Sistem Rekomendasi Penerima Beasiswa pada Calon Mahasiswa Menggunakan Algoritma *XGBoost* di Universitas Islam Darul 'Ulum Lamongan

M.Tsalits Nururrohman[1*], Endang Setyati[2]

[1,2]Program Studi Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Jawa Timur
Email: *salisrohman@unisda.ac.id

Abstrak. Peningkatan pesat dalam teknologi informasi telah memungkinkan pemanfaatan teknik data mining dalam berbagai bidang, termasuk dalam seleksi kelayakan penerima beasiswa. Universitas Islam Darul 'Ulum Lamongan (Unisda) berkomitmen mendukung akses pendidikan yang merata melalui berbagai program beasiswa berbasis seleksi data akademik, ekonomi, dan demografis. Dengan menerapkan teknologi *machine learning*, Unisda berupaya meningkatkan akurasi dan keadilan dalam menentukan kelayakan penerima beasiswa. Metodologi yang digunakan dalam penelitian ini adalah framework CRISP-DM yang terdiri dari enam tahapan: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment*. Data yang digunakan mencakup informasi akademik dan ekonomi mahasiswa dari tahun ajaran 2023/2024 dan 2024/2025 dengan total 1327 entri data. Penelitian ini bertujuan untuk mengembangkan model rekomendasi kelayakan beasiswa bagi mahasiswa Universitas Islam Darul 'Ulum Lamongan menggunakan algoritma *Extreme Gradient Boosting* (*XGBoost*) dengan menggunakan fitur/atribut sebanyak 10 atribut/fitur termasuk kelas targetnya terdiri dari: Status_DTKS, Status_P3KE, Penghasilan_Ayah, Penghasilan_Ibu, Sumber_Daya_Listrik, Prestasi, Kepemilikan_Rumah, Nilai_Test, dan Rata-rata_Skor_Kuesioner, dan Output_Class. Hasil penelitian menunjukkan bahwa algoritma *XGBoost* memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan kelayakan beasiswa mahasiswa. Algoritma *XGBoost* mencapai tingkat akurasi 98,99% dengan nilai AUC 99,31%.

Kata kunci: *sistem rekomendasi, beasiswa, data mining, XGBoost, CRISP-DM*

1 Pendahuluan

Perkembangan pesat teknologi informasi telah membawa perubahan besar di berbagai sektor kehidupan, termasuk di dunia pendidikan, baik di tingkat sekolah menengah maupun perguruan tinggi. Universitas Islam Darul 'Ulum Lamongan (Unisda) sebagai salah satu institusi pendidikan tinggi juga turut memanfaatkan teknologi ini untuk mengelola berbagai data mahasiswa dan

Diterima 16 Maret 2025 Direvisi 15 Mei 2025 Diterima untuk publikasi 30 Mei 2025

aktivitas akademik secara lebih efektif untuk menghasilkan informasi berharga yang dapat mendukung peningkatan mutu pendidikan di lingkungan kampus.

Unisda juga berkomitmen mendukung akses pendidikan yang merata melalui berbagai program beasiswa berbasis seleksi data akademik, ekonomi, dan demografis mahasiswa. Proses seleksi beasiswa yang tradisional seringkali memakan waktu dan rentan terhadap bias atau kesalahan manusia. Dalam menghadapi tantangan ini, pemanfaatan machine learning dalam seleksi beasiswa telah terbukti meningkatkan akurasi dan keadilan, sebagaimana ditunjukkan oleh penelitian sebelumnya yang berhasil memprediksi penerima beasiswa berdasarkan faktor akademik dan sosial ekonomi [1], [2], [3].

