ANALISIS PERBANDINGAN XGBOOST DAN CATBOOST PADA SISTEM PENDUKUNG KEPUTUSAN KLASIFIKASI PENERIMA BEASISWA PESANTREN

COMPARATIVE ANALYSIS OF XGBOOST AND CATBOOST IN THE DECISION SUPPORT SYSTEM FOR CLASSIFYING ISLAMIC BOARDING SCHOOL SCHOLARSHIP RECIPIENTS

Rizal Rachman[1], Popon Dauni[2], Sri Erina Damayanti[3] , Hanhan Maulana[4] , Bobi Kurnia[5], Sri Ednawati Rainarli[6], Adam Mukharil Bachtiar[7]
E-mail: [1*]rizal.75725020@mahasiswa.unikom.ac.id, [2]popon.75725026@mahasiswa.unikom.ac.id, [3]sri.75725027@mahasiswa.unikom.ac.id
[1] Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya
[2]Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Sistem Informasi, Universitas Kebangsaan Republik Indonesia
[3]Program Studi Teknik Informatika, Fakultas Industri Kreatif, Universitas Teknologi Bandung
[4,5,6,7] Doktor Sistem Informasi, Fakultas Pascasarjana, Universitas Komputer Indonesia

Abstrak

Penelitian ini membandingkan kinerja algoritma XGBoost dan CatBoost dalam klasifikasi penerima beasiswa pada lingkungan pesantren. Model dirancang untuk menghasilkan rekomendasi kategori beasiswa A (Tahfidz Unggulan), B (Akademik dan Prestasi), dan C (Kebutuhan Ekonomi) yang dapat digunakan sebagai alat bantu keputusan bagi admin dan pengurus pesantren dalam proses seleksi. Tantangan utama pada data adalah ketidakkonsistenan penilaian manual, keberadaan fitur yang berpotensi menimbulkan kebocoran informasi, serta distribusi kelas yang tidak seimbang, terutama pada kategori B. Untuk mengatasinya, dilakukan pembersihan data, eliminasi fitur yang bersifat deterministik terhadap label, rekayasa tujuh fitur turunan, serta penerapan class weighting pada kelas minoritas. Evaluasi menggunakan 10-fold stratified cross-validation untuk menilai stabilitas performa model. Hasil menunjukkan bahwa XGBoost memiliki kinerja paling stabil, dengan akurasi 86,42% serta F1-score yang lebih konsisten pada sebagian besar kelas dibandingkan CatBoost. Model juga mampu memanfaatkan fitur turunan secara lebih efektif pada kategori A dan C yang memiliki jumlah data lebih besar. Sementara itu, kategori B masih menjadi tantangan karena keterbatasan sampel yang membatasi kemampuan model mengenali pola secara optimal. Hasil penelitian ini menunjukkan bahwa pendekatan berbasis machine learning dapat mengurangi subjektivitas dan mempercepat proses seleksi, karena admin cukup memasukkan data santri dan model menyediakan rekomendasi yang kemudian diverifikasi oleh pengurus. Pendekatan ini dapat direplikasi pada lembaga pendidikan lain dengan karakteristik populasi yang serupa.

Kata kunci: *Klasifikasi beasiswa, XGBoost, CatBoost, Machine learning, Ketidakseimbangan kelas*

Abstract

This study compares the performance of XGBoost and CatBoost for scholarship recipient classification in Islamic boarding schools (pesantren). The models generate recommendations for three scholarship categories: A (Qur'an Memorization Excellence), B (Academic and Achievement), and C (Economic Need). These recommendations are intended to support decision-making by administrative staff and pesantren management, who remain responsible for the final approval. The dataset presents several challenges, including inconsistency in manual assessments, potential information leakage in several attributes, and imbalanced class distribution, particularly in Category B. To address these issues, the study applies data cleaning, removal of leakage-prone features, engineering of seven derived features, and class weighting to strengthen learning on minority classes. Model performance is evaluated using 10-fold stratified

cross-validation. The results show that XGBoost provides the most stable performance, achieving 86.42% accuracy and more consistent F1-scores across most classes compared with CatBoost. The engineered features particularly improve classification in Categories A and C, while Category B remains challenging due to its limited sample size. The findings indicate that machine learning-based classification can reduce subjectivity and improve efficiency in scholarship screening, where admin staff input student data and the model produces recommendations for review by pesantren management. The proposed approach can be replicated in similar educational environments.

Keywords: *Scholarship classification, XGBoost, CatBoost, Machine learning, Class imbalance*

1. PENDAHULUAN

Distribusi beasiswa pendidikan di Indonesia terus berkembang seiring meningkatnya kebutuhan siswa untuk mengakses layanan pendidikan yang lebih merata [1], [2]. Arus ini tampak jelas pada lembaga pendidikan berbasis pesantren yang mengelola ribuan santri dengan latar sosial ekonomi beragam [3]. Kompleksitas pengelolaan beasiswa di lingkungan pesantren meningkat karena setiap santri membawa kombinasi karakter akademik, nonakademik, dan kondisi keluarga yang berbeda [4], [5]. Situasi ini menciptakan kebutuhan sistem seleksi yang benar-benar objektif, cepat, dan mampu menangkap variasi karakter santri secara lebih presisi [6]. Pondok pesantren memegang peran penting bagi mobilitas pendidikan banyak pelajar dari keluarga menengah ke bawah [7]. Fenomena relevan yang muncul ialah bertambahnya jumlah pengajuan beasiswa yang tidak diimbangi dengan mekanisme penilaian yang efisien [8], [9]. Banyak pesantren masih mengandalkan penilaian manual yang rawan bias, lambat, dan menguras sumber daya [10]. Beban administratif meningkat karena setiap pengurus perlu membaca ulang data santri, memeriksa riwayat akademik maupun nonakademik, lalu menilai kelayakan secara subjektif [11]. Situasi ini menurunkan akurasi penentuan penerima beasiswa dan memperlambat pengambilan keputusan [12].

