

Komparasi Algoritma Decision Tree dan Random Forest dalam Prediksi Kelayakan Mahasiswa Penerima Beasiswa

Bentar Candra Perdana*¹, Elis Anita Farida²

¹Sistem Informasi, Fakultas Ilmu Komputer, Universitas PGRI Jombang, Jombang

²Kewirausahaan, Fakultas Ekonomi dan Bisnis, Universitas Anwar Medika, Sidoarjo

e-mail: *¹bentar.candra@upjb.ac.id, ²elisanitafarida@uam.ac.id

Abstrak

Pemberian beasiswa merupakan salah satu bentuk dukungan institusi pendidikan dalam meningkatkan motivasi belajar dan membantu mahasiswa yang kurang mampu dalam pembiayaan selama proses kuliah. Setiap mahasiswa memiliki peluang untuk mendapatkan beasiswa tersebut. Dalam proses seleksi penerima beasiswa sering kali insitusi atau perguruan tinggi menghadapi kendala dalam menentukan kelayakan secara objektif dan efisien. Oleh karena itu, diperlukan sistem pendukung keputusan berbasis data mining untuk membantu proses prediksi kelayakan mahasiswa penerima beasiswa. Penelitian ini dilakukan bertujuan untuk melakukan komparasi antara algoritma Decision Tree dan Random Forest dalam memprediksi kelayakan mahasiswa penerima beasiswa berdasarkan beberapa kriteria seperti program studi, penghasilan orang tua, pekerjaan orang tua, alat transportasi, alamat rumah dan jenis kelamin. Data penelitian ini diperoleh dari Perguruan Tinggi X yang ada di Jawa timur. Evaluasi performa model dilakukan dengan menggunakan nilai akurasi yang diperoleh. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mendapatkan nilai akurasi sebesar 82,67%, sedangkan dengan algoritma Random Forest memiliki nilai akurasi sebesar 86,00%. Dengan demikian algoritma Random Forest memiliki kinerja yang lebih baik dibandingkan dengan algoritma Decision Tree, dengan nilai akurasi yang lebih tinggi. Dengan demikian, dapat disimpulkan bahwa Algoritma Random Forest direkomendasikan sebagai algoritma yang lebih efektif dalam pengembangan sistem prediksi kelayakan mahasiswa penerima beasiswa di masa mendatang.

Kata Kunci— Decision Tree, Random Forest, Prediksi, Beasiswa, Akurasi

Submitted : 1 November 2025 | Accepted : 11 November 2025 | Published : 12 November 2025

1. PENDAHULUAN

Dunia pendidikan merupakan aspek yang penting dalam kehidupan mengingat Ki Hajar Dewantara (bapak pendidikan nasional) menyebutkan bahwa pendidikan itu mengalami proses yang panjang hingga akhir hayat mengingat kehidupan adalah bagian dari proses belajar [1]. Beasiswa merupakan salah satu bentuk penghargaan dan dukungan finansial yang diberikan kepada mahasiswa dengan tujuan mendorong prestasi akademik serta membantu mereka yang mengalami keterbatasan ekonomi. Dalam konteks pendidikan tinggi, beasiswa berperan penting dalam meningkatkan motivasi belajar dan pemerataan kesempatan pendidikan bagi seluruh lapisan masyarakat [2]. Beasiswa internal merupakan beasiswa yang berasal dari institusi itu sendiri baik itu secara penuh maupun sebagian sedangkan beasiswa eksternal merupakan beasiswa yang berasal dari luar institusi yang sumber pendanaannya biasanya diadakan oleh pemerintah dan swasta [3].

Namun, proses seleksi penerima beasiswa sering kali masih dilakukan secara manual dan subjektif, bergantung pada penilaian administratif maupun rekomendasi pihak tertentu. Kondisi ini dapat menimbulkan ketidakakuratan serta potensi ketidakefisienan dalam menentukan calon penerima yang benar-benar layak [4]. Hal ini dikarenakan banyaknya calon mahasiswa yang mengajukan beasiswa tersebut dan sistem seleksi yang dilakukan oleh pihak universitas masih bersifat konvensional. Sehingga pemanfaatan beasiswa masih kurang optimal [5].

Perkembangan teknologi informasi, khususnya dalam bidang *data mining* dan *machine learning*, memungkinkan dilakukannya analisis data secara lebih objektif dan sistematis. Teknik *machine learning* mampu mempelajari pola dari data historis untuk melakukan prediksi atau klasifikasi terhadap data baru [6]. Salah satu penerapan yang relevan adalah pada sistem pendukung keputusan untuk prediksi kelayakan penerima beasiswa berdasarkan berbagai variabel seperti nilai akademik, penghasilan orang tua, prestasi non-akademik, serta aktivitas organisasi mahasiswa [7].