Penelitian sebelumnya dilakukan oleh Asriyanik [4] yang membandingkan algoritma *Decision Tree* (DT) dan *Support Vector Machine* (SVM) dalam klasifikasi kelayakan beasiswa. Dari eksplorasi atribut tersebut menunjukkan kinerja SVM mencapai akurasi lebih tinggi 80,17% sedangkan DT mencapai tingkat akurasi sebesar 78,44%. Penelitian [1], [2], [3], [4], [5] mengembangkan sistem rekomendasi kelayakan penerima beasiswa menggunakan algoritma *Boosting Tree* yaitu *Extreme Gradient Boosting* (*XGBoost*). Proses pengembangan meliputi pemilihan *feature importance* [6], melakukan *preprocessing* pada *dataset*, serta penerapan *hyperparameter tuning* [7]. Selanjutnya berdasarkan penelitian Seo et al. [5], performa *XGBoost* menunjukkan akurasi yang sangat baik mencapai nilai AUC sebesar 87% dalam prediksi resiko *dropout* mahasiswa. Prediksi penerima beasiswa pada calon mahasiswa di Unisda menggunakan algoritma *XGBoost* dengan memanfaatkan teknik *hyperparameter tuning* dari *library GridSearchCV* untuk mengoptimalkan model yang dibangun. Penelitian ini bertujuan memberikan prediksi yang lebih akurat dalam menentukan calon mahasiswa yang layak menerima beasiswa di Unisda. Hasil yang diperoleh diharapkan dapat menjadi acuan bagi pengambil kebijakan kampus untuk menyempurnakan proses seleksi beasiswa dan meningkatkan pemerataan akses pendidikan.

2 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif untuk mengembangkan sistem rekomendasi kelayakan beasiswa pada calon mahasiswa di Unisda.

2.1. *Dataset*

Data yang digunakan berjumlah 1.327 *record* berasal dari penggabungan dua jenis *dataset* dan telah melalui sebagian tahap *preprocessing* yaitu penghapusan *record outlier*, cek data yang memiliki *record null* dan *record* yang mengandung duplikasi. *Dataset* primer diperoleh dari hasil penyebaran

kuesioner terdiri dari 20 pertanyaan dengan skala Likert 1-10, yang bertujuan untuk mengukur minat mahasiswa terhadap beasiswa serta kesiapan melanjutkan studi jika tidak lolos seleksi. Sementara itu, *dataset* sekunder terdiri dari data pendaftar program KIP-Kuliah di Unisda pada tahun 2022, 2023, dan 2024, yang mencakup informasi demografis, kondisi ekonomi, status sosial keluarga, serta prestasi akademik calon mahasiswa.

2.2. *Preprocessing*

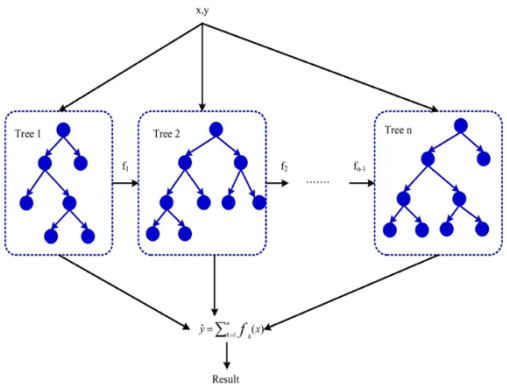
Tahap *preprocessing* data meliputi beberapa langkah penting. Pertama, dilakukan pembersihan data (*data cleaning*) secara manual untuk menghilangkan *record* yang mengandung *outlier*, karena *outlier* ini berpotensi mengganggu kinerja model prediksi, terutama akibat perbedaan pertimbangan konvensional dalam seleksi internal universitas dan juga dipengaruhi oleh ketidakwajaran dalam *entry* data oleh calon mahasiswa dengan keadaan yang sebenarnya. Kedua, pengecekan data kosong (*missing values*) dilakukan secara manual untuk memastikan kelengkapan setiap *record*. Ketiga, pemeriksaan data duplikat dilakukan dengan memanfaatkan fungsi ***df.duplicated().sum()*** pada bahasa pemrograman Python. Terakhir, transformasi data kategorikal menjadi data numerik dilakukan dengan metode label *encoding*, agar seluruh atribut bertipe objek dapat diproses oleh algoritma *Extreme Gradient Boosting* (*XGBoost*).