Ekosistem pengelolaan beasiswa di pesantren melibatkan beberapa pihak berkepentingan. Pengurus pesantren bertindak sebagai pengambil keputusan utama dalam penetapan penerima beasiswa, sedangkan admin atau petugas akademik berperan dalam pengumpulan, validasi, dan pengelolaan data santri. Unit keuangan bertanggung jawab menyalurkan dana sesuai hasil seleksi yang ditetapkan, sementara wali santri menerima dampak langsung dari bantuan yang diberikan. Pada akhirnya, santri menjadi pihak utama yang merasakan manfaat program. Ketepatan hasil seleksi sangat memengaruhi persebaran dana, kualitas tata kelola, rasa keadilan di lingkungan pesantren, serta keberlanjutan program beasiswa itu sendiri.
Secara umum, proses seleksi beasiswa di pesantren masih berjalan melalui tahapan pengajuan berkas oleh santri atau wali santri, dilanjutkan dengan verifikasi administrasi oleh admin, kemudian penilaian manual oleh pengurus terhadap aspek hafalan, prestasi, nilai akademik, aktivitas keagamaan, dan kondisi ekonomi keluarga. Keputusan akhir biasanya ditentukan melalui rapat internal, lalu diserahkan kepada unit keuangan untuk proses pencairan dana. Mekanisme ini masih sangat bergantung pada subjektivitas penilai dan belum sepenuhnya terdokumentasi secara kuantitatif. Kondisi ini menyebabkan konsistensi keputusan antarperiode seleksi sulit dijaga dan risiko bias tetap tinggi.

Penelitian terdahulu menunjukkan adanya potensi kuat pemanfaatan data dan algoritma untuk mengatasi masalah seleksi beasiswa. Studi pertama berhasil mengembangkan model prediksi berbasis pembelajaran mesin dengan akurasi tertinggi pada Logistic Regression sebesar 62,29% dan AUC 0,818 yang memberikan indikator peluang penerimaan beasiswa secara kuantitatif [13]. Studi kedua menegaskan adanya ketidakefisienan dalam pengelolaan beasiswa manual di tingkat sekolah karena proses panjang dan risiko kesalahan penilaian [14]. Dua hasil ini memberi gambaran jelas bahwa pendekatan berbasis data mampu meningkatkan konsistensi keputusan serta mempercepat proses seleksi.

Pemilihan metode prediktif dalam penelitian ini mengarah pada penggunaan XGBoost dan CatBoost karena keduanya terbukti andal ketika bekerja dengan data tabular yang berisi campuran fitur numerik dan kategorikal [15], [16]. XGBoost dikenal efisien dalam membangun

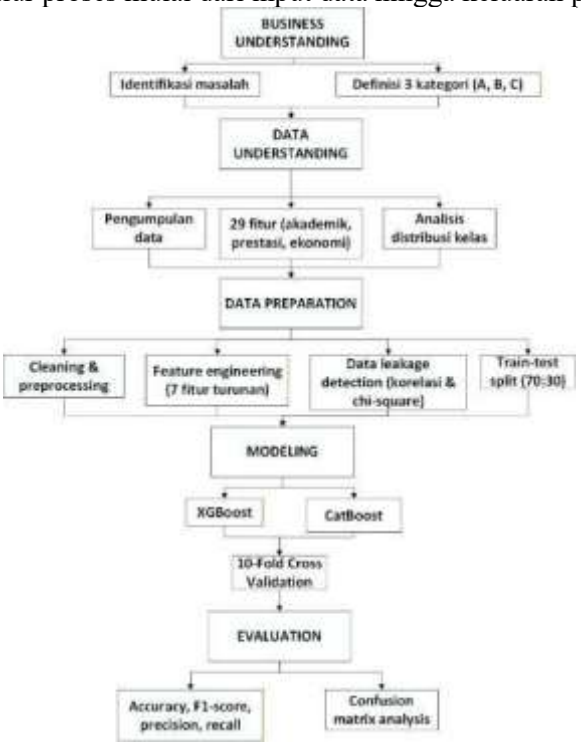
model berbasis pohon yang mampu menangkap pola kompleks melalui proses boosting yang terkontrol [17]. CatBoost memberikan keunggulan lain dengan kemampuannya mengolah variabel kategorikal secara langsung tanpa proses encoding tambahan sehingga risiko degradasi informasi dapat ditekan. Dua algoritma ini memberi ruang bagi proses pemodelan yang lebih konsisten dan stabil, terutama ketika jumlah data bertambah dan variasi karakter santri makin luas [18].

Dari segi kontribusi, penelitian ini tidak hanya menerapkan algoritma yang sudah ada, tetapi juga mengadaptasikannya pada kekhasan data pesantren. Pertama, penelitian ini merancang sistem pelabelan yang menempatkan aspek tahfidz dan nilai-nilai pesantren sebagai komponen utama, sehingga menghasilkan skema seleksi yang lebih selaras dengan kebutuhan institusi berbasis pendidikan karakter. Kedua, penelitian ini mengembangkan strategi rekayasa fitur yang dirancang untuk mengurangi risiko data *leakage* dan menjaga ketepatan evaluasi model. Ketiga, penelitian ini melakukan perbandingan terstruktur terhadap performa dua algoritma yang banyak digunakan pada domain pendidikan karakter, sehingga menghasilkan rekomendasi yang dapat diadopsi oleh lembaga dengan karakteristik serupa.

2. METODOLOGI

2.1 Kerangka Kerja Penelitian

Penelitian ini menerapkan tahapan eksperimen pemodelan machine learning untuk membangun model klasifikasi penerima beasiswa. Proses yang dilakukan mencakup persiapan data, rekayasa fitur, pelatihan model XGBoost dan CatBoost, serta evaluasi kinerja menggunakan skema validasi terstruktur. Alur proses pemodelan ditunjukkan pada Gambar 1, yang menggambarkan alur proses mulai dari input data hingga keluaran prediksi model.



Gambar 1. Diagram alir penelitian

Proses dimulai dari *Problem Definition*, yaitu penetapan tujuan untuk membangun sistem klasifikasi yang mengelompokkan siswa ke dalam tiga kategori beasiswa: A (Tahfidz Unggulan), B (Akademik Prestasi), dan C (Kebutuhan Ekonomi). Tahapan ini menetapkan ruang lingkup, batasan, serta fokus variabel yang akan digunakan pada pemodelan.

