



Pengaruh Hyperparameter Tuning Gradient Boosting Terhadap Prediksi Pemilihan Program Studi Mahasiswa Baru

Harminto Mulyo*, Akhmad Khanif Zyen

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Nahdlatul Ulama, Jepara, Indonesia

Email: ^{1,*}minto@unisnu.ac.id, ²khanif.zyen@unisnu.ac.id

Email Penulis Korespondensi: minto@unisnu.ac.id

Abstrak—Penelitian ini bertujuan untuk meningkatkan akurasi prediksi pemilihan program studi mahasiswa baru dengan menggunakan *algoritma Gradient Boosting* yang dioptimalkan melalui *hyperparameter tuning*. *Gradient Boosting* dipilih karena kemampuannya dalam menangani data kompleks dan beragam, yang penting dalam konteks prediksi program studi. Data yang digunakan berasal dari *database* penerimaan mahasiswa baru Universitas Islam Nahdlatul Ulama Jepara periode 2013–2023, dengan *preprocessing* yang mencakup pembersihan data, imputasi nilai kosong, dan transformasi fitur kategorikal. Akurasi awal model *Gradient Boosting* dengan konfigurasi *default* mencapai 99,01%, yang menunjukkan bahwa dataset memiliki pola yang relatif jelas dan terstruktur, sehingga model dasar mampu memberikan performa tinggi. Namun, untuk memastikan generalisasi dan menghindari risiko *overfitting*, dilakukan *hyperparameter tuning* menggunakan *Randomized Search CV*. Hasil tuning menunjukkan peningkatan akurasi menjadi 99,84% dengan konfigurasi optimal meliputi learning rate sebesar 0,1, jumlah estimators sebanyak 300, dan kedalaman maksimum pohon 4. Analisis fitur juga mengungkap bahwa atribut seperti "jenis_sekolah", "asal_sekolah", dan "kelamin" memiliki pengaruh besar terhadap hasil prediksi. Penelitian ini membuktikan bahwa tuning hyperparameter dapat meningkatkan performa model secara signifikan, memberikan solusi prediktif yang lebih akurat dan relevan untuk proses seleksi program studi. Meskipun demikian, keterbatasan penelitian terletak pada lingkup dataset yang berasal dari satu institusi, sehingga eksplorasi lebih lanjut dengan data yang lebih beragam dan metode tuning canggih seperti *Bayesian Optimization* sangat disarankan. Temuan ini memberikan kontribusi nyata bagi institusi pendidikan dalam mengembangkan sistem berbasis data untuk mendukung pengambilan keputusan strategis.

Kata Kunci: Gradient Boosting; tuning hyperparameter; prediksi pemilihan program studi; penerimaan mahasiswa baru; optimasi model.

Abstract—This study aims to improve the accuracy of predicting new student major selection using the Gradient Boosting algorithm optimized through hyperparameter tuning. Gradient Boosting was chosen for its ability to handle complex and diverse data, which is crucial in the context of major prediction. The data used was sourced from the new student admissions database of Universitas Islam Nahdlatul Ulama Jepara for the 2013–2023 period, with preprocessing including data cleaning, imputation of missing values, and transformation of categorical features. The initial accuracy of the Gradient Boosting model with default configuration reached 99.01%, indicating that the dataset had relatively clear and structured patterns, enabling the baseline model to perform highly. However, to ensure generalization and avoid the risk of overfitting, hyperparameter tuning was performed using Randomized Search CV. The tuning results showed an increase in accuracy to 99.84% with optimal configurations including a learning rate of 0.1, 300 estimators, and a maximum tree depth of 4. Feature analysis also revealed that attributes such as "school_type," "school_origin," and "gender" significantly influenced the prediction outcomes. This study demonstrates that hyperparameter tuning can significantly enhance model performance, providing a more accurate and relevant predictive solution for the major selection process. Nevertheless, the study's limitation lies in the scope of the dataset, which originated from a single institution, suggesting the need for further exploration with more diverse data and advanced tuning methods like Bayesian Optimization. These findings provide valuable contributions to educational institutions in developing data-driven systems to support strategic decision-making.

Keywords: Gradient Boosting; hyperparameter tuning; program selection prediction; student admission; model optimization.