2.3. Metode *Data Mining* dan Algoritma *XGBoost*

Dalam penelitian ini, metode pengembangan *data mining* yang digunakan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM), karena metode ini menyediakan kerangka kerja sistematis dari pemahaman data hingga penerapan model, serta fleksibel terhadap berbagai jenis proyek *data mining* [8]. Tahapan metode *data mining* CRISP-DM dapat dilihat pada Gambar 1.

Gambar 1. *Workflow* CRISP-DM

Algoritma *Extreme Gradient Boosting* (*XGBoost*), sebagai metode *ensemble* yang kuat, telah menunjukkan kinerja superior dalam berbagai tugas klasifikasi dan prediksi di bidang pendidikan, termasuk dalam penentuan kelayakan beasiswa [9]. Diagram perancangan sistem rekomendasi kelayakan beasiswa menggunakan algoritma *XGBoost* di Unisda dapat dilihat pada Gambar 2.



Gambar 2. Pembuatan Pohon Keputusan pada *XGBoost*

Meskipun *XGBoost* awalnya dikembangkan untuk regresi, juga banyak diadaptasi untuk klasifikasi karena fleksibilitas dalam mendefinisikan fungsi objektif dan transformasi *output*. Dalam klasifikasi biner, fungsi *loss* yang digunakan adalah *binary cross-entropy* (*logistic loss*), yang dirumuskan sebagai:

$$Loss = -[y \log(p) + (1-y)\log(1-p)] \quad (1)$$

Karena sistem rekomendasi beasiswa ini merupakan klasifikasi biner (0 atau 1), maka *logloss* menjadi pilihan fungsi objektif yang paling sesuai. Untuk klasifikasi multikelas, digunakan *softmax loss*. Pada transformasi *output*, regresi menghasilkan nilai kontinu, sedangkan klasifikasi mengubah *output* menjadi probabilitas menggunakan fungsi sigmoid:

$$p = \frac{1}{1 + e^{-f(x)}} \quad (2)$$

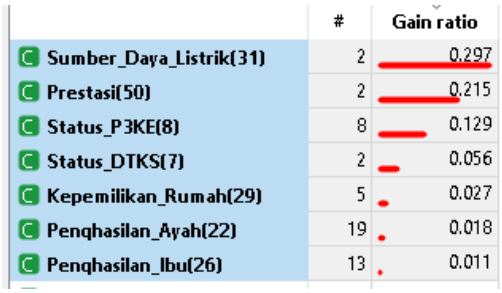
untuk klasifikasi biner, sedangkan fungsi *softmax* digunakan untuk multikelas. Evaluasi model klasifikasi menggunakan metrik seperti *logloss*, *error rate*, dan AUC, berbeda dengan regresi yang menggunakan RMSE, MAE, atau R^2. Penanganan label dalam klasifikasi juga dilakukan dengan encoding: 0 dan 1 untuk biner, dan 0 hingga K–1 untuk multikelas.

Studi komparatif oleh Hossain et al. [10] dan Islam et al. [11] menunjukkan bahwa algoritma ensemble seperti *XGBoost* seringkali mengungguli metode lain

seperti *Support Vector Machine* (SVM), *Naive Bayes*, atau *Decision Tree* dalam hal akurasi dan robusta. Hal ini disebabkan oleh kemampuan *XGBoost* untuk membangun model prediktif yang kuat melalui kombinasi *weak learners*, serta kemampuannya menangani fitur-fitur yang bervariasi (numerik dan kategorikal) secara efektif. Berdasarkan keunggulan tersebut dan bukti empiris dari penelitian sebelumnya, *XGBoost* dipilih sebagai algoritma utama dalam penelitian ini