Tahap berikutnya adalah *Data Acquisition*, yang dilakukan dengan mengumpulkan data dari 310 siswa beserta 29 atribut yang menggambarkan aspek akademik, prestasi nonakademik,

kepemimpinan, aktivitas organisasi, hafalan Al-Qur'an, dan kondisi ekonomi keluarga.

2.2 Persiapan dan Pembersihan Data

Langkah berikutnya adalah *Preparation*, yang meliputi pembersihan dataset, Standarisasi nilai, dan penanganan atribut yang tidak konsisten. Seluruh nilai akademik dinormalisasi ke skala 0–100 untuk menjaga homogenitas rentang dan memudahkan pemodelan berbasis pohon.

Proses dilanjutkan dengan *Feature Engineering*, yaitu pengembangan tujuh fitur turunan berdasarkan agregasi prestasi, tingkat partisipasi organisasi, interaksi antar atribut akademik–kepemimpinan, serta indikator biner untuk mengidentifikasi siswa berprofil unggul. Tujuan dari rekayasa fitur ini adalah menambahkan informasi yang tidak tampak langsung pada fitur asli agar model mampu menangkap pola kombinatif secara lebih presisi.

Tabel 1. Strategi Rekayasa Fitur untuk Kompensasi Data Leakage

No.	Fitur Rekayasa	Sumber Data	Kompensasi untuk Leakage
1.	Total_Prestasi	4 jenis prestasi (akademik & non-akademik)	-
2.	Total_Organisasi	4 organisasi ekstrakurikuler	Menggantikan Keterlibatan Organisasi Kesenian
3.	Prestasi x Organisasi	Total_Prestasi × Total_Organisasi	Mendukung kompensasi organisasi kesenian
4.	Nilai × Kepemimpinan	Nilai Akademik × Skor Kepemimpinan	-
5.	Academic_Score	Nilai Akademik × (Total_Prestasi + 1)	Menggantikan standar tinggi dari Kefasihan Bahasa Asing
6.	Leadership_Index	Kepemimpinan × (Total_Organisasi + 1)	Menggantikan Aktivitas di Pesantren
7.	High_Achiever	Nilai ≥80 dan Prestasi ≥2	Mengkompensasi kriteria unggul dari Kefasihan Bahasa Asing

Rekayasa fitur menghasilkan tujuh fitur turunan dengan dua tujuan utama: (1) menangkap pola kombinatif antar variabel, dan (2) mengkompensasi informasi yang hilang akibat eliminasi tiga fitur *leakage*. Seperti diilustrasikan pada Tabel 1, strategi kompensasi diterapkan secara selektif berdasarkan karakteristik *leakage* setiap fitur.

Total_Organisasi menggantikan secara langsung Keterlibatan Organisasi Kesenian yang dihapus karena korelasi sempurna dengan *outcome*. Dengan mengagregasi empat jenis organisasi menjadi satu metrik, informasi partisipasi dipertahankan tanpa mengungkapkan spesifisitas organisasi kesenian yang bias. Prestasi x Organisasi mendukung kompensasi ini dengan menangkap sinergi antara prestasi dan aktivitas organisasi.

Leadership_Index mengambil alih peran Aktivitas di Pesantren yang menyebabkan *leakage* deterministik terhadap label "A". Fitur ini mempertahankan esensi pengukuran kepemimpinan namun menghilangkan unsur-unsur spesifik yang berpotensi menimbulkan bias, dengan mengintegrasikan skor kepemimpinan dan pengalaman organisasi.

Academic_Score dan High_Achiever secara kolektif mengkompensasi Kefasihan Bahasa Asing yang berfungsi sebagai *proxy leakage* untuk kompetensi unggul. Academic_Score menciptakan skor akademik kontekstual yang memperhitungkan prestasi tambahan, sementara High_Achiever menetapkan *threshold* biner untuk identifikasi siswa berprestasi tinggi. Kedua fitur ini mempertahankan kemampuan prediktif kompetensi unggul tanpa menyebabkan *leakage*.

Empat fitur lainnya (Total_Prestasi, Nilai x Kepemimpinan, serta dua fitur pendukung kompensasi) berfungsi memperkaya representasi data dengan agregasi, interaksi, dan transformasi yang meningkatkan kemampuan model menangkap pola kompleks

2.3 Pelabelan Data

Dataset terdiri dari 310 sampel siswa pesantren dengan 29 atribut awal yang dikategorikan menjadi empat domain utama: (1) demografi, (2) kompetensi akademik, (3) keterampilan non-akademik, dan (4) kondisi sosio-ekonomi. Setelah *preprocessing*, dataset berkembang menjadi 36 fitur termasuk 7 fitur turunan hasil rekayasa. Sebanyak 41 sampel (13,2%) tidak memenuhi kriteria pelabelan dan dieksklusi dari proses training, menghasilkan 269 sampel valid.

Proses pelabelan menghasilkan distribusi kelas yang tidak seimbang (*imbalanced class distribution*), seperti ditunjukkan pada Tabel 2.

Tabel 2. Distribusi Kelas Beasiswa

Kelas	Deskripsi	Jumlah Siswa	Persentase
A	Tahfidz Unggulan	138	44,5%
B	Akademik & Prestasi	19	6,1%
C	Kebutuhan Ekonomi	112	36,1%
Tidak terlabel	Tidak memenuhi kriteria	41	13,2%

Ketidakseimbangan signifikan pada Kelas B (hanya 19 sampel) ditangani dengan teknik *class weighting* selama pelatihan model. Pelabelan dilakukan secara berjenjang menggunakan aturan berbasis domain (*rule- based labeling*) dengan prioritas A → B → C. Kriteria setiap kelas disajikan pada Tabel 3.