1. PENDAHULUAN

Pemilihan program studi mahasiswa baru merupakan keputusan penting yang memengaruhi perjalanan pendidikan tinggi seseorang. Keputusan ini tidak hanya menentukan arah akademik mahasiswa tetapi juga memiliki dampak jangka panjang terhadap kesuksesan mereka dalam dunia karir dan kehidupan profesional. Salah satu tantangan utama dalam proses ini adalah bagaimana memastikan bahwa calon mahasiswa memilih program studi yang sesuai dengan minat [1], bakat, dan potensi akademiknya. Pemilihan yang tidak sesuai sering kali berujung pada masalah serius, seperti penurunan motivasi belajar, hasil akademik yang buruk, hingga risiko putus sekolah yang lebih tinggi. Sebaliknya, pemilihan program studi yang sesuai dapat meningkatkan kemungkinan kesuksesan akademik dan karir secara signifikan [2] [3].

Dalam beberapa tahun terakhir, perkembangan teknologi dan ketersediaan data yang lebih besar telah membuka peluang untuk mengembangkan sistem yang lebih efektif dalam membantu calon mahasiswa menentukan pilihan program studi. Pendekatan berbasis data memungkinkan universitas dan lembaga pendidikan tinggi untuk menyediakan rekomendasi yang lebih akurat berdasarkan analisis historis dan prediktif. Beberapa penelitian sebelumnya telah menggunakan algoritma *machine learning* seperti *Random Forest* untuk memprediksi preferensi program studi berdasarkan data demografi dan hasil tes masuk [3]. Meskipun hasilnya cukup akurat, penelitian tersebut tidak melakukan *tuning hyperparameter* secara mendalam, sehingga potensi optimalisasi model belum sepenuhnya dieksplorasi.

Gradient Boosting adalah salah satu *algoritma ensemble learning* berbasis *boosting* yang mampu meningkatkan akurasi prediksi dengan memanfaatkan kesalahan prediksi dari model sebelumnya. Algoritma ini bekerja dengan membangun model secara bertahap, di mana setiap model baru berusaha memperbaiki kesalahan dari model sebelumnya. Dalam berbagai penelitian, *Gradient Boosting* telah terbukti menjadi salah satu algoritma terbaik untuk tugas prediksi berbasis data yang kompleks [4][5]. Namun, keunggulan *Gradient Boosting* sangat bergantung pada pengaturan



hyperparameter yang optimal. Proses *tuning hyperparameter* adalah langkah penting untuk memastikan bahwa model dapat memberikan performa terbaiknya. *Hyperparameter* seperti *learning rate*, jumlah *estimators*, dan kedalaman maksimum pohon keputusan (*max depth*) memiliki pengaruh besar terhadap hasil prediksi. Pengaturan yang salah dapat menyebabkan model menjadi *overfitting* atau *underfitting*, yang pada akhirnya akan menurunkan akurasi prediksi [6][7].

Penelitian ini bertujuan untuk mengeksplorasi pengaruh *hyperparameter tuning* terhadap performa algoritma *Gradient Boosting* dalam memprediksi pemilihan program studi mahasiswa baru. Secara khusus, penelitian ini akan mengidentifikasi *hyperparameter* penting yang memengaruhi akurasi prediksi, menganalisis pengaruh setiap *hyperparameter* terhadap performa model, serta mengevaluasi peningkatan akurasi yang dapat dicapai melalui *tuning hyperparameter*. Beberapa *hyperparameter* utama yang akan dieksplorasi mencakup *learning rate*, jumlah *estimators*, kedalaman maksimum pohon, serta ukuran sampel minimum untuk pemisahan dan daun (*min samples split* dan *min samples leaf*).

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem prediksi berbasis data yang lebih akurat dan efektif untuk konteks pemilihan program studi mahasiswa baru. Selain itu, penelitian ini juga bertujuan untuk memberikan rekomendasi konfigurasi *hyperparameter optimal* yang dapat digunakan oleh institusi pendidikan dalam mengimplementasikan algoritma *Gradient Boosting*. Dengan demikian, penelitian ini tidak hanya relevan bagi akademisi, tetapi juga bagi pengambil kebijakan di institusi pendidikan tinggi yang ingin meningkatkan kualitas proses seleksi dan penempatan mahasiswa baru.