2.4. Pemilihan Atribut/*Feature*

Untuk pemilihan atribut yang digunakan dalam model, dilakukan analisis *feature importance* menggunakan metode *gain ratio* pada aplikasi *Orange Data Mining*. Dalam data mahasiswa terdapat 28 atribut/fitur mahasiswa yang terdiri dari: ID_Pendaftar_PMB, Nama_Mahasiswa, NIK, NISN, Jenis_Kelamin, Tanggal _Lahir, Asal_sekolah, No.Handphone, Email, Status_DTKS, Status P3KE, Pekerjaan_Ayah, Penghasilan_Ayah, Pendidikan_Ayah, Status_Ayah, Pekerjaan_Ibu, Penghasilan_Ibu, Pendidikan_Ibu, Status_Ibu, Jumlah_Tanggungan, Kepemilikan_Rumah, Sumber_Daya_Listrik, Prestasi, Pilihan_ Prodi1, Pilihan_Prodi2, Nilai_Tes, Rata-rata_Skor_Kuesioner, dan Output_ Kelas. Pemilihan *feature* kategorikal dari aplikasi *Orange* dapat dilihat pada Gambar 3.



	#	Gain ratio
Sumber_Daya_Listrik(31)	2	0.297
Prestasi(50)	2	0.215
Status_P3KE(8)	8	0.129
Status_DTKS(7)	2	0.056
Kepemilikan_Rumah(29)	5	0.027
Penghasilan_Ayah(22)	19	0.018
Penghasilan_Ibu(26)	13	0.011

Gambar 3. *Gain Ratio* pada Fitur/Atribut Kategorikal yang Dipilih

Dari hasil seleksi ini, diperoleh 10 atribut utama, yang terdiri atas 7 atribut bertipe kategorikal (Sumber_Daya_Listrik, Prestasi, Status_P3KE, Status_DTKS, Kepemilikan_Rumah, Penghasilan_Ayah, Penghasilan_Ibu), 2 atribut bertipe numerik (Nilai_Tes dan Rata-rata_Skor_Kuisioner), dan 1 atribut *output class* bertipe kategorikal.

2.5. *Hyperparameter Tuning*

Dalam tahap *hyperparameter tuning*, digunakan *library GridSearchCV* dengan menetapkan tiga opsi nilai untuk setiap parameter yang diuji, sehingga didapatkan kombinasi parameter optimal yang mampu meningkatkan akurasi

model. Parameter yang disesuaikan dapat dilihat pada Tabel 2. *Learning rate* penting disesuaikan karena mengontrol seberapa besar kontribusi setiap pohon baru dalam membentuk model, sehingga mempengaruhi kecepatan dan ketelitian pembelajaran. *Max depth* perlu diatur untuk menyeimbangkan kompleksitas model agar tidak terjadi *overfitting* atau *underfitting*. Sedangkan *n_estimators* menentukan jumlah total pohon yang dibangun, dimana jumlah yang tepat memastikan model cukup belajar tanpa memperpanjang waktu pelatihan secara berlebihan.

2.6. Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan dengan menghitung nilai akurasi, *precision*, *recall*, dan *F1-Score*. Selain itu, untuk mengukur keseimbangan antara *true positive rate* dan *false positive rate*, dilakukan analisis *Area Under Curve* (AUC) menggunakan grafik *Receiver Operating Characteristic* (ROC), sebagaimana pendekatan evaluasi dalam penelitian oleh Saito dan Rehmsmeier [8]. Seluruh tahapan penelitian ini diimplementasikan menggunakan perangkat lunak Python dengan dukungan *library Scikit-learn*, *XGBoost*, dan *GridSearchCV*, serta lingkungan pengembangan *Jupyter Notebook*.

3 Hasil dan Pembahasan

Pada penelitian ini, implementasi tahapan CRISP-DM digunakan untuk membangun sistem rekomendasi kelayakan beasiswa dengan algoritma *XGBoost*. Setiap tahap dilakukan dengan sistematis, mulai dari pemahaman bisnis hingga evaluasi dan usulan penerapan model.