Tabel 3. Kriteria Pelabelan Beasiswa

Kelas	Kriteria Utama	Kondisi Tambahan
A	1. Hafalan ≥ 3 juz 2. Bahasa asing FASIH 3. Aktif di pesantren	Semua kondisi harus terpenuhi
B	1. Nilai akademik ≥ 75 2. Rencana lanjut kuliah	Plus salah satu: 1. Prestasi ≥ 3 2. Organisasi ≥ 1 3. Kepemimpinan ≥ 2
C	1. Motivasi pendidikan TINGGI	Plus salah satu: 1. Kondisi ekonomi TIDAK MAMPU 2. Sering menunggak pembayaran

2.4 Pemeriksaan dan Pengendalian Leakage

Sebelum memasuki tahap pemodelan, dilakukan langkah *Label Leakage Check*. Pemeriksaan dilakukan dengan mengevaluasi hubungan antara masing-masing fitur terhadap label menggunakan koefisien korelasi untuk atribut numerik dan uji chi-square untuk atribut kategorikal. Tiga fitur teridentifikasi memiliki hubungan yang terlalu dominan terhadap label sehingga berpotensi menyebabkan kebocoran informasi selama training. Ketiga fitur tersebut dihapus agar model tidak belajar dari sinyal yang tidak seharusnya.

2.5 Pembagian Data

Setelah data dinyatakan bersih dan bebas dari potensi *leakage*, proses dilanjutkan pada tahap pembagian data (*train–test* split). Dari 310 sampel awal, terdapat 269 sampel yang berhasil dilabeli sesuai kriteria yang ditetapkan. Pembagian dilakukan menggunakan *stratified sampling* dengan rasio 70:30 untuk mempertahankan proporsi setiap kelas pada data *training* dan *testing*. Hasilnya, 188 sampel dialokasikan sebagai data *training* dan 81 sampel sebagai data *testing*. Pemisahan ini membuat proses pengujian benar-benar menggunakan data yang belum pernah dilihat model sebelumnya. Dengan cara ini, penilaian performa tidak tercampur oleh informasi dari tahap pelatihan, sehingga hasil evaluasi lebih mencerminkan kemampuan model ketika menghadapi data baru yang muncul di penggunaan sebenarnya.

2.6 Pengembangan Model

Tahapan berikutnya adalah Model Training, yang dilakukan terhadap dua algoritma utama: XGBoost dan CatBoost. Keduanya dipilih karena kemampuan adaptifnya terhadap data tabular dengan fitur numerik dan kategorikal.

Konfigurasi XGBoost menggunakan 1500 *trees*, *depth* maksimal 8, *learning rate* 0.02, serta *sample weighting* untuk menstabilkan pembelajaran pada kelas minoritas. CatBoost dilatih menggunakan 1500 iterasi, *depth* 7, *learning rate* 0.02, dan penanganan *native* pada fitur kategorikal. Strategi pembobotan kelas diterapkan pada kedua model untuk mengurangi dominasi Kelas A dan C, sekaligus memberi penekanan pada Kelas B yang memiliki jumlah sampel kecil.

2.7 Validasi Model

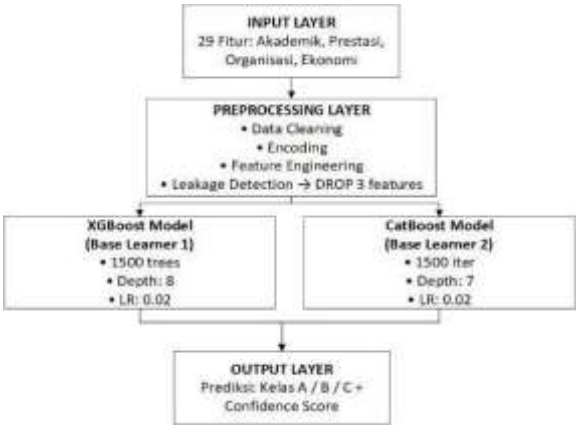
Langkah selanjutnya adalah *Cross-Validation* untuk memperoleh estimasi performa yang lebih stabil. Penelitian ini menggunakan 10-*fold stratified cross-validation* agar komposisi kelas tetap seimbang di setiap *fold*. Metode ini memberikan gambaran ketahanan model terhadap variasi subset data dan membantu memprediksi seberapa baik performa model di dunia nyata.

2.8 Evaluasi Model

Tahapan terakhir adalah *Model Evaluation.* Pengujian dilakukan pada data testing menggunakan empat metrik utama: *accuracy*, *precision*, *recall*, dan *F1-score.* Seluruh metrik ini digunakan untuk membandingkan kemampuan kedua algoritma dalam mengenali tiap kategori beasiswa. Hasil evaluasi ini digunakan untuk menentukan model dengan kinerja terbaik yang direkomendasikan sebagai basis pengembangan sistem pendukung keputusan seleksi beasiswa.

2.9 Arsitektur Model

Gambar 2 menampilkan arsitektur model yang digunakan dalam penelitian, mulai dari tahap *input* hingga proses keluaran prediksi. Struktur ini dirancang agar seluruh proses mulai dari pengolahan fitur, deteksi *leakage*, hingga pelatihan dua *base learner* berjalan sebagai satu pipeline yang konsisten. Setiap komponen pada diagram menunjukkan bagaimana data bergerak melalui tahapan *preprocessing* hingga diolah oleh XGBoost dan CatBoost untuk menghasilkan prediksi akhir.



Gambar 2. Arsitektur Model Klasifikasi / Machine Learning Pipeline untuk Seleksi Beasiswa

