Model Dengan Random Forest**
Setelah data siap, tahap selanjutnya yaitu pembangunan model dengan Random Forest. Model dibangun dengan cara:
- 1) Menentukan parameter awal Random Forest, seperti jumlah pohon keputusan (n_estimators), pengaturan pembobotan kelas (class_weight) untuk menangani ketidakseimbangan data, serta pengaturan random_state untuk memastikan konsistensi hasil eksperimen, n_jobs untuk mempercepat proses komputasi dengan memanfaatkan seluruh core CPU, dan kedalaman pohon (max_depth).
 - 2) Melatih model dengan data latih, di mana Random Forest akan membangun banyak pohon keputusan menggunakan subset acak dari data latih.
 - 3) Melakukan pengujian dan evaluasi dengan data uji, untuk melihat seberapa akurat model dalam memprediksi kelayakan penerima beasiswa.
- d. **Pengujian dan Evaluasi Model Random Forest**
Setelah model Random Forest dibangun, pengujian dilakukan menggunakan cross-validation untuk membagi data dan menghindari overfitting, serta confusion matrix untuk menghitung akurasi, presisi, recall, F1-score, dan AUC-ROC. AUC-ROC mengukur kemampuan model dalam membedakan antara kelas positif dan negatif, memberikan gambaran tentang keakuratan keseluruhan model pada berbagai threshold. Dilakukan juga analisis feature importance untuk mengidentifikasi faktor-faktor yang paling mempengaruhi keputusan klasifikasi oleh model Random Forest, guna memberikan wawasan lebih dalam mengenai kualitas dan kontribusi fitur dalam prediksi kelayakan.
- e. **Implementasi dan Penyajian Hasil**
Setelah model Random Forest dibangun dan dilakukan pengujian, tahap selanjutnya adalah implementasi model serta penyajian hasil klasifikasi. Implementasi dilakukan dengan menerapkan model untuk menghasilkan prediksi kelayakan penerima Beasiswa Luwu Timur Tahun 2025 berupa kelas layak atau tidak layak beserta nilai probabilitas pendukung. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan ROC-AUC. Confusion matrix dan ROC curve untuk menggambarkan kemampuan klasifikasi model. Analisis feature importance juga digunakan untuk mengenali kontribusi masing-masing faktor dalam proses pengambilan keputusan, sehingga hasil yang diperoleh dapat digunakan sebagai bahan pendukung dalam proses seleksi penerima beasiswa.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Bagian Kesejahteraan Rakyat (Kesra) Kabupaten Luwu Timur. Dataset berisi data pendaftar Beasiswa Luwu Timur Tahun 2025 sebanyak 8.789 data. Dataset tersebut terdiri dari beberapa atribut yang digunakan sebagai variabel dalam proses klasifikasi kelayakan penerima beasiswa.

Tabel 1 Deskripsi Data

Variabel	Keterangan
Umur	Usia mahasiswa
Alamat (Kecamatan)	Domisili mahasiswa sesuai KTP wajib Luwu Timur

Variabel	Keterangan
IPK	Nilai prestasi akademik mahasiswa
Semester	Semester aktif mahasiswa

3.2 Preprocessing Data

Pada penelitian ini, *preprocessing* data dilakukan secara bertahap untuk memastikan bahwa data yang digunakan dalam proses pelatihan model *Random Forest* berada dalam kondisi bersih, konsisten, dan siap dianalisis. Tahapan ini penting dilakukan untuk menghindari kesalahan dalam proses klasifikasi serta meningkatkan kualitas hasil prediksi model. Adapun tahapan *preprocessing* data yang dilakukan adalah sebagai berikut:

a. Pembersihan Data (*Data Cleaning*)

Tahap pertama dalam *preprocessing* adalah melakukan pembersihan data untuk memastikan bahwa dataset yang digunakan tidak mengandung data duplikat, data kosong, maupun kesalahan penulisan. Proses ini dilakukan dengan memeriksa seluruh atribut dataset seperti nama pemohon, tanggal lahir, kecamatan, IPK, semester, pendidikan, tahun angkatan, dan jenis beasiswa.

