

# ANALISIS PREDIKTIF DROPOUT MAHASISWA BERDASARKAN KINERJA AKADEMIK SEMESTER AWAL MENGGUNAKAN *MACHINE LEARNING*

Putu Satya Saputra

Politeknik Negeri Bali

Jalan Raya Kampus Bukit, Jimbaran, Kec. Kuta Selatan, Kabupaten Badung, Bali 80364

satya@pnb.ac.id

## Abstrak

Fenomena dropout mahasiswa menjadi tantangan signifikan dalam pendidikan tinggi. Penelitian ini bertujuan membangun model prediksi dropout berdasarkan kinerja akademik semester awal menggunakan algoritma Random Forest dan Gradient Boosting. Data bersumber dari UCI Repository, terdiri dari 4424 data mahasiswa. Fokus analisis berada pada atribut seperti jumlah mata kuliah yang diambil, evaluasi, nilai rata-rata, dan usia masuk kuliah. Hasil menunjukkan bahwa algoritma Gradient Boosting mencapai akurasi 70,05%, sementara Random Forest sebesar 70,16%, dengan performa tertinggi pada kategori mahasiswa lulus. Model berhasil mengidentifikasi mahasiswa berisiko tinggi dropout, meskipun masih ada tantangan pada klasifikasi status “enrolled”. Temuan ini membuktikan bahwa pendekatan machine learning dapat diterapkan untuk mendeteksi dini risiko putus studi dan mendukung intervensi akademik yang lebih tepat sasaran.

**Kata Kunci :** dropout, kinerja akademik, machine learning, Random Forest, Gradient Boosting

## Abstract

*Student dropout is a critical issue in higher education. This study aims to develop a predictive model of dropout based on early academic performance using Random Forest and Gradient Boosting algorithms. The dataset, sourced from the UCI Repository, contains 4,424 student records. Key features analyzed include the number of enrolled courses, evaluations, average grades, and enrollment age. Results show that the Gradient Boosting algorithm achieved 70.05% accuracy, while Random Forest reached 70.16%, both performing best in classifying graduates. The model successfully identifies high-risk students, although challenges remain in predicting “enrolled” status. These findings highlight the potential of machine learning for early dropout detection and support more targeted academic interventions.*

**Keyword :** dropout, academic performance, machine learning, Random Forest, Gradient Boosting

## PENDAHULUAN

Tingkat keberhasilan pendidikan tinggi menjadi indikator penting dalam mengukur efektivitas sistem pendidikan suatu negara (Syafii et al., 2023). Salah satu tantangan besar yang dihadapi institusi pendidikan adalah tingginya angka *dropout* mahasiswa, terutama pada tahun-tahun awal masa studi. Fenomena ini tidak hanya berdampak pada citra lembaga pendidikan, tetapi juga membawa implikasi sosial dan ekonomi yang signifikan, baik bagi individu maupun masyarakat.

*Dropout* mahasiswa sering kali dipengaruhi oleh banyak faktor kompleks (Nurmalitasari et al., 2023). Kinerja akademik pada semester awal menjadi salah satu indikator utama yang dapat memberikan sinyal dini tentang potensi risiko putus studi. Mahasiswa yang gagal beradaptasi dengan tuntutan akademik cenderung mengalami kesulitan dalam mempertahankan komitmen belajarnya. Studi oleh Herbaut menegaskan bahwa kegagalan akademik menjadi penyebab dominan terjadinya *dropout* pada tahun pertama perkuliahan (Herbaut, 2021). Kemajuan teknologi di bidang *data science* dan *machine learning* membuka peluang baru untuk melakukan analisis prediktif terhadap fenomena *dropout*. Model prediktif berbasis *machine learning* mampu mengolah data historis kinerja akademik mahasiswa secara lebih efektif, sehingga memungkinkan deteksi risiko *dropout* lebih cepat dan akurat. Menurut Feng et al (Feng & Fan, 2024), aplikasi *machine learning* dalam pendidikan, atau yang dikenal sebagai *educational data mining*, telah terbukti meningkatkan kemampuan

lembaga dalam memahami perilaku mahasiswa. Penggunaan pendekatan prediktif menawarkan manfaat strategis bagi lembaga pendidikan tinggi. Langkah ini sejalan dengan rekomendasi Addison & Williams (2023) yang menyatakan bahwa intervensi dini berbasis data merupakan salah satu kunci dalam meningkatkan retensi mahasiswa.