Dalam penelitian klasifikasi, algoritma **Decision Tree** sering digunakan karena memiliki interpretabilitas yang tinggi dan mampu menghasilkan model yang mudah dipahami oleh pengguna non-teknis [8]. Namun, Decision Tree juga memiliki kelemahan, yaitu kecenderungan untuk mengalami *overfitting* terhadap data latih. Untuk mengatasi hal tersebut, algoritma **Random Forest** dikembangkan sebagai metode *ensemble learning* yang menggabungkan sejumlah pohon keputusan guna meningkatkan akurasi dan stabilitas hasil prediksi [9].

Beberapa penelitian terdahulu menunjukkan bahwa Random Forest cenderung memberikan performa yang lebih baik dibandingkan Decision Tree dalam berbagai kasus klasifikasi, termasuk pada bidang pendidikan [10]. Meskipun demikian, hasil tersebut dapat bervariasi tergantung pada karakteristik data dan parameter yang digunakan. Oleh karena itu, penting dilakukan studi komparatif untuk menilai sejauh mana perbedaan kinerja kedua algoritma tersebut dalam konteks prediksi kelayakan mahasiswa penerima beasiswa.

Dengan latar belakang tersebut, penelitian ini berfokus pada komparasi antara algoritma Decision Tree dan Random Forest dalam memprediksi kelayakan mahasiswa penerima beasiswa. Hasil penelitian diharapkan dapat memberikan kontribusi bagi pengembangan sistem pendukung keputusan berbasis *machine learning* yang lebih akurat, efisien, dan adil dalam proses seleksi beasiswa di lingkungan perguruan tinggi.

2. METODE

2.1 Jenis Penelitian

Penelitian ini termasuk dalam kategori penelitian kuantitatif dengan pendekatan **eksperimen** komparatif. Penelitian kuantitatif digunakan karena berfokus pada pengolahan dan analisis data numerik guna memperoleh hasil yang terukur secara objektif (Sugiyono, 2019). Pendekatan eksperimen komparatif digunakan untuk membandingkan kinerja dua algoritma *machine learning*, yaitu **Decision Tree (C4.5)** dan **Random Forest**, dalam melakukan prediksi kelayakan mahasiswa penerima beasiswa berdasarkan data historis yang tersedia.

2.2 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di lingkungan perguruan tinggi X dengan menggunakan data mahasiswa selama beberapa tahun akademik. Proses penelitian meliputi tahap pengumpulan data, prapemrosesan, pelatihan model, pengujian, serta analisis hasil. Adapun waktu pelaksanaan penelitian direncanakan berlangsung selama 1 bulan, dimulai dari tahap persiapan data hingga penyusunan laporan hasil penelitian.

2.3 Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini mencakup atribut-atribut seperti, program studi, Jenis kelamin, alamat rumah, penghasilan orang tua (ayah), pekerjaan orang tua (ayah), alat transportasi, NIM, dan layak atau tidak layak. Jenis data ini termasuk dalam kategori data terstruktur dengan tipe klasifikasi biner (binary classification), di mana variabel target adalah *kelayakan mahasiswa penerima beasiswa*. Menurut Han, Pei, dan Kamber (2012), penggunaan data terstruktur memungkinkan penerapan algoritma klasifikasi seperti Decision Tree dan Random Forest secara efisien untuk menemukan pola-pola prediktif.

2.4 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari:

1. Variabel Independen (X) meliputi program studi, Jenis kelamin, alamat rumah, penghasilan orang tua (ayah), pekerjaan orang tua (ayah), alat transportasi, NIM.
2. Variabel Dependen (Y) Adalah layak atau tidak layak

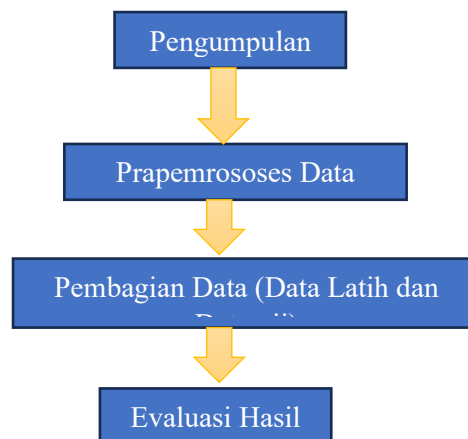
variabel independen adalah faktor yang bisa dianggap mempengaruhi variabel dependen, sementara itu variabel dependen merupakan sebuah hasil atau keluaran yang akan diprediksi.

2.5 Metode Pengumpulan Data

Metode pengumpulan data dilakukan melalui dokumentasi, yaitu pengambilan data mahasiswa dari Perguruan Tinggi X dan wawancara, untuk memperoleh informasi tambahan dari pihak pengelola beasiswa terkait kriteria penilaian kelayakan.