3.1 Implementasi Tahapan CRISP-DM

Pada penelitian ini, tahapan kerja mengikuti standar metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk mengembangkan sistem rekomendasi kelayakan beasiswa menggunakan algoritma *XGBoost*. Implementasi setiap tahapan dijelaskan sebagai berikut:

- ***Bussiness Understanding***
  Pada tahap ini, ditetapkan bahwa tujuan utama penelitian adalah membangun sistem prediksi untuk menentukan kelayakan calon mahasiswa dalam menerima beasiswa, berdasarkan informasi demografis, kondisi ekonomi, prestasi akademik, dan minat terhadap beasiswa.
- ***Data Understanding***
  Data yang digunakan terdiri dari data primer hasil kuesioner serta data sekunder dari pendaftar program KIP-Kuliah di Unisda pada tahun 2022 hingga 2024. Analisis awal dilakukan untuk memahami struktur data, karakteristik setiap atribut, serta memeriksa distribusi kelas target.

- ***Data Preparation***
  Tahap ini meliputi proses pembersihan data, seperti penghapusan *outlier* berdasarkan hasil analisis manual, penanganan *missing value*, serta penghapusan data duplikat menggunakan fungsi df.duplicated() dalam Python. Transformasi data dilakukan melalui label *encoding* untuk mengubah atribut kategorikal menjadi numerik, sehingga dapat diolah oleh algoritma *XGBoost*.
- ***Modeling***
  Pemodelan dilakukan menggunakan algoritma *XGBoost* dengan proses *tuning hyperparameter* melalui GridSearchCV. Parameter yang disesuaikan antara lain *learning rate*, *max depth*, dan *n_estimators* untuk memperoleh kombinasi model terbaik dalam meningkatkan akurasi prediksi.
- ***Evaluation***
  Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, *F1-score*, serta *Area Under Curve* (AUC) dari kurva ROC. Hasil evaluasi menunjukkan bahwa model yang dikembangkan mampu mengklasifikasikan kelayakan beasiswa dengan performa prediksi yang baik.
- ***Deployment***
  Model yang telah dihasilkan diusulkan untuk diimplementasikan sebagai sistem rekomendasi yang mendukung proses seleksi penerima beasiswa di Unisda, sehingga keputusan dapat dilakukan secara lebih objektif dan berbasis data.

3.2 Hasil Evaluasi *Preprocessing Data*

Berikut adalah hasil dari *preprocessing data* yang telah dilakukan:

- ***Cleaning Data* (Penanganan *Record Data Outlier*)**
  Data *outlier* yang mempengaruhi performa model, diidentifikasi secara manual berdasarkan analisis distribusi data dan kebijakan internal universitas dan ketidakwajaran *entry data* oleh calon mahasiswa yang tidak sesuai dengan keadaan sebenarnya, kemudian dihapus secara manual untuk meningkatkan keakuratan prediksi. Sehingga keseluruhan data yang digunakan sejumlah 1327 *record* adalah data setelah proses *data cleaning*.
- ***Missing Data* (Penanganan *Record Data* Bernilai *Null*)**
  Data yang memiliki nilai kosong (null) dilakukan secara manual. Setiap *record* yang tidak terisi sempurna pada atribut-atribut penting dianalisis, dan jika tidak memungkinkan untuk diperbaiki atau dilengkapi, maka *record* tersebut dihapus dari *dataset*. Pemeriksaan data yang bernilai *null* dilakukan dengan segmen kode Python ***df.isnull().sum()***.

- ***Duplicate Data*** **(Penanganan *Record* Ganda/Lebih dari Satu *Record*)**
  Pemeriksaan data ganda dilakukan secara otomatis menggunakan segmen kode ***df.duplicated().sum()*** dalam Python. *Record* duplikat yang terdeteksi kemudian dihapus untuk menghindari bias dan redundansi dalam proses pelatihan model.
- **Pemilihan Atribut dan *Labeling Data* Kategorikal & Kontinu**
  Diawali dengan menghapus 18 atribut yang tidak relevan atau tidak memiliki nilai penting untuk penilaian menggunakan segmen kode ***df.drop(['Nama_Mahasiswa(2)', … , Program_Studi_Pilihan 2(42)',], axis=1, inplace = True)***. Selanjutnya, melakukan pembagian label data terhadap atribut bertipe kategorikal (*object*) dengan segmen kode Python ***cate__val,*** sehingga muncul visualisasi *output* seperti Gambar 4.

```
['Status_DTKS(7)',
 'Status_P3KE(8)',
 'Penghasilan_Ayah(22)',
 'Penghasilan_Ibu(26)',
 'Kepemilikan_Rumah(29)',
 'Sumber_Daya_Listrik(31)',
 'Prestasi(50)',
 'Output_Class']
```