<pre>===== PREPROCESSING: DATA BERLABEL ===== Ukuran awal: (8789, 10) Ukuran setelah cleaning: (8789, 18) Missing value (kolom utama): IPK 0 SEMESTER 0 TANGGAL LAHIR 0 KECAMATAN 0 PENDIDIKAN 0 dtype: int64</pre>	<pre>===== PREPROCESSING: DATA TANPA LABEL ===== Ukuran awal: (8789, 9) Ukuran setelah cleaning: (8789, 17) Missing value (kolom utama): IPK 0 SEMESTER 0 TANGGAL LAHIR 0 KECAMATAN 0 PENDIDIKAN 0 dtype: int64</pre>
--	--

Gambar 2 Hasil *Preprocessing* Data

Gambar 2 menunjukkan hasil proses *preprocessing* pada dataset berlabel dan dataset tanpa label. Berdasarkan hasil pemeriksaan, jumlah data awal pada dataset berlabel adalah 8.789 data dengan 10 atribut, kemudian setelah proses pembersihan dan transformasi data jumlah atribut bertambah menjadi 18 atribut. Selain itu, dilakukan juga pemeriksaan terhadap nilai yang hilang pada beberapa kolom utama seperti IPK, semester, tanggal lahir, kecamatan, dan pendidikan. Hasil pemeriksaan menunjukkan bahwa seluruh kolom tersebut tidak memiliki missing value (nilai = 0) sehingga dataset dapat digunakan langsung dalam proses pelatihan model klasifikasi.

b. Transformasi Data

Transformasi data dilakukan untuk menyesuaikan dataset agar dapat digunakan oleh algoritma *Random Forest*.

Delimiter: . , v									
	NO	NAMA PEMOHON	TANGGAL LAHIR	KECAMATAN	IPK	SEMESTER	PENDIDIKAN	TAHUN ANGKATAN	JENIS BEASISWA
1	1	ARDIANSYAH JAMIL	1997-07-07	WASUPONDA	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
2	2	STEVANY SAKLIRESY	2003-09-27	MALILI	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
3	3	NUR FAJRI	2002-02-13	BURAU	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
4	4	JIHAN ALFIYYAH	2003-08-09	MALILI	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
5	5	J ANISA AMALIA SAHAR	2002-05-03	MALILI	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
6	6	RETTY WASPITA INDAH	2003-07-07	WASUPONDA	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
7	7	HAERUL UMAR	2003-03-22	BURAU	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
8	8	AGIL NABILA	2002-05-31	MANGKUTANA	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
9	9	AQILA DWI SALSABILA	2002-03-15	NUHA	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
10	10	K SINTIA SUPUTRI GIRI	2002-03-08	NUHA	0.0	2	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
11	11	A HASNAWATI AB	2002-07-07	MALILI	3.4	2	S1 PROFESI	2024	Program Pengembangan Keahlian (Profesi)
12	12	YEYEN PUSPITA SARI	2003-07-20	WASUPONDA	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
13	13	NIKEN AYU S. PUJI	2003-05-01	TOWUTI	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
14	14	SAVITRI WERE ULENG	2000-08-07	MALILI	3.8	3	S1 PROFESI	2024	Program Pengembangan Keahlian (Profesi)
15	15	AULIA ZAHIRANI	2003-06-19	MALILI	0.0	1	D4 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
16	16	NASWA NABILA PUTRI	2004-03-09	WOTU	0.0	1	D4 PROFESI	2025	Program Pengembangan Keahlian (Profesi)
17	17	UUN ZAKINAH	2002-07-12	MANGKUTANA	0.0	1	S1 PROFESI	2025	Program Pengembangan Keahlian (Profesi)

Gambar 3 Dataset Sebelum Transformasi

Pada tahap ini atribut tanggal lahir diubah menjadi umur, beberapa variabel dikonversi ke format numerik, serta dibuat variabel validasi persyaratan seperti IPK_OK,

SEMESTER_OK, UMUR_OK, dan ALAMAT_OK. Dataset hasil transformasi kemudian digunakan sebagai fitur dalam proses pelatihan model.