Peningkatan akurasi prediksi dalam pemilihan program studi dapat memberikan dampak positif yang luas. Dari sisi mahasiswa, prediksi yang akurat akan membantu mereka memilih program studi yang sesuai dengan minat, kemampuan, dan potensi mereka, sehingga dapat mengurangi tingkat putus sekolah dan meningkatkan keberhasilan akademik. Dari sisi institusi pendidikan, penggunaan sistem berbasis data memungkinkan pengelolaan sumber daya yang lebih efisien dan strategi penerimaan yang lebih tepat sasaran.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan data historis pemilihan program studi mahasiswa baru di Universitas Islam Nahdlatul Ulama Jepara, yang dikumpulkan dari database pendaftaran sejak tahun 2013 hingga 2023. Data tersebut mencakup 54 atribut seperti nilai ujian, jenis sekolah asal, jurusan sekolah, penghasilan orang tua, alat transportasi, dan atribut lainnya. Berikut adalah tahapan penelitian yang dirancang secara sistematis:

a. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari database penerimaan mahasiswa baru universitas, yang telah tersedia dalam bentuk terstruktur. Proses pengumpulan data melibatkan pengolahan file untuk memastikan data siap digunakan dalam analisis. Setiap atribut diproses untuk menangani data yang hilang, tidak konsisten, atau redundan melalui data cleaning [8][9]. Langkah-langkah utama dalam pengumpulan data meliputi:

1. Menghapus atribut dengan proporsi data hilang lebih dari 50% untuk mengurangi redundansi.
2. Mengisi data hilang pada kolom penting dengan metode statistik seperti *mean*, *median*, atau nilai *modus*, bergantung pada jenis datanya.
3. Melakukan encoding pada atribut kategori (misalnya, alat transportasi, jenis sekolah, atau jurusan sekolah) menggunakan metode *one-hot encoding* atau *label encoding* untuk memungkinkan analisis berbasis algoritma *machine learning*.

b. Eksplorasi Data

Pada tahap eksplorasi data, dilakukan analisis deskriptif untuk memahami distribusi data dan hubungan antarvariabel [10]. Langkah-langkah utama yang dilakukan adalah:

1. Visualisasi distribusi data numerik (seperti nilai sekolah dan penghasilan) menggunakan *histogram* atau *boxplot* untuk mengidentifikasi pola distribusi dan *outlier*.
2. Evaluasi hubungan antarvariabel melalui heatmap korelasi untuk mendeteksi korelasi yang signifikan.
3. Penghapusan atribut yang tidak relevan atau memiliki variabilitas rendah untuk mengurangi *noise* dan meningkatkan efisiensi analisis data.

c. Pengembangan Model Dasar

Algoritma *Gradient Boosting* diterapkan menggunakan pustaka *scikit-learn* [11]. Model dasar dikembangkan dengan menggunakan konfigurasi *default hyperparameter* [9]. Tahapan meliputi:

1. Memuat data yang telah diproses (*preprocessing*).
2. Membagi dataset menjadi data latih (80%) dan data uji (20%).
3. Melatih model *Gradient Boosting* pada data latih.
4. Mengevaluasi performa model awal menggunakan data uji. Fungsi *loss* yang digunakan adalah fungsi *log-loss* :

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1)$$



d. Hyperparameter Tuning

Untuk meningkatkan performa model, dilakukan *tuning hyperparameter* dengan dua pendekatan: *Grid Search* dan *Random Search*

1. *Grid Search* [7] mencoba semua kombinasi *hyperparameter* yang telah ditentukan, dengan ruang pencarian sebagai berikut:

- a) *Learning rate* : [0.01, 0.05, 0.1]
- b) *Number of estimators* : [50, 100, 200]
- c) *Max depth* : [3, 5, 10]

2. *Random Search* [12] memilih kombinasi *hyperparameter* secara acak dalam ruang pencarian berikut:

- a) *Learning rate* : [0.01, 0.05, 0.1]
- b) *Number of estimators* : [100, 200, 300]
- c) *Max depth* : [3, 4, 5]
- d) *Min samples split* : [2, 5, 10]
- e) *Min samples leaf* : [1, 2, 4]
- f) *Subsample* : [0.8, 0.9, 1.0]

Dalam penelitian ini, *Random Search* dipilih sebagai metode utama karena dataset memiliki ruang pencarian *hyperparameter* yang cukup besar. *Grid Search*, meskipun komprehensif, memerlukan waktu komputasi yang jauh lebih lama karena mencoba semua kombinasi *hyperparameter*. Sebaliknya, *Random Search* lebih efisien dalam waktu karena hanya memilih kombinasi secara acak dalam ruang pencarian yang luas. Hal ini memungkinkan penelitian untuk mengeksplorasi lebih banyak variasi *hyperparameter* tanpa terlalu membebani sumber daya komputasi [13].

e. Evaluasi Model

Model dievaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* [14]. Perhitungan *F1-score* diberikan oleh rumus:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

Metrik evaluasi yang digunakan adalah:

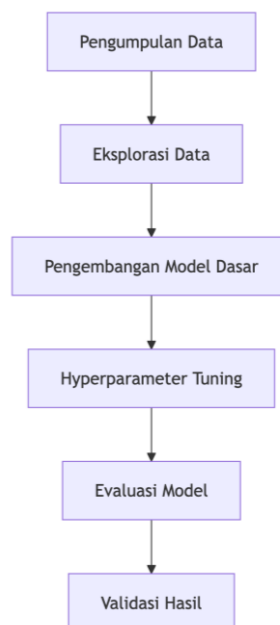
1. Akurasi: Mengukur persentase prediksi yang benar.
2. Precision: Mengukur proporsi prediksi benar pada kelas positif.
3. Recall: Mengukur kemampuan model dalam mendeteksi semua sampel positif.
4. F1-Score: Merupakan rata-rata harmonis antara precision dan recall.

f. Validasi Hasil

Teknik *k-fold cross-validation* [15][16] digunakan untuk memastikan hasil yang lebih generalisasi. Dalam penelitian ini, data dibagi menjadi 5 lipatan (*folds*), dengan setiap lipatan bergiliran menjadi data uji. Skor evaluasi akhir diperoleh dengan menghitung rata-rata skor dari semua lipatan.

2.2 Bagan Alur Tahapan Penelitian

Berikut adalah gambaran alur tahapan penelitian dalam bentuk bagan:



Gambar 1. Alur Tahapan Penelitian



2.3 Kajian Pustaka Algoritma/Metode

Gradient Boosting adalah salah satu algoritma *ensemble learning* berbasis *boosting* yang mampu meningkatkan akurasi prediksi dengan memanfaatkan kesalahan prediksi dari model sebelumnya [17]. Algoritma ini bekerja dengan membangun model secara bertahap, di mana setiap model baru berusaha memperbaiki kesalahan dari model sebelumnya. Keunggulan utama *Gradient Boosting* adalah kemampuannya menangani data kompleks dan fleksibilitas dalam *tuning hyperparameter*.

Dalam konteks penelitian ini, *Gradient Boosting* dipilih karena:

- Kemampuan untuk menangkap interaksi non-linear antar fitur dalam dataset.
- Fleksibilitas dalam *tuning hyperparameter* untuk meningkatkan performa model.
- Ketangguhan dalam menangani data dengan dimensi tinggi dan kompleksitas yang bervariasi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Penelitian ini menghasilkan beberapa temuan signifikan terkait penerapan *hyperparameter tuning* pada algoritma *Gradient Boosting* untuk meningkatkan akurasi prediksi pemilihan program studi mahasiswa baru. Proses pengolahan data dan hasil evaluasi model dijelaskan secara terperinci dalam bagian ini.

a. Data Preprocessing

Penelitian ini menghasilkan beberapa temuan signifikan terkait penerapan *hyperparameter tuning* pada algoritma *Gradient Boosting* untuk meningkatkan akurasi prediksi pemilihan program studi mahasiswa baru. Tahap awal penelitian dimulai dengan *preprocessing* data yang mencakup langkah-langkah berikut:

- Pembersihan data*: Dataset awal terdiri dari 54 kolom, beberapa di antaranya memiliki nilai kosong yang signifikan. Kolom seperti "kebangsaan" dan "alat_transport" dihapus karena lebih dari 50% nilainya kosong.
- Imputasi nilai kosong*: Kolom dengan data hilang diisi menggunakan pendekatan median untuk data numerik dan modus untuk data kategorikal, memastikan integritas data terjaga.
- Transformasi data kategorikal*: Atribut-atribut kategori seperti jenis sekolah dan alat transportasi dikonversi menjadi bentuk numerik menggunakan teknik *one-hot encoding* atau *label encoding*.

Hasil dari preprocessing data ini menghasilkan dataset yang bersih dengan total 42 kolom, yang siap digunakan untuk tahap analisis lebih lanjut. Proses ini memastikan bahwa dataset memiliki kualitas yang baik dan relevan untuk mendukung pengembangan model prediktif.

b. Implementasi Gradient Boosting

Setelah preprocessing, algoritma *Gradient Boosting* diterapkan dengan pengaturan *hyperparameter default* untuk membangun model dasar. Proses ini dimulai dengan membagi dataset menjadi data latih (80%) dan data uji (20%) untuk memastikan bahwa performa model dapat dievaluasi pada data yang belum pernah dilihat sebelumnya. Setelah data diproses, model dilatih menggunakan data latih untuk mempelajari pola hubungan antara fitur input dan target, yaitu program studi yang dipilih.

Pada tahap evaluasi, model dasar ini diuji pada data uji untuk menilai performanya berdasarkan berbagai metrik. Model mencapai akurasi sebesar 99,02%, yang menunjukkan kemampuan tinggi dalam mengklasifikasikan program studi mahasiswa baru. Selain itu, metrik seperti *presisi*, *recall*, dan *F1-score*, juga menunjukkan performa yang tinggi, sebagaimana dirangkum dalam Tabel 1.

Tabel 1. Performansi Model Gradient Boosting Tanpa Tuning

Metrik	Nilai (%)
Akurasi	99,02
Presisi	98,91
Recall	99,02
F1-Score	98,95

Hasil ini menunjukkan bahwa Gradient Boosting sebagai model dasar sudah cukup efektif dalam menangani data kompleks. Namun, potensi peningkatan performa melalui optimasi hyperparameter menjadi langkah berikutnya yang signifikan dalam penelitian ini.

c. Hyperparameter Tuning

Hyperparameter tuning bertujuan untuk mengoptimalkan kinerja model Gradient Boosting dengan menemukan kombinasi pengaturan parameter yang memberikan hasil terbaik. Proses tuning dilakukan menggunakan metode Randomized Search CV, di mana parameter yang diuji meliputi:

Tabel 2. Cross Validation

Parameter	Nilai yang Dicoba	Kombinasi Terbaik
n_estimators	[100, 200, 300]	300
learning_rate	[0.01, 0.05, 0.1]	0.1



max_depth	[3, 4, 5]	4
min_samples_split	[2, 5, 10]	5
min_samples_leaf	[1, 2, 4]	2
subsample	[0.8, 0.9, 1.0]	0.8

Proses ini menggunakan tiga lipatan *cross-validation* dengan 10 kombinasi parameter yang diuji secara acak. Setiap kombinasi diuji sebanyak tiga kali untuk memastikan hasil yang stabil. Model dengan konfigurasi terbaik tersebut mencapai skor rata-rata *cross-validation* sebesar **99,46%**, menunjukkan peningkatan yang signifikan dalam performa model. Hasil tuning menunjukkan bahwa parameter seperti *learning_rate* sebesar 0.1 dan *max_depth* sebesar 4 memberikan keseimbangan antara bias dan varians, menghasilkan performa yang optimal pada data validasi.