Gambar 4. Labeling Data Kategorikal

Lalu atribut yang bersifat kontinu (*numeric*) dengan segmen kode Python ***cont_val,*** sehingga muncul visualisasi output seperti Gambar 5.

```
['Rata-rata_Skor_Kuesioner', 'Nilai_Tes(51)']
```

Gambar 5. Labeling Data Kontinu

- **Transformasi Data (Label *Encoding* untuk Tipe Data Kategorikal)**
  Selanjutnya proses transformasi data dengan mengunakan label *encoding* pada atribut yang memiliki tipe data objek. Tujuan diubahnya data objek menjadi menjadi numerik agar data objek dapat diproses oleh algoritma *machine learning*. Label *Encoding* untuk data kategorikal dapat dilihat pada Tabel 1.
- **Melakukan Normalisasi pada Atribut Numerik**
  Melakukan *scaling* pada atribut numerik yang bersifat kontinu menggunakan *Standard Scaler*, untuk menstandarkan nilai fitur (atribut) sehingga memiliki rata-rata (*mean*) = 0 dan standar deviasi = 1.
- **Menentukan *Target Class* dan Pembagian *Dataset***
  Memilih atribut *Output_Class* yang akan menjadi kelas target pada prediksi kelayakan beasiswa, selanjutnya dilakukan pembagian *dataset* menjadi dua set: satu set untuk *data train* dan satu set untuk *data test*.

Tabel 1. Transformasi Data *Object* ke *Numeric*

No	Nama Atribut	Nilai	Label *Encoding*
1	Status_DTKS	Terdata; Belum Terdata	1; 0
2	Status_P3KE	Belum Terdata,; Desil 1……; Desil 8	0; 1; ……; 8
3	Penghasilan Ayah	Tidak Berpenghasilan; <250.000; 250.001 – 500.000; …. ; 4.750.001 – 5.000.000; >5.000.000	0; 1; 2; …. ; 17; 18
4	Penghasilan Ibu	Tidak Berpenghasilan; <250.000; 250.001 – 500.000; …. ; 4.750.001 – 5.000.000; >5.000.000	0; 1; 2; …. ; 17; 18
6	Sumber Daya Listrik	450; 900	0; 1
7	Prestasi	Ada; Tidak Ada	1; 0
8	*Output Class*	Layak; Tidak Layak	1; 0

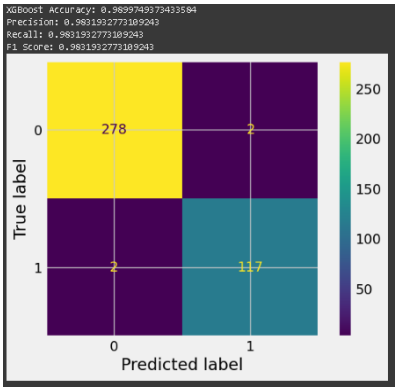
3.3 Hasil Evaluasi Model

Model *XGBoost* yang dibangun dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-Score*. Tabel 2 menunjukkan hasil evaluasi yang diperoleh.

Tabel 2. Hasil Evaluasi Model dengan *Hyperparameter Tuning*

No	Nama Parameter	Nilai	*Accuracy*	*Precision*	*Recall*	*F1-Score*	AUC
1	learning_rate	0.01					
2	n_estimators	100					
3	max_depth	3					
4	min_child_weight	1	98.99%	98.31%	98.31%	98.31%	99.24%
5	alpha, lamda	0					
6	colsample_bytree	0.7					
7	subsample	0.7					