Delimit:											
	NO	NAMA PEMOHON	TAHUN LAHIR	UMUR	JENJANG	IPK_OK	SEMESTER_OK	UMUR_OK	ALAMAT_OK	LOLOS_SYARAT	label
1	1	ARDIANSYAH JAMIL	1997	28	PROFESI	1	1	1	1	1	1
2	2	STEVANY SAKLIRESY	2003	22	PROFESI	1	1	1	1	1	1
3	3	NUR FAJRI	2002	23	PROFESI	1	1	1	1	1	1
4	4	JIHAN ALFIYYAH	2003	22	PROFESI	1	1	1	1	1	1
5	5	PUTRI ANISA AMALIA SAHAR	2002	23	PROFESI	1	1	1	1	1	1
6	6	PRETTY WASPITA INDAH	2003	22	PROFESI	1	1	1	1	1	1
7	7	HAERUL UMAR	2003	22	PROFESI	1	1	1	1	1	1
8	8	AGIL NABILA	2002	23	PROFESI	1	1	1	1	1	1
9	9	AQILA DWI SALSABILA	2002	23	PROFESI	1	1	1	1	1	1
10	10	KADEK SINTIA SUPUTRI GIRI	2002	23	PROFESI	0	1	1	1	0	1
11	11	A.HASNAWATI AB	2002	23	PROFESI	1	1	1	1	1	1
12	12	YEYEN PUSPITA SARI	2003	22	PROFESI	1	1	1	1	1	1
13	13	NIKEN AYU S. PUJI	2003	22	PROFESI	1	1	1	1	1	1
14	14	DIATHALIA SAVITRI WERE ULENG	2000	25	PROFESI	1	1	1	1	1	1
15	15	AULIA ZAHIRANI	2003	22	PROFESI	1	1	1	1	1	1
16	16	NASWA NABILA PUTRI	2004	21	PROFESI	1	1	1	1	1	1
17	17	UUN ZAKINAH	2002	23	PROFESI	1	1	1	1	1	1
18	18	SHOFWATUN NISA	2001	24	PROFESI	1	1	1	1	1	1

Gambar 4 Dataset Setelah Transformasi

Gambar 4 menunjukkan perubahan struktur dataset setelah dilakukan proses transformasi data. Dataset yang sebelumnya masih berisi atribut mentah telah diubah menjadi dataset yang lebih terstruktur dengan tambahan variabel baru yang digunakan dalam proses klasifikasi menggunakan algoritma *Random Forest*.

c. Pembagian Data (Data Split)

Setelah tahapan *preprocessing* selesai dilakukan, data kemudian dibagi kedalam dua bagian utama menggunakan metode *train-test split* yang disediakan oleh library *Scikit-Learn* [14], [15]. Metode ini membagi dataset secara acak dengan proporsi tertentu, dalam penelitian ini 80% data digunakan sebagai data latih (*training data*) dan 20% data digunakan sebagai data uji (*testing data*).

```

=== JUMLAH DATA LATIH & DATA UJI ===
Total data      : 8789
Data latih (80%) : 7031
Data uji (20%)  : 1758

```

Gambar 5 Hasil Pembagian data Latih dan Data Uji

Gambar 5 menunjukkan pembagian dataset menjadi data latih sebesar 80% (7.031 data) dan data uji sebesar 20% (1.758 data) dari total 8.789 data yang digunakan dalam penelitian. Pembagian ini bertujuan untuk melatih dan mengevaluasi kinerja model *Random Forest*.

3.3 Pembentukan Model Random Forest

Setelah proses *preprocessing* dan pembagian dataset selesai dilakukan, tahap selanjutnya adalah membangun model klasifikasi menggunakan algoritma *Random Forest*. Algoritma ini merupakan metode *ensemble* yang bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir yang lebih akurat dan stabil. Dalam proses pelatihan model digunakan beberapa parameter awal yang bertujuan untuk mengoptimalkan kinerja algoritma *Random Forest*. Parameter tersebut ditunjukkan pada Gambar 6.

```

=== PARAMETER AWAL RANDOM FOREST ===
n_estimators : 300
random_state : 42
n_jobs       : -1
class_weight : balanced_subsample