3.2 Evaluasi dan Validasi Model

Implementasi model Gradient Boosting dilakukan dengan membagi dataset menjadi data latih (80%) dan data uji (20%). Model diuji menggunakan metrik akurasi, presisi, recall, dan F1-score. Selain itu, evaluasi juga dilakukan terhadap pentingnya fitur dalam memengaruhi hasil prediksi. Fitur "jenis_sekolah", "asal_sekolah" dan "kelamin" ditemukan sebagai variabel paling berpengaruh dalam menentukan program studi mahasiswa.

a. Perbandingan Performa Sebelum dan Sesudah Tuning

Setelah mendapatkan konfigurasi optimal, model diuji kembali pada data uji untuk membandingkan performa dengan model default (tanpa tuning).

Tabel 3. Perbandingan Performansi Sebelum dan Sesudah Tuning

Metrik	Tanpa Tuning (%)	Dengan Tuning (%)
Akurasi	99,02	99,84
Presisi	98,91	99,83
Recall	99,02	99,83
F1-Score	98,95	99,83

Performa model menunjukkan peningkatan signifikan setelah tuning hyperparameter, terutama dalam hal akurasi, presisi, recall, dan F1-score. Peningkatan ini mencerminkan kemampuan tuning dalam mengurangi kesalahan model sekaligus meningkatkan generalisasi terhadap data uji.

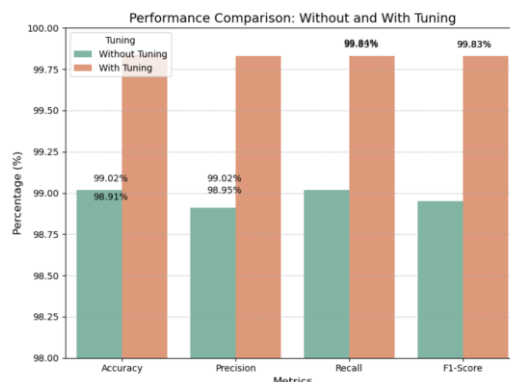
b. Evaluasi Stabilitas Model

Stabilitas model setelah tuning diuji menggunakan tiga lipatan *cross-validation* pada data latih. Hasil evaluasi *cross-validation* disajikan pada tabel berikut:

Tabel 4. Hasil *Cross-Validation* Model Setelah Tuning

Fold	Akurasi (%)
Fold 1	99,66
Fold 2	99,73
Fold 3	95,62
Rata-rata	98,33

Cross-validation menunjukkan bahwa model memiliki performa yang stabil di sebagian besar data latih, dengan rata-rata akurasi sebesar 98,33%. Variasi pada beberapa lipatan menunjukkan bahwa dataset mungkin memiliki distribusi kelas yang tidak seimbang, tetapi tuning berhasil meningkatkan generalisasi model.



Gambar 2. Diagram Perbandingan Akurasi Model Sebelum dan Sesudah Tuning

Variasi akurasi yang cukup signifikan pada Fold 3 (95,62%) dibandingkan *fold* lainnya menunjukkan adanya kemungkinan ketidakseimbangan kelas dalam dataset. Ketidakseimbangan kelas dapat menyebabkan model kesulitan memprediksi kelas minoritas, sehingga menghasilkan variasi performa yang tidak stabil. Untuk mengatasi masalah ini, teknik seperti *resampling* (*undersampling* atau *oversampling*) atau **SMOTE** (*Synthetic Minority Over-sampling*)



Technique) dapat digunakan untuk menyeimbangkan distribusi kelas dalam dataset. Penerapan teknik ini diharapkan dapat meningkatkan kestabilan model lebih lanjut.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa *tuning hyperparameter* pada *Gradient Boosting* mampu meningkatkan akurasi prediksi sebesar 0,83% dibandingkan model *default*. Meskipun peningkatan ini terlihat kecil, hal ini signifikan mengingat *dataset* memiliki dimensi yang tinggi dan karakteristik variabel yang kompleks.

Dibandingkan penelitian sebelumnya, seperti yang dilakukan oleh [3], yang menggunakan *Random Forest* untuk prediksi program studi dengan akurasi 88% , penelitian ini menunjukkan keunggulan *Gradient Boosting* dalam menangani data kompleks. Selain itu, metode ini menawarkan fleksibilitas dalam konfigurasi *hyperparameter*, yang memungkinkan adaptasi terhadap dataset dengan pola yang berbeda.