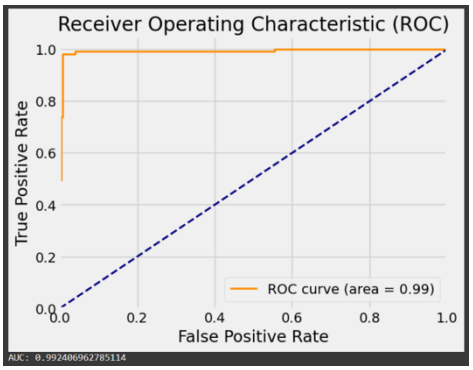
Berdasarkan Tabel 2, dengan mengatur 3 set nilai *tuning* parameter GridSearchCV sebelumnya, set *tuning* yang menunjukkan hasil paling optimal yaitu *learning_rate 0.01, n_estimators 100, max_depth 3, min_child_weight 1, alpha lamda 0, colsample_bytree 0.7, subsample 0.7*. Model menunjukkan kinerja yang baik dalam mengklasifikasikan kelayakan beasiswa, dengan nilai akurasi di 98.99%. Hasil *confusion matrix* dari model algoritma *XGBoost* dapat dilihat pada Gambar 6.



Gambar 6. *Confusion Matrix XGBoost*

3.4 Analisis Kurva ROC dan AUC

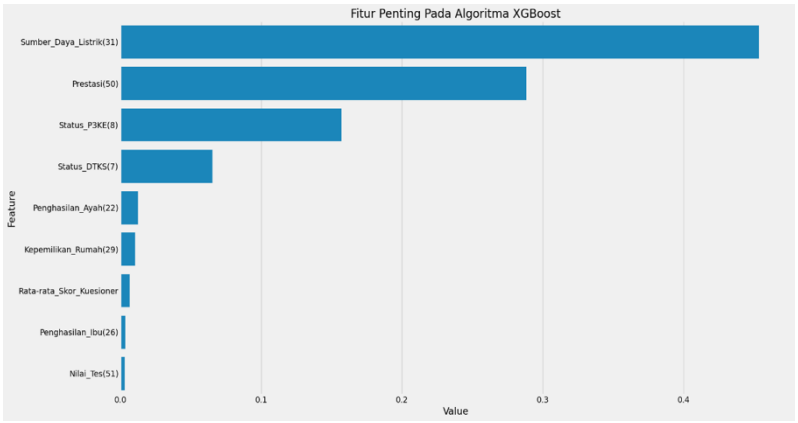
Selain menggunakan metrik evaluasi biasa, performa model juga dinilai berdasarkan kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under Curve* (AUC). Grafik ROC menunjukkan bahwa model mampu membedakan antara kelas layak dan tidak layak dengan baik. Nilai AUC yang diperoleh sebesar 99.24% mengindikasikan bahwa model memiliki kemampuan prediksi yang sangat baik. Visualisasi kurva ROC dapat dilihat pada Gambar 7.



Gambar 7. ROC *XGBoost*

3.5 *Intrepretasi Feature Importance*

Berdasarkan analisis *feature importance* yang dilakukan menggunakan aplikasi *Orange*, terdapat 10 atribut utama yang berkontribusi besar terhadap hasil prediksi. Tujuh atribut bertipe kategorikal, dua atribut bertipe numerik, dan satu atribut sebagai target klasifikasi. Atribut-atribut ini menunjukkan bahwa faktor kondisi ekonomi dan prestasi akademik memiliki pengaruh dominan terhadap peluang kelayakan beasiswa. Hasil visualisasi *feature importance XGBoost* dapat dilihat pada Gambar 8.



Gambar 8. *Feature Importance* Algoritma *XGBoost*

Fitur paling penting adalah Sumber_Daya_Listrik(31), yang memiliki bobot jauh lebih besar dibandingkan fitur lainya. Sedangkan beberapa fitur lain, seperti Prestasi(50) dan Status_P3KE(8) juga penting tetapi dengan nilai lebih rendah dibandingkan fitur Sumber_Daya_Listrik(31).