```

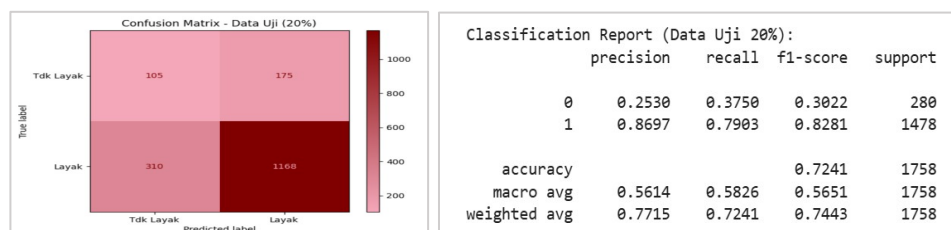
Gambar 6 Parameter Awal Model Random Forest

Berdasarkan gambar tersebut, model *Random Forest* dibangun dengan menggunakan 300 pohon keputusan ($n_estimators = 300$), nilai $random_state = 42$ untuk menjaga konsistensi hasil eksperimen, serta $n_jobs = -1$ yang memungkinkan proses komputasi memanfaatkan seluruh inti prosesor yang tersedia. Selain itu, parameter $class_weight = balanced_subsample$ digunakan

untuk menyeimbangkan distribusi kelas pada setiap subset data sehingga model dapat mempelajari pola klasifikasi dengan lebih baik.

3.4 Evaluasi Kinerja Model

Setelah model *Random Forest* berhasil dibangun, tahap selanjutnya adalah melakukan evaluasi terhadap performa model dalam melakukan klasifikasi kelayakan penerima beasiswa. Evaluasi dilakukan menggunakan data uji sebesar 20% dari total dataset, yang sebelumnya telah dipisahkan dari data latih dengan metode *train-test split*. Evaluasi ini bertujuan untuk mengetahui kemampuan model dalam memprediksi kelas Layak dan Tidak Layak secara akurat. Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi yaitu *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh dari hasil *confusion matrix*. Metrik tersebut digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan data ke dalam kelas Layak dan Tidak Layak.

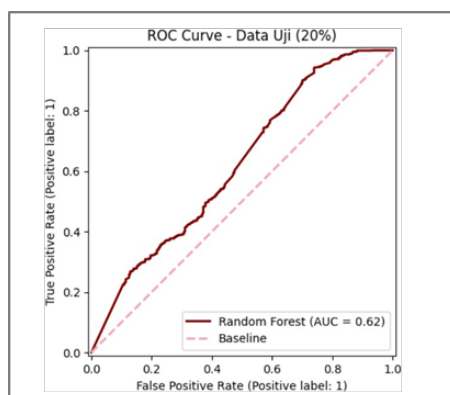


Gambar 7 Hasil *Confusion Matrix* dan *Classification Report*

Berdasarkan hasil evaluasi menggunakan *confusion matrix* pada data uji sebesar 20%, model *Random Forest* menghasilkan nilai *accuracy* sebesar 72,41%, *precision* 86,97%, *recall* 79,03%, dan *F1-score* 82,81%. Hasil tersebut menunjukkan bahwa model mampu mengidentifikasi kelas Layak dengan baik.

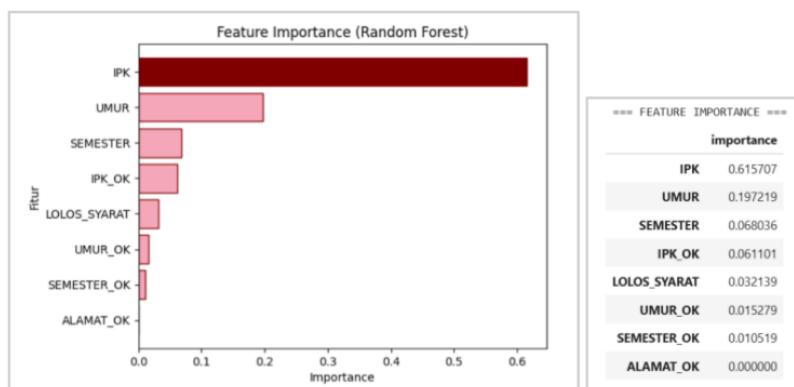
Berdasarkan confusion matrix, diperoleh *True Positive* (TP) sebanyak 1.168 data, *True Negative* (TN) 105 data, *False Positive* (FP) 175 data, dan *False Negative* (FN) 310 data. Jumlah *false negative* yang lebih tinggi dibandingkan *false positive* menunjukkan bahwa model masih menghasilkan beberapa kesalahan dalam mengidentifikasi mahasiswa yang benar-benar layak, sehingga nilai *recall* belum optimal.