2.6 Tahapan Penelitian

Adapun penelitian dilakukan dengan tahapan sebagai berikut:



Gambar 1. Tahapan penelitian

2.6.1 Pengumpulan Data

Data mahasiswa dikumpulkan dari Perguruan Tinggi X diambil dari beberapa tahun mahasiswa yang ada di Perguruan Tinggi Tesebut.

2.6.2 Prapemrosesan Data (Data Preprocessing)

Pada tahap ini data di *cleaning*, atribut yang tidak perlu atau tidak berpengaruh dihapus, data yang dihapus seperti jenis kelamin, Alamat rumah, NIM.

2.6.3 Pembagian Data (Data Latih dan Data Uji)

Kemudian pada tahan paembagian data, dataset dibagi menjadi dua, yaitu data latih dan data uji. Data latih yaitu sebesar 70% dan data uji sebesar 30%. Pembagian dataset ini dilakukan secara acak atau random untuktetap menjaga agar ada keseimbangan kelas.

2.6.4 Evaluasi Hasil

Pada tahap ini hasil nilai akurasi kedua algoritma di komparasi mana yang menghasilkan akurasi yang lebih baik.

2.7 Alat dan Perangkat yang Digunakan

Penelitian ini menggunakan perangkat keras dan lunak sebagai berikut: Perangkat keras: Laptop Lenovo Ideapad Intel Core i3 1215U dan RAM 8192MB. Perangkat lunak yaitu Bahasa pemrograman Python Libraries: Scikit-learn, Pandas, NumPy, Matplotlib. Tools visualisasi yaitu menggunakan Google Colab

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini terdiri dari 1083 record. Terdiri dari 943 data layak penerima beasiswa dan 142 layak penerima beasiswa. Sebelum dilakukan proses pelatihan model, data melewati beberapa tahapan prapemrosesan diantaranya. Peneliti mempersingkat nama atribut biar lebih ringkas dalam pemrosesan data yaitu PS (Program Studi), AT (Alat Transportasi), PA (Penghasilan Orang Tua), PEA (Pekerjaan Orang Tua), dan LT (Layak atau tidak layak).

Tabel 1. Karakteristik Dataset

No	Nama	Tipe
1	Nama Sesuai Ijazah	Polynomial
2	Program Studi	Polynomial
3	Jenis Kelamin	Binomial
4	Alamat	Polynomial
5	Kecamatan	Polynomial
6	Alat Transportasi	Polynomial
7	Pekerjaan Ayah	Polynomial
8	Penghasilan Ayah	Polynomial
9	Layak atau Tidak Layak	Binomial

Kemudian dataset dilakukan Transformasi data (data transformation) untuk mengubah format atribut kategorikal menjadi numerik dengan teknik *label encoding*.

	PS	AT	PA	PEA	LT
0	8	10	2	3	tidak layak
1	8	10	12	3	tidak layak
2	8	10	5	0	tidak layak
3	4	10	1	3	tidak layak
4	4	10	1	5	tidak layak
...
1078	17	10	1	2	tidak layak
1079	17	10	1	4	tidak layak
1080	17	10	3	3	tidak layak
1081	17	10	6	3	tidak layak
1082	17	1	7	2	tidak layak

1083 rows x 5 columns

Gambar 3. Transformasi Data

Proses ini dilakukan agar data bisa di proses menggunakan machine learning, baik itu algoritma decision tree maupun algoritma random forest. Setelah itu data dibagi menjadi dua secara random, data latih dan data uji. Untuk data latih 70% dan untuk data uji 30% dalam program python seperti pada gambar berikut.

```
#membagi dataset dengan data testing yakni 70% data training dan 30% data testing  
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size=0.3, random_state=100)
```

Gambar 4. Pembagian data latih dan data uji

Proses berikutnya evaluasi hasil dengan dua algoritma, yaitu algoritma decision tree dan algoritma random forest. Untuk hasil akurasi menggunakan algoritma decision tree didapatkan seperti gambar berikut.

```
#Menampilkan akurasi dari model dan menampilkan confusion matrix  
import numpy as np  
print("Accuracy :", metrics.accuracy_score(Y_test,Y_pred))  
cm = np.array(confusion_matrix(Y_test, Y_pred))  
print(cm)  
#Classification Report  
cr = classification_report(Y_test, Y_pred)  
print(cr)  
  
Accuracy : 0.8276923076923077  
[[ 16  28]  
 [ 28 253]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| layak        | 0.36      | 0.36   | 0.36     | 44      |
| tidak layak  | 0.90      | 0.90   | 0.90     | 281     |
| accuracy     |           |        | 0.83     | 325     |
| macro avg    | 0.63      | 0.63   | 0.63     | 325     |
| weighted avg | 0.83      | 0.83   | 0.83     | 325     |