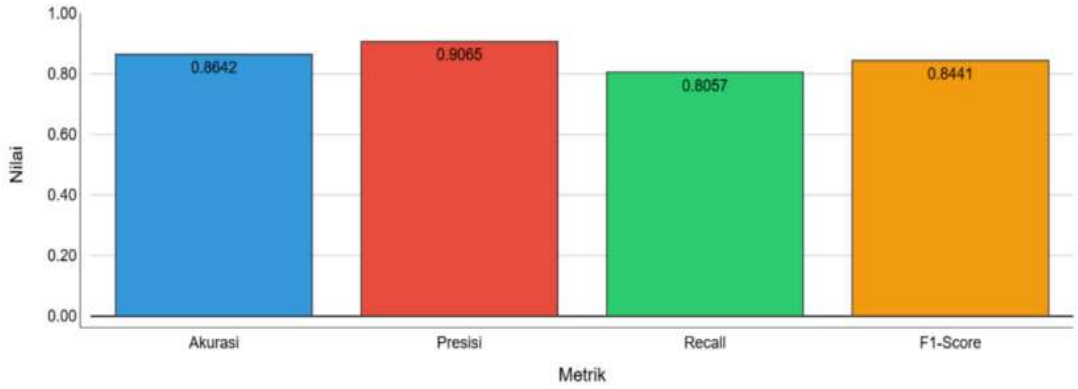
Arsitektur model dirancang dalam empat *layer* berurutan. *Input Layer* menerima 29 fitur siswa yang dikategorikan menjadi numerik (10 fitur: usia, tingkat pendidikan, nilai akademik, kepemimpinan, empat jenis prestasi, jumlah hafalan, dan kegiatan hafalan) serta kategorikal (19 fitur: jenis kelamin, status keluarga, kondisi ekonomi, keterlibatan organisasi, rencana masa depan, dan lainnya). *Preprocessing Layer* melakukan transformasi data melalui pembersihan nilai hilang, enkoding ordinal untuk fitur kategorikal, pembuatan tujuh fitur *engineered*, dan eliminasi tiga fitur berkorelasi tinggi hasil deteksi leakage menggunakan uji korelasi Pearson (numerik) dan chi-square (kategorikal) dengan threshold p-value <0.001. *Model Layer* terdiri dari dua base learner: XGBoost dengan konfigurasi 1500 estimator, *depth* maksimal 8, *learning rate* 0.02, dan bobot kelas {A:1, B:5, C:2.5}, serta CatBoost dengan 1500 iterasi, *depth* 7, *learning rate* 0.02,

dan penanganan native untuk fitur kategorikal. *Output Layer* menghasilkan prediksi kelas final (A/B/C) beserta skor kepercayaan probabilistik untuk setiap kelas.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan evaluasi menyeluruh terhadap model yang dikembangkan. Analisis dimulai dari pengukuran performa pada data testing untuk menilai kemampuan generalisasi model terhadap sampel yang tidak terlihat selama pelatihan. Hasil ini menjadi dasar bagi pembahasan lanjutan terkait konsistensi model, stabilitas melalui validasi silang, serta interpretasi performa pada masing-masing kelas beasiswa.

3.1 Performa Model pada Data Testing

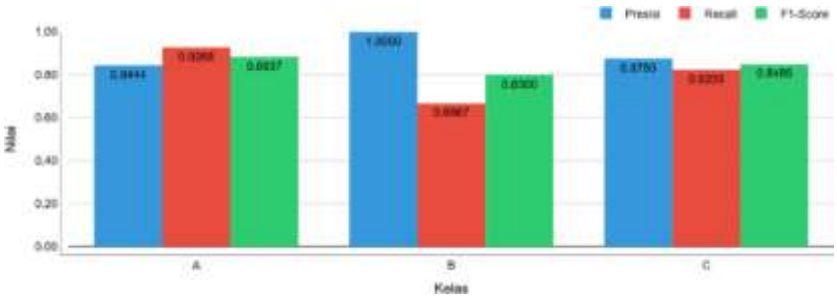


Gambar 3. Diagram Performa Model XGBoost

Pengujian menggunakan 81 sampel data *testing* menunjukkan bahwa XGBoost memberikan performa kuat dan relatif stabil. Diagram pada Gambar 3 memperlihatkan akurasi 86.42% dengan *F1-score* 0.8441, *precision* 0.9065, dan *recall* 0.8057. Radar chart menunjukkan seluruh metrik berada di atas 0.80 sehingga model memiliki keseimbangan yang baik antara kemampuan mengenali kelas positif dan menjaga tingkat kesalahan.

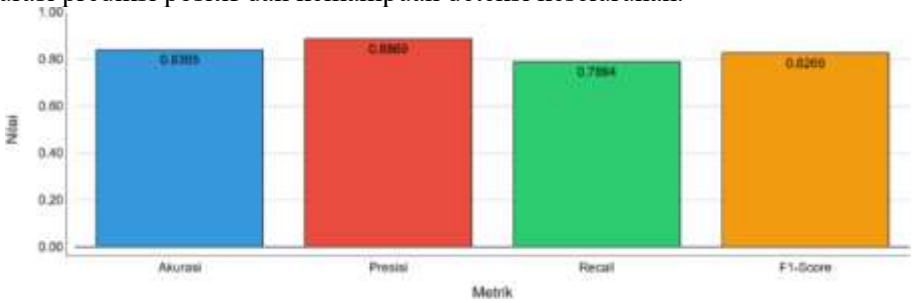
Gambar 4. Confusion Matrix Model XGBoost

Confusion matrix pada Gambar 4 memvisualisasikan performa klasifikasi model XGBoost dalam memprediksi tiga kategori beasiswa (A, B, C). Model menunjukkan kemampuan terbaik dalam mengklasifikasikan kategori A dengan 38 prediksi benar, sementara kategori C memperoleh 14 prediksi benar dengan 3 kesalahan sebagai kategori A. Kategori B mengalami tantangan terbesar dengan hanya 1 prediksi benar, mengindikasikan kecenderungan model untuk lebih sering memprediksi kategori B sebagai kategori lainnya. Distribusi ini merefleksikan ketidakseimbangan data atau kompleksitas pola dalam kategori B yang memerlukan perhatian lebih dalam optimasi model.



Gambar 5. Metrik per Kelas Model XGBoost

Gambar 5 menyajikan analisis *precision*, *recall*, dan *F1-score* untuk masing-masing kategori. Kategori A mencapai performa terbaik dengan *F1-score* 0.8837, didukung oleh *recall* tertinggi (0.9268) yang mengindikasikan kemampuan model dalam mendeteksi sebagian besar sampel kategori A. Kategori C menunjukkan keseimbangan antara *precision* (0.8750) dan *recall* (0.8235) dengan *F1-score* 0.8485. Kategori B memperoleh presisi sempurna (1.0000) namun *recall* terendah (0.6667), menghasilkan *F1-score* 0.8000 yang merepresentasikan *trade-off* antara akurasi prediksi positif dan kemampuan deteksi keseluruhan.