Selain menggunakan *confusion matrix*, evaluasi kinerja model juga dilakukan menggunakan *ROC Curve* untuk melihat kemampuan model dalam membedakan kelas Layak dan Tidak Layak pada berbagai nilai *threshold*.



Gambar 8 ROC Curve

Hasil evaluasi menunjukkan nilai ROC-AUC sebesar 62,13%, yang menunjukkan bahwa model memiliki kemampuan dalam membedakan kelas Layak dan Tidak Layak. Nilai ROC-AUC yang berada di atas 0,5 menunjukkan bahwa model memiliki performa yang lebih baik dibandingkan dengan klasifikasi secara acak.



Gambar 9 Hasil *Feature Importance*

3.5 Feature Importance

Berdasarkan hasil analisis *feature importance* pada algoritma *Random Forest*, diketahui bahwa IPK memiliki kontribusi terbesar 61,57%, diikuti oleh umur sebesar 19,72%, dan semester 6,80% dalam menentukan kelayakan penerima beasiswa. Hasil ini menunjukkan bahwa prestasi akademik (IPK) menjadi faktor utama yang dipelajari model dalam proses klasifikasi. Sementara itu, variabel alamat 0% karena pemohon beasiswa 100% berasal dari Luwu Timur.

4. KESIMPULAN

- Pengelompokan penerima beasiswa dilakukan menggunakan algoritma *Random Forest* dengan memanfaatkan data akademik dan persyaratan administratif, seperti IPK, semester, umur, alamat, serta fitur validasi persyaratan. Model mampu mengelompokkan data menjadi dua kelompok, yaitu layak dan tidak layak, sehingga dapat menjadi alternatif pendukung seleksi yang lebih sistematis dan berbasis data.
- Berdasarkan analisis *feature importance*, variabel yang paling berpengaruh dalam menentukan kelayakan penerima beasiswa adalah IPK dengan kontribusi sebesar 61,57%, diikuti oleh umur sebesar 19,72% dan semester sebesar 6,80%. Hasil ini menunjukkan bahwa aspek akademik, khususnya IPK, menjadi faktor utama dalam proses klasifikasi kelayakan penerima beasiswa.
- Hasil evaluasi kinerja algoritma *Random Forest* pada data uji menunjukkan nilai *accuracy* sebesar 72,41%, *presisi* 86,97%, *recall* 79,03%, *F1-score* 82,81%, dan ROC-AUC 62,13%. Nilai *presisi* dan *F1-score* yang relatif tinggi menunjukkan bahwa algoritma mampu mengidentifikasi penerima yang layak. Sementara itu, nilai akurasi, *recall*, dan ROC-AUC menunjukkan bahwa model masih menghasilkan beberapa kesalahan klasifikasi dan memiliki ruang untuk peningkatan dalam membedakan kelas layak dan tidak layak.

5. SARAN

- Dapat dilakukan pengujian menggunakan algoritma *machine learning* lain atau proses penyetelan parameter (*hyperparameter tuning*) pada *Random Forest* untuk memperoleh performa model yang lebih optimal.