```

Gambar 5. Hasil akurasi menggunakan algoritma decision tree

Seperti yang ditunjukkan pada gambar 5. Bahwa algoritma decision tree mendapatkan hasil akurasi sebesar 0,827%. Proses selanjutnya evaluasi hasil menggunakan algoritma random forest dalam Bahasa pemrograman python seperti berikut.

```
accuracy = accuracy_score(Y_test, y_pred)  
print(f"Model Accuracy: {accuracy:.2f}")  
  
Model Accuracy: 0.86  
  
import numpy as np  
cm = np.array(confusion_matrix(Y_test, y_pred))  
print(cm)  
#Classification Report  
cr = classification_report(Y_test, y_pred)  
print(cr)  
  
[[ 13  31]  
 [ 16 265]]  


|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| layak        | 0.45      | 0.30   | 0.36     | 44      |
| tidak layak  | 0.90      | 0.94   | 0.92     | 281     |
| accuracy     |           |        | 0.86     | 325     |
| macro avg    | 0.67      | 0.62   | 0.64     | 325     |
| weighted avg | 0.83      | 0.86   | 0.84     | 325     |


```

Gambar 6. Hasil akurasi menggunakan random forest

KESIMPULAN

Dalam penelitian ini peneliti dapat menyimpulkan bahwa algoritma random forest memperoleh hasil akurasi yang lebih baik dibandingkan dengan menggunakan algoritma decision tree dalam memprediksi kelayakan mahasiswa penerima beasiswa di suatu perguruan tinggi X. Hasil evaluasi menggunakan algoritma decision tree memperoleh nilai akurasi sebesar 82,67% sedangkan algoritma random forest memperoleh nilai akurasi sebesar 86,0%.

DAFTAR PUSTAKA

- [1] A. Hidayat, S. Fatimah, dan D. N. Rosidin, "Challenges and Prospects of Islamic Education Institutions and Sustainability in The Digital Era," *Nazhruna J. Pendidik. Islam*, vol. 5, no. 2, hal. 351–366, 2021.
- [2] A. A. Putri and K. Kusumawati, "Sistem Pendukung Keputusan Menentukan Penerima Beasiswa Menggunakan Metode Comparative Performance Indeks Pada SMK Al-Hikmah," *J. Satya Inform.*, vol. 6, no. 01, pp. 10–21, 2021.
- [3] M. Hatta, "Implementasi Metode Simple Additive Weighting Pada Sistem DSS Seleksi Penerimaan Beasiswa Perguruan Tinggi," *J. Ilm. INTECH Inf. Technol. J. UMUS*, vol. 2, no. 01, hal. 31–40, 2020.
- [4] N. K. Daulay, A. T. Hidayat, and S. Shepty, "Penentuan Program Indonesia Pintar (PIP) Pada Siswa Kurang Mampu dengan Metode Preference Selection Index (PSI) Berbasis Web," *Bull. Comput. Sci. Res.*, vol. 4, no. 1, pp. 121–128, 2023.
- [5] S. Yunita dan V. N. Alaeysda, "Penerapan Algoritma C4. 5 Untuk Prediksi Penerimaan Beasiswa di SD 4 Pelangsian," *ICIT J.*, vol. 8, no. 2, hal. 181–193, 2022.
- [6] L. A. Supriyono *et al.*, *Buku Ajar Big Data dan Data Mining: Konsep, Metodologi, dan Aplikasi*. PT. Sonpedia Publishing Indonesia, 2025.
- [7] D. Kurniadi, F. Nuraeni, and A. F. Hazar, "Aplikasi Sistem Prediksi Mahasiswa Penerima Beasiswa Berbasis Web dengan Menerapkan Model Klasifikasi K-Nearest Neighbors," *J. Algoritm.*, vol. 21, no. 1, pp. 68–79, 2024.
- [8] M. Kristanaya, M. P. Azzahra, and M. Idhom, "Klasifikasi Status Rujukan Pasien Poliklinik Bandara Berbasis Random Forest dan Interpretabilitas Model Menggunakan SHAP," *Data Sci. Indones.*, vol. 5, no. 1, pp. 60–74, 2025.
- [9] A. Rifai, S. E. Permana, and R. Hamonangan, "Optimalisasi Klasifikasi Indeks Desa Membangun Menggunakan Metode Ensemble Dan Algoritma Random Forest," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 4, pp. 8226–8234, 2024.
- [10] D. Hartama and N. Amalya, "Perbandingan Algoritma Decision Tree, ID3, dan Random Forest dalam Klasifikasi Faktor-Faktor yang Mempengaruhi Karier Mahasiswa Ilmu Komputer," *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 1, pp. 72–80, 2025.