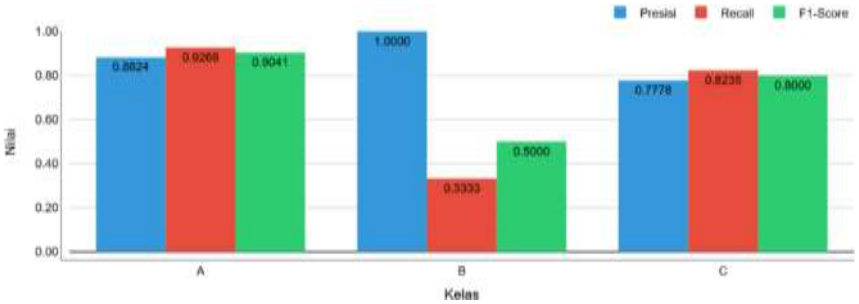


Gambar 6. Diagram Performa Model CatBoost

Evaluasi performa keseluruhan Model CatBoost dalam Gambar 6 menghasilkan akurasi 83,95% dengan *F1-score* 0,8269. Model ini mencapai presisi makro tertinggi (0,8869) namun mengalami penurunan *recall* (0,7894) dibandingkan model XGBoost. Profil ini mengindikasikan bahwa CatBoost cenderung lebih konservatif dalam memprediksi kategori positif, sehingga mengurangi *false positive* namun berpotensi meningkatkan *false negative*. Perbedaan distribusi metrik ini menyarankan karakteristik algoritmik yang berbeda dalam menangani *trade-off* antara *precision* dan *recall*.

Gambar 7. Confusion Matrix Model CatBoost

Confusion matrix Model CatBoost pada Gambar 7 menunjukkan pola yang serupa dengan XGBoost dalam klasifikasi kategori A dengan 38 prediksi benar, namun mengalami peningkatan kesalahan pada kategori B menjadi 2 prediksi salah sebagai kategori lain. Kategori C mempertahankan konsistensi dengan 14 prediksi benar. Peningkatan kesalahan pada kategori B mengindikasikan bahwa model CatBoost mengalami kesulitan yang lebih signifikan dalam membedakan pola karakteristik kategori B dibandingkan model XGBoost. Hal ini merefleksikan perbedaan sensitivitas kedua algoritma terhadap karakteristik data kategori minoritas.



Gambar 8. Metrik per Kelas Model CatBoost

Gambar 8 menunjukkan analisis per kategori menunjukkan bahwa CatBoost mencapai *precision* sempurna (1,0000) untuk kategori B namun dengan *recall* terendah (0,3333), menghasilkan *F1-score* 0,5000 yang signifikan lebih rendah dibandingkan XGBoost. Kategori A mempertahankan performa kuat dengan *F1-score* 0,9041, sementara kategori C menunjukkan *F1-score* 0,8000. Pola ini mengonfirmasi bahwa meskipun CatBoost mampu menghindari *false positive* pada kategori B, algoritma ini kurang efektif dalam mendeteksi keseluruhan sampel kategori B. Temuan ini menggarisbawahi pentingnya penyesuaian *threshold* klasifikasi untuk kategori minoritas.

3.2 Cross Validation dan Stabilitas Model

Hasil *cross-validation* menegaskan perbedaan karakter antara dua algoritma. XGBoost mencatat rata-rata akurasi 0.7807 ± 0.1051 dan *F1-score* 0.7987 ± 0.1059. Deviasi metrik yang relatif rendah menunjukkan performa yang stabil antar-fold. CatBoost mencatat akurasi 0.7173 ± 0.1027 dan *F1-score* 0.6925 ± 0.1608, deviasi *F1-score* yang lebih besar memperlihatkan variabilitas performa yang lebih tinggi.

Perbandingan antara hasil *cross-validation* dan testing memberikan gambaran pola generalisasi. XGBoost menunjukkan peningkatan moderat dari CV ke *testing*, sedangkan CatBoost mengalami lonjakan yang lebih besar. Perbedaan ini mencerminkan sifat CatBoost yang cenderung menghasilkan estimasi konservatif pada data *training* yang dipecah menjadi beberapa *fold*, lalu meningkat ketika dihadapkan pada data *testing* yang tidak terlalu bising. Sementara itu, XGBoost dengan regularisasi yang lebih agresif cenderung menjaga keseimbangan sehingga performanya lebih konsisten di dua skenario evaluasi.

3.3 Implikasi Model terhadap Sistem Informasi Seleksi Beasiswa

Penerapan model klasifikasi ini berpotensi mengurangi tingkat subjektivitas dalam proses seleksi beasiswa. Selama ini keputusan kelayakan sangat bergantung pada penilaian manual pengurus yang tidak selalu konsisten antarperiode seleksi. Dengan memanfaatkan model pembelajaran mesin yang dilatih pada data historis, proses penilaian dilakukan berdasarkan pola numerik dan indikator yang terukur. Model ini tidak dimaksudkan untuk menggantikan keputusan pengurus, namun berfungsi sebagai alat bantu untuk memberikan rekomendasi awal yang lebih objektif dan terstandar.

Model yang dikembangkan juga berpotensi mempercepat proses seleksi. Pada mekanisme manual, pengurus harus membaca dan menilai berkas setiap santri secara individual, sehingga proses administrasi menjadi panjang dan melelahkan. Dengan adanya model klasifikasi, data santri yang telah direkap oleh admin dapat langsung diproses untuk menghasilkan rekomendasi kategori beasiswa. Pengurus hanya perlu melakukan verifikasi akhir terhadap rekomendasi tersebut, sehingga waktu yang dibutuhkan untuk proses seleksi dapat berkurang secara signifikan.

Dari sisi pengembangan sistem informasi, model ini memiliki peluang untuk diintegrasikan dengan sistem akademik pesantren apabila telah tersedia basis data nilai, hafalan, dan aktivitas santri dalam format digital. Model dapat dijalankan sebagai modul klasifikasi yang secara otomatis memberikan rekomendasi kategori beasiswa berdasarkan data pada sistem. Dengan demikian, proses evaluasi santri dapat dilakukan secara lebih terstruktur dan terdokumentasi.