Hasil evaluasi pentingnya fitur juga memberikan wawasan yang dapat digunakan oleh institusi pendidikan untuk memahami faktor-faktor yang memengaruhi pemilihan program studi mahasiswa. Sebagai contoh, atribut "*jenis_sekolah*", "*asal_sekolah*", dan "*kelamin*" yang mencerminkan jenis sekolah, asal sekolah, dan jenis kelamin calon mahasiswa berhubungan erat dengan pola seleksi program studi tertentu, menunjukkan potensi penyesuaian sistem penerimaan berdasarkan analisis data historis.

Secara keseluruhan, penelitian ini memberikan kontribusi nyata dalam pengembangan sistem berbasis data untuk mendukung proses penerimaan mahasiswa baru. Dengan optimalisasi model prediksi, universitas dapat lebih efektif dalam merancang strategi seleksi dan memberikan rekomendasi program studi yang sesuai dengan potensi dan minat calon mahasiswa.

4. KESIMPULAN

Penelitian ini berhasil mengoptimalkan prediksi pemilihan program studi mahasiswa baru melalui penerapan *hyperparameter tuning* pada algoritma *Gradient Boosting*. Hasil penelitian menunjukkan bahwa *tuning hyperparameter* secara signifikan meningkatkan akurasi prediksi dibandingkan model *default*, dengan akurasi naik dari **99,01%** menjadi **99,84%**. Konfigurasi optimal yang diperoleh meliputi *learning rate sebesar 0,1*, jumlah *estimators 300*, kedalaman maksimum *4*, serta parameter lain seperti *min_samples_split* dan *subsample*, yang membuktikan pentingnya pengaturan *hyperparameter* untuk meningkatkan performa model, terutama pada data kompleks. Selain itu, penelitian ini mengidentifikasi fitur-fitur penting seperti "*jenis_sekolah*", "*asal_sekolah*", dan "*kelamin*" sebagai faktor utama yang memengaruhi prediksi program studi. Informasi ini memberikan wawasan bagi institusi pendidikan untuk merancang strategi penerimaan mahasiswa baru berbasis data. Meskipun demikian, penelitian ini memiliki keterbatasan karena data berasal dari satu institusi sehingga hasilnya mungkin tidak sepenuhnya generalisasi ke institusi lain, serta eksplorasi *hyperparameter* masih terbatas pada kombinasi tertentu. Untuk penelitian selanjutnya, disarankan menggunakan dataset lebih beragam dan metode *tuning* canggih seperti Bayesian Optimization atau AutoML untuk eksplorasi parameter lebih luas. Secara keseluruhan, penelitian ini membuktikan keunggulan *Gradient Boosting* dalam prediksi pemilihan program studi serta pentingnya *tuning hyperparameter* untuk meningkatkan akurasi model, yang dapat dijadikan acuan bagi institusi pendidikan dalam mengembangkan sistem prediksi adaptif untuk mendukung pengambilan keputusan strategis.