- b. Penambahan variabel lain yang relevan, seperti kondisi sosial dan ekonomi mahasiswa, diharapkan dapat memperkaya informasi yang digunakan oleh model dalam menentukan kelayakan penerima beasiswa.
- c. Model yang dihasilkan pada penelitian ini dapat dikembangkan lebih lanjut menjadi sistem pendukung keputusan berbasis aplikasi untuk membantu proses seleksi penerima beasiswa secara lebih efektif dan efisien.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Dipa Makassar dan semua pihak yang telah memberikan dukungan dan bantuan selama penelitian ini berlangsung. Penghargaan khusus disampaikan kepada Bapak Dr. Aprizal, S.Kom., S.E., M.M. dan Suci Rahma Dani Rachman S.Kom., M.T. atas bimbingan dan inspirasinya dalam penyusunan penelitian ini. Terima kasih juga kepada orang tua yang selalu memberikan doa, dukungan, dan semangat tanpa henti dalam setiap langkah penulis. Serta semua pihak yang telah membantu dengan cara apa pun.

DAFTAR PUSTAKA

- [1] B. P. Abd Rahman, S. A. Munandar, A. Fitriani, Y. Karlina, and Y. Yumriani, "Pengertian pendidikan, ilmu pendidikan dan unsur-unsur pendidikan," *Al-Urwatul Wutsqa Kaji. Pendidik. Islam*, vol. 2, no. 1, pp. 1–8, 2022.
- [2] A. R. Almuqorobin, "Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan Menggunakan Random Forest," 2021.
- [3] S. Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijawa Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022.
- [4] A. F. Azmi and A. Voutama, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix," *Komputa J. Ilm. Komput. Dan Inform.*, vol. 13, no. 1, pp. 111–119, 2024.
- [5] J. H. Gea, A. R. Himamunanto, and H. Budiati, "Analisis Kinerja Random Forest Dalam Deteksi Gejala Alergi Rongga Mulut Berbasis Warna Gusi," *Bull. Comput. Sci. Res.*, vol. 5, no. 4, pp. 734–744, 2025.
- [6] R. Novrian, T. Agustiani, M. Fikri, M. F. Hikmatulloh, M. E. Gunawan, and U. Firdaus, "Penerapan Algoritma Random Forest dalam Prediksi Status Penerima PIP pada Siswa: Studi Kasus pada SMK Amaliah 1," *Karimah Tauhid*, vol. 3, no. 2, pp. 1791–1799, 2024, doi: 10.30997/karimahtauhid.v3i2.11937.
- [7] Z. Setiawan *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2023.
- [8] Yustian Servanda, D. Djumhadi, and V. Vidy, "Analisis Metode Smart Dalam Penentuan Kelayakan Penerima Beasiswa Kip Kuliah," *J. Rekayasa Sist. Inf. dan Teknol.*, vol. 1, no. 4, pp. 220–228, 2024, doi: 10.59407/jrsit.v1i4.550.
- [9] I. Purnama Oktavia and N. Lestari Anggreini, "Implementasi Algoritma Naive Bayes Dengan Metode Klasifikasi Dalam Menentukan Siswa Penerima Bantuan Program Indonesia Pintar," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 6, pp. 11152–11158, 2024, doi: 10.36040/jati.v8i6.11241.
- [10] R. A. Pora, D. Maulina, and N. Trihartanti, "Komparasi XGBoost dan C4. 5 dalam Klasifikasi Risiko Kredit untuk Kelayakan Pinjaman di India," *Indones. J. Comput. Sci. Res.*, vol. 4, no. 2, pp. 53–61, 2025.
- [11] D. N. Handayani and S. Qutub, "Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 405–412, 2025, doi: 10.31004/riggs.v4i2.512.

- [12] S. Junaidi *et al.*, *Buku Ajar Machine Learning*. PT. Sonpedia Publishing Indonesia, 2024.
- [13] O. Pahlevi, A. Amrin, and Y. Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *J. Infortech*, vol. 5, no. 1, pp. 71–76, 2023.
- [14] S. Rahman, A. Sembiring, D. Siregar, I. G. Prahmana, R. Puspadini, and M. Zen, "Python: Dasar Dan Pemrograman Berorientasi Objek," *Penerbit Tahta Media*, 2023.
- [15] H. L. Sari, D. Tribuana, J. Sumah, and A. N. Indahsari, *Implementasi Python Untuk Machine Learning*. Serasi Media Teknologi, 2025.