Selain itu, hasil klasifikasi juga dapat dihubungkan dengan sistem keuangan pesantren untuk keperluan penyaluran dana. Setiap kategori beasiswa yang dihasilkan model dapat menjadi

acuan nominal bantuan yang diterima santri, sehingga proses distribusi dana menjadi lebih transparan, terukur, dan memiliki jejak data yang jelas untuk keperluan audit serta pelaporan.

3.4 Analisis Per Kelas dan Implikasi Praktis

Diagram metrik per kelas pada Gambar 5 dan 8 memperlihatkan bahwa tiap kategori memiliki pola yang berbeda. Kelas A menunjukkan *recall* tinggi pada XGBoost (0.9268), menandakan kemampuan model dalam menangkap sebagian besar kandidat tahfidz unggulan. Tingginya *recall* menjadi penting agar kandidat potensial tidak terabaikan [19]. Kelas B menunjukkan *precision* sempurna (1.0000) pada kedua model, tetapi *recall* rendah (0.6667). Pola ini wajar pada kategori dengan jumlah data sangat sedikit [20]. Model menjadi terlalu selektif untuk menekan *false positives*, sehingga beberapa contoh yang sebenarnya layak tidak terdeteksi. Kelas C menunjukkan konsistensi terbaik, baik *precision* maupun *recall* berada di kisaran 0.82 pada kedua model. Konsistensi tersebut menunjukkan bahwa fitur-fitur terkait karakter ekonomi memiliki pemisahan yang relatif jelas sehingga lebih mudah dipelajari [21]. Penggunaan *class weight* membantu distribusi performa agar tidak terlalu berat ke kategori dengan jumlah sampel dominan [22].

3.5 Analisis Hasil Pemodelan dan Keputusan Model Final

Distribusi kelas B yang hanya mencakup 19 dari 269 sampel disebabkan oleh karakteristik populasi pesantren pada penelitian ini. Kategori beasiswa B diperuntukkan bagi siswa dengan kombinasi capaian akademik tinggi (≥75), aktivitas ekstrakurikuler yang terkelola dengan baik, dan minat melanjutkan studi ke perguruan tinggi. Pada pesantren tradisional, mayoritas siswa umumnya terfokus pada dua jalur utama, yaitu penguatan tahfidz dan bahasa Arab (kelas A) serta siswa dengan prioritas kebutuhan ekonomi (kelas C). Proporsi siswa yang memenuhi seluruh kriteria kategori B menjadi lebih kecil karena struktur kurikulum pesantren memang memberikan porsi besar pada pengembangan kompetensi keagamaan. Oleh sebab itu, jumlah kelas B yang minim mencerminkan kondisi lapangan secara apa adanya, bukan akibat kesalahan dalam proses pelabelan.

Hasil perbandingan menunjukkan bahwa XGBoost muncul sebagai model dengan performa paling kompetitif. *F1-score* mencapai 0.8441, lebih tinggi dibanding CatBoost (0.8269). Perbedaan paling mencolok terlihat pada kelas A dan C, di mana XGBoost memperoleh *F1-score* yang lebih baik pada keduanya.

Kelas B tidak menunjukkan perbedaan antara dua model karena keterbatasan sampel yang memang mempengaruhi batas kemampuan pembelajaran algoritma mana pun. Namun, dominasi XGBoost pada dua kelas terbesar menjadikannya pilihan paling layak untuk implementasi akhir, terutama karena stabilitas performanya yang tercermin dari nilai deviasi *cross-validation* yang lebih rendah.

Tabel 4. Perbandingan Fitur paling Berpengaruh untuk Kedua Model

Urutan	Fitur	XGBoost (%)	CatBoost (%)	Rata-rata (%)
1	Nilai x Kepemimpinan	18.41	15.67	17.04
2	High_Achiever	14.72	16.88	15.80
3	Jumlah Juz Hafal (Tahfidz)	12.08	11.92	12.00
4	Total_Prestasi	9.83	9.14	9.49
5	Prestasi x Organisasi	8.47	10.21	9.34

Analisis *feature importance* berbasis *gain* pada model XGBoost dalam Tabel 4 menunjukkan bahwa dua fitur *engineered*, yaitu Nilai Akademik x Kepemimpinan (gain 18,4%) dan High_Achiever (*gain* 14,7%), termasuk dalam lima besar fitur paling berpengaruh, diikuti oleh Total_Prestasi dan Prestasi x Organisasi. Pola serupa juga teramati pada CatBoost, di mana High_Achiever dan interaksi Nilai x Kepemimpinan secara konsisten mendapat skor *importance* tertinggi. Keberadaan fitur-fitur ini berhasil mengkompensasi penghapusan tiga fitur *leakage* (Aktivitas di Pesantren, Keterlibatan Organisasi Kesenian, dan Kefasihan Bahasa Asing) sehingga performa model tetap tinggi (akurasi 86,42% pada XGBoost) dan *recall* Kelas A tetap mencapai 0,93 meskipun menggunakan hanya informasi yang tersedia setelah penghapusan data *leakage*.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengevaluasi model klasifikasi penerima beasiswa pada lingkungan pesantren dengan membandingkan dua algoritma machine learning berbasis boosting, yaitu XGBoost dan CatBoost. Model yang dibangun mampu mengelola permasalahan ketidakseimbangan kelas sekaligus meminimalkan potensi bias akibat kebocoran informasi melalui penerapan rekayasa fitur dan penghapusan atribut yang bersifat deterministik terhadap label.

Hasil pengujian menunjukkan bahwa XGBoost merupakan model dengan kinerja paling stabil, dengan akurasi sebesar 86,42% dan nilai F1-score yang lebih konsisten pada sebagian besar kelas dibandingkan CatBoost. Temuan ini menunjukkan bahwa model yang mampu menangkap hubungan non-linier sekaligus memanfaatkan fitur berbasis domain lebih efektif untuk memetakan karakter santri yang bersifat multidimensional. Pendekatan ini berpotensi direplikasi pada institusi pendidikan lain yang memiliki karakteristik populasi serupa.