REFERENCES

- [1] N. Chaerani *et al.*, "Strategi Dalam Mengembangkan Minat dan Bakat Mahasiswa Melalui Analisis SWOT (Studi Kasus: Program Studi Kehutanan Fakultas Pertanian Universitas Mataram)," *Sasambo: Jurnal Abdimas (Journal of Community Service)*, vol. 5, no. 2, pp. 430–449, May 2023, doi: 10.36312/sasambo.v5i2.1210.
- [2] F. R. N. Rian and M. Hafiyusholeh, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kepuasan Mahasiswa Terhadap Pemilihan Program Studi Matematika UINSA Surabaya Menggunakan Metode Regresi Logistik Ordinal," *AXIOM: Jurnal Pendidikan dan Matematika*, vol. 10, no. 1, p. 26, Jul. 2021, doi: 10.30821/axiom.v10i1.8072.
- [3] H. Mulyo and N. A. Maori, "Peningkatan Akurasi Prediksi Pemilihan Program Studi Calon Mahasiswa Baru Melalui Optimasi Algoritma Decision Tree dengan Teknik Pruning dan Ensemble," *Jurnal Disprotek*, vol. 15, no. 1, pp. 15–25, Jan. 2024, doi: 10.34001/jdpt.v15i1.5585.
- [4] M. Setiawati, I. N. Y. A. Wijaya, and N. M. Estiyanti, "Rancang Bangun Sistem Informasi Penjualan, Pembelian Dan Persediaan Berbasis Web (Studi Kasus Resto Jinggo Tutu)," *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 10, no. 3, p. 533, Dec. 2021, doi: 10.35889/jutisi.v10i3.723.
- [5] A. R. Pratama, R. Rizky Aryanto, A. Taufiq, M. Pratama, and P. Korespondensi, "Model Klasifikasi Calon Mahasiswa Baru Untuk Sistem Rekomendasi Program Studi Sarjana Berbasis Machine Learning," vol. 9, no. 4, Aug. 2022, doi: 10.25126/jtiik.202294311.
- [6] A. R. Al-Aizari *et al.*, "Uncertainty Reduction in Flood Susceptibility Mapping Using Random Forest and eXtreme Gradient Boosting Algorithms in Two Tropical Desert Cities, Shibam and Marib, Yemen," *Remote Sens (Basel)*, vol. 16, no. 2, p. 336, Jan. 2024, doi: 10.3390/rs16020336.
- [7] W. Nugraha and A. Sasongko, "Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search," May 2022. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [8] I. I. Indra, U. Rizki, P. M. Jakak, M. B. Prayogi, and M. Rahman, "Penerapan Metode K-Means Clustering Dalam Pengembangan Strategi Promosi Berbasis Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Nurul Huda)," *Jurnal Nasional Ilmu Komputer*, vol. 5, no. 1, pp. 25–43, Feb. 2024, doi: 10.47747/jurnalnik.v5i1.1656.



- [9] P. Sejati, M. Munawar, M. Pilliang, and H. Akbar, "Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan Random Forest untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 7, pp. 1341–1348, Dec. 2022, doi: 10.25126/jtiik.2022976737.
- [10] N. Rahmadani, S. Wardhani, and A. S. W. Sumah, "Hubungan kemampuan awal, pemahaman konsep, dan hasil belajar siswa dalam pembelajaran daring di SMAN Sumatera Selatan," *Bioma : Jurnal Ilmiah Biologi*, vol. 11, no. 1, pp. 1–9, Apr. 2022, doi: 10.26877/bioma.v11i1.8425.
- [11] A. J. Liu, A. Mukherjee, L. Hu, J. Chen, and V. N. Nair, "Performance and Interpretability Comparisons of Supervised Machine Learning Algorithms: An Empirical Study," Apr. 2022.
- [12] Y. Yennimar, A. Rasid, and S. Kenedy, "Implementation Of Support Vector Machine Algorithm With Hyper-Tuning Randomized Search In Stroke Prediction," *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, vol. 6, no. 2, pp. 61–65, Mar. 2023, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v6i2.3479.
- [13] A. V. Konstantinov and L. V. Utkin, "Interpretable machine learning with an ensemble of gradient boosting machines," *Knowl Based Syst*, vol. 222, p. 106993, Jun. 2021, doi: 10.1016/j.knosys.2021.106993.
- [14] S. M. Natzir, "Perbandingan Kinerja Model Pembelajaran Mesin Dalam Prediksi Banjir Menggunakan KNN, Naive Bayes, Dan Random Forest," *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, vol. 14, no. 2, pp. 59–64, Dec. 2023, doi: 10.52972/hoaq.vol14no2.p59-64.
- [15] S. Lee, T. P. Vo, H.-T. Thai, J. Lee, and V. Patel, "Strength prediction of concrete-filled steel tubular columns using Categorical Gradient Boosting algorithm," *Eng Struct*, vol. 238, p. 112109, Jul. 2021, doi: 10.1016/j.engstruct.2021.112109.
- [16] J. Yu, F. Zhou, K. Zhu, C. Yue, J. Wang, and C. Xie, "Transmission Line Loss Prediction by Cross Validation and Gradient Boosting Decision Tree," *IOP Conf Ser Earth Environ Sci*, vol. 440, no. 3, p. 032099, Feb. 2020, doi: 10.1088/1755-1315/440/3/032099.
- [17] P. Florek and A. Zagdański, "Benchmarking state-of-the-art gradient boosting algorithms for classification," May 2023.