3.6 Pembahasan Hasil

Hasil penelitian menunjukkan bahwa implementasi *XGBoost* mampu menghasilkan model klasifikasi yang kuat untuk rekomendasi beasiswa. Penyesuaian *hyperparameter* seperti menggunakan GridSearchCV berpengaruh penting dalam meningkatkan kinerja model. Selain itu, pemilihan atribut penting melalui analisis *feature importance* berkontribusi dalam meningkatkan akurasi dan efisiensi model. Dengan demikian, sistem ini dapat menjadi alat pendukung keputusan yang efektif dalam proses seleksi penerima beasiswa di Unisda.

4 Kesimpulan

Hasil metrik pengukuran algoritma *XGBoost* bernilai *Accuracy_score* sebesar 98.99%, *Precision*_score 98.31%, *Recall* 98.31%, *F1_score* 98.31% dan AUC_*score* 99.24%.

Urutan dari *feature importance* pada algoritma *XGBoost* dari yang terbesar hingga yang terkecil yang paling berpengaruh pada penilaian kelayakan beasiswa adalah Sumber_Daya_Listrik(31), Prestasi(50), Status_P3KE(8), Status_DTKS(7), Penghasilan_Ayah(22), Kepemilikan_Rumah(29), Rata-rata_Skor_Kuesioner, Penghasilan_Ibu(26), dan Nilai_Test(51).

5 Referensi

[1] R. D. Putri, M. B. Jauhar, M. K. Akbar, and R. J. Rifa, "A machine learning approach to predict student scholarship awardees based on academic and socioeconomic factors," *J. Phys.: Conf. Ser.*, vol. 2393, no. 1, p. 012053, 2022. doi: 10.1088/1742-6596/2393/1/012053.

[2] A. H. Khan, M. I. Khan, M. Aslam, and M. S. Uddin, "Predicting student performance and scholarship eligibility in higher education using data mining techniques," in *2020 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2020, pp. 1-6, doi: 10.1109/ICASERT49906.2020.9345020.

[3] N. R. K. Siregar, M. R. Effendi, and R. A. Nasution, "Predictive analytics for student scholarship selection using machine learning," *J. Phys.: Conf. Ser.*, vol. 1339, no. 1, p. 012061, 2019. doi: 10.1088/1742-6596/1339/1/012061.

[4] Asriyanik and A. Pambudi, "Comparative SVM and decision tree algorithm in identifying the eligibility of KIP scholarship awardee," *Conf. Ser.*, vol. 4, no. 1, pp. 49–57, Dec. 2023, doi: 10.34306/conferenceseries.v4i1.625.

[5] E.-Y. Seo, J. Yang, J.-E. Lee, and G. So, "Predictive modelling of student dropout risk: Practical insights from a South Korean distance university," *Heliyon*, vol. 10, no. 11, p. e30960, Jun. 2024, doi: 10.1016/j.heliyon.2024.e30960.

[6] R. M. Syafei and D. A. Efrilianda, "Machine learning model using extreme gradient boosting (XGBoost) feature importance and light gradient boosting machine (LightGBM) to improve accurate prediction of bankruptcy," *Recursive J. Inform.*, vol. 1, no. 2, pp. 64–72, Sep. 2023, doi: 10.15294/rji.v1i2.71229.

[7] X. Wang, J. Li, S. Pang, and B. Tan, "Predicting student achievement based on fusion models," in *2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems (CCIS)*, Dali, China, Aug. 2023, pp. 82–92, doi: 10.1109/CCIS59572.2023.10263032.

[8] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.

[9] A. T. Haryono, S. B. Prabowo, and F. S. B. Putra, "Predicting student scholarship eligibility using ensemble machine learning methods," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1297, no. 1, p. 012028, 2023, doi: 10.1088/1757-899X/1297/1/012028.

[10] S. M. Hossain, M. Z. Islam, M. M. Islam, and S. B. Hasan, "A comparative study of machine learning algorithms for scholarship prediction," in *2023 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, 2023, pp. 1-6. doi: 10.1109/ICAICT59620.2023.10427386.

[11] S. R. Islam, M. A. Hasan, M. A. Islam, and M. S. Rahman, "Scholarship recommendation system using machine learning techniques: A comparative analysis," in *2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 2021, pp. 1-5. doi: 10.1109/IC4ME253218.2021.9678129.