Penelitian selanjutnya dapat diarahkan pada peningkatan representasi kelas minoritas melalui pendekatan data-centric, serta penguatan interpretabilitas model menggunakan teknik XAI seperti SHAP atau LIME. Upaya tersebut diharapkan dapat meningkatkan transparansi model dan mendukung penerapannya sebagai komponen analitik pada sistem pendukung keputusan seleksi beasiswa di masa mendatang.

5. DAFTAR RUJUKAN

[1] Y. S. H. Hidayat, "Pemodelan Data Mining Menggunakan Neural Network untuk Seleksi Mahasiswa Penerima Bantuan Beasiswa: Studi Kasus STMIK Pasim Sukabumi," *Seminalu*, vol. 1, no. 1, pp. 602–613, 2023.

[2] I. Kusumawati, "Integrasi kurikulum pesantren dalam kurikulum nasional pada pondok pesantren modern," *Sanskara Pendidikan Dan Pengajaran*, vol. 2, no. 01, pp. 1–7, 2024.

[3] I. P. Permatasari, A. F. Cobantoro, and D. Mustikasari, "Penggunaan Algoritma K-means untuk Menentukan Calon Penerima Beasiswa," *SinarFe7*, vol. 7, no. 1, pp. 524–533, 2025.

[4] S. F. Jasmine, "Pengaruh beasiswa KIP-K terhadap prestasi belajar mahasiswa Manajemen Pendidikan angkatan 2021 Universitas Negeri Surabaya," *Jurnal Pendidikan, Bahasa dan Budaya*, vol. 2, no. 2, pp. 61–70, 2023.

[5] A. Fadlan, "Pengaruh latar belakang ekonomi keluarga dan biaya pendidikan terhadap motivasi belajar peserta didik di SMA Negeri 1 Linggabayu," *Jurnal Pamator: Jurnal Ilmiah Universitas Trunojoyo*, vol. 15, no. 1, pp. 81–88, 2022.

[6] D. J. Lubis and M. B. Tamam, "Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor," *TeknoIS: Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 12, no. 1, pp. 7–20, 2022.

[7] M. Rozaidin and H. H. Adinugraha, "Penerapan Akuntansi Pondok Pesantren (Studi pada Koperasi Pondok Pesantren Al Hasyimi Kabupaten Pekalongan)," *ekonomika syariah: Journal of Economic Studies*, vol. 4, no. 2, pp. 123–135, 2020.

[8] R. Zainal, K. Joesyiana, H. Zainal, S. Wahyuni, and A. Adriyani, "Manajemen Pengelolaan Keuangan bagi Mahasiswa Penerima Beasiswa KIP Kuliah pada Perguruan Tinggi di Lingkungan Yayasan Pendidikan Persada Bunda (STIE–STISIP–STBA–STIH)," *JIPM: Jurnal Inovasi Pengabdian Masyarakat*, vol. 1, no. 1, pp. 1–5, 2023.

[9] L. Liesnaningsih, R. Taufiq, R. Destriana, and A. P. Suyitno, "Sistem Pendukung Keputusan Penerima Beasiswa Berbasis WEB Menggunakan Metode Simple Additive Weighting (SAW) pada Pondok Pesantren Daarul Ahsan," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 1, pp. 54–60, 2020.

[10] A. Aliansah, "penerapan metode profile matching untuk merekomendasikan penerima beasiswa di pondok pesantren tahfidz al-qur'an/Andy Aliansah/14187053/SISTEM INFORMASI/Pembimbing I: Irmayansyah/Pembimbing II: Wahyu Hidayat," 2022.

[11] R. N. S. Syifa, A. Wibowo, E. Marsusanti, N. Purwati, and R. Riniawati, "Sistem

Pendukung Keputusan Pemilihan Calon Penerima Beasiswa Tahfidz Menggunakan Metode SAW," *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, vol. 5, no. 1, pp. 19–26, 2022.

[12] Z. Maulana and A. Rofiq, "mengoptimalkan beasiswa pendidikan berkelanjutan di sekolah tinggi pesantren: MENGUBAH PARADIGMA KESEJAHTERAAN DOSEN," *Benchmarking*, vol. 8, no. 1, pp. 43–49, 2024.

[13] G. Budiyanto and D. Suryadi, "Model Prediksi dengan Pembelajaran Mesin dalam Pemberian Program Beasiswa kepada Calon Mahasiswa Baru Program S1 di Perguruan Tinggi Swasta.," *Jurnal Rekayasa Sistem Industri*, vol. 12, no. 2, pp. 187–200, 2023.

[14] S. Salmia and E. T. Siregar, "Seleksi Siswa Yang Memperoleh Beasiswa Pada SMK 12 Saentis menggunakan Metode SMART," in *SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, 2023, pp. 276–289.

[15] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J Big Data*, vol. 7, no. 1, p. 94, 2020.

[16] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," *Int J Distrib Sens Netw*, vol. 18, no. 6, p. 15501329221106936, 2022.

[17] J. Chen, F. Zhao, Y. Sun, and Y. Yin, "Improved XGBoost model based on genetic algorithm," *International Journal of Computer Applications in Technology*, vol. 62, no. 3, pp. 240–245, 2020.

[18] J. Hancock and T. M. Khoshgoftaar, "Performance of catboost and xgboost in medicare fraud detection," in *2020 19th IEEE international conference on machine learning and applications (ICMLA)*, IEEE, 2020, pp. 572–579.

[19] T. Prexawanprasut and T. Banditwattanawong, "Improving minority class recall through a novel cluster-based oversampling technique," in *Informatics*, MDPI, 2024, p. 35.

[20] C. K. I. Williams, "The effect of class imbalance on precision-recall curves," *Neural Comput*, vol. 33, no. 4, pp. 853–857, 2021.

[21] L. Xue, X. Zhang, W. Jiang, K. Huo, and Q. Shen, "A classification performance evaluation measure considering data separability," in *International Conference on Artificial Neural Networks*, Springer, 2023, pp. 1–13.

[22] B. Bakirarar and A. H. Elhan, "Class weighting technique to deal with imbalanced class problem in machine learning: Methodological research," *Türkiye Klinikleri Biyoistatistik*, vol. 15, no. 1, pp. 19–29, 2023.