Penelitian sebelumnya telah banyak mengeksplorasi faktor-faktor yang berkontribusi terhadap *dropout*, seperti kondisi ekonomi, faktor sosial, dan motivasi individu. Namun, masih sedikit yang secara spesifik mengkaji keterkaitan langsung antara kinerja akademik pada semester awal dengan keputusan *dropout* menggunakan pendekatan *machine learning*. Gap penelitian ini menjadi peluang untuk mengembangkan model prediktif yang lebih fokus dan praktis diterapkan. Faktor-faktor akademik seperti jumlah mata kuliah yang diambil, tingkat kelulusan mata kuliah, dan nilai rata-rata semester pertama menjadi variabel-variabel penting yang dapat diolah dalam model. Menurut Crowther (Crowther & Briant, 2021), performa akademik awal memiliki korelasi kuat terhadap keberhasilan studi jangka panjang mahasiswa. Mahasiswa yang menunjukkan kinerja buruk pada semester pertama memiliki kemungkinan lebih besar untuk tidak menyelesaikan pendidikannya [7]. Implementasi analisis prediktif membutuhkan pendekatan sistematis. Dimulai dari tahap pengumpulan data akademik semester awal, dilanjutkan dengan pembersihan dan transformasi data, pemilihan fitur relevan, pembangunan model *machine learning*, hingga evaluasi kinerja model berdasarkan metrik seperti akurasi, *recall*, dan *precision*. Studi oleh Kabathova (Kabathova & Drlik, 2021) menunjukkan bahwa kombinasi tahapan ini mampu menghasilkan sistem prediksi *dropout* yang handal dan efektif.

Pemilihan algoritma *machine learning* yang tepat menjadi faktor krusial dalam membangun model prediktif yang optimal (Ouadah et al., 2022). Algoritma Random Forest dan Gradient Boosting menjadi pilihan populer karena kinerja yang kuat dan kemampuan menangani data dengan banyak variabel. Random Forest, sebagai *ensemble method*, membangun banyak *decision tree* dan menggabungkan hasilnya untuk meningkatkan akurasi dan mengurangi risiko *overfitting* (Salman et al., 2024). Algoritma ini sesuai dengan prediksi *dropout* karena mampu menangkap interaksi kompleks antar variabel akademik mahasiswa. Gradient Boosting menawarkan pendekatan bertahap untuk membangun model prediktif. Setiap model baru dalam Gradient Boosting difokuskan untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya. Teknik ini telah terbukti efektif dalam banyak kompetisi *data science*, termasuk klasifikasi berbasis pendidikan (Colpo et al., 2024). Studi oleh Alshboul et al. (2022) menegaskan bahwa Gradient Boosting Machines memberikan performa yang kompetitif dalam berbagai skenario prediksi kompleks. Penggunaan kedua algoritma ini memungkinkan pengembangan model prediktif mampu menangani ketidakseimbangan data, *noise*, serta hubungan *non-linear* antara fitur-fitur akademik dan risiko *dropout*. Penggunaan data akademik semester awal sebagai fokus utama analisis juga memberikan nilai praktis yang tinggi. Institusi tidak perlu menunggu hingga mahasiswa menunjukkan gejala *dropout* yang nyata. Data nilai dan progres akademik yang rutin dikumpulkan setiap akhir semester sudah cukup untuk membangun sistem peringatan dini yang adaptif.

Penerapan model prediktif berbasis *machine learning* juga memberikan peluang integrasi dengan sistem informasi akademik yang sudah ada. Model prediktif tidak dimaksudkan untuk menghakimi mahasiswa, melainkan untuk mendukung mahasiswa. Tujuan utama penggunaan *machine learning* adalah mengidentifikasi kebutuhan bantuan lebih dini, sehingga mahasiswa mendapatkan dukungan akademik dan emosional yang diperlukan untuk menyelesaikan studinya. Prinsip ini sejalan dengan etika penggunaan teknologi dalam pendidikan yang menempatkan kesejahteraan peserta didik sebagai prioritas utama (Laksita & Sasi, 2024). Adopsi analitik prediktif berbasis *machine learning* dalam pendidikan tinggi juga mendorong pergeseran budaya institusi ke arah *evidence-based decision making*. Keputusan manajerial yang sebelumnya bersifat intuitif kini dapat diperkuat dengan data konkret. Pengelolaan *dropout* tidak lagi sekadar upaya reaktif, melainkan menjadi bagian dari strategi keberlanjutan institusi.

Tantangan dalam membangun model prediktif tidak bisa diabaikan. Permasalahan seperti ketidakseimbangan kelas (*class imbalance*), *noise* pada data, dan perubahan dinamika perilaku

mahasiswa antar generasi perlu diantisipasi. Oleh karena itu, penting untuk terus mengevaluasi dan memperbarui model secara berkala agar tetap relevan dan akurat. Pengembangan model prediktif *dropout* berbasis kinerja akademik semester awal juga membuka peluang penelitian lanjutan. Integrasi variabel-variabel non-akademik seperti keterlibatan dalam kegiatan ekstrakurikuler, interaksi sosial, dan tingkat kesejahteraan mental mahasiswa dapat memperkaya model prediksi. Kombinasi pendekatan *supervised* dan *unsupervised learning* juga patut dieksplorasi untuk mengungkap pola-pola tersembunyi yang belum terdeteksi.

### **PENELITIAN RELEVAN**

Penelitian sebelumnya yang dilakukan oleh Andrianof et al. (2025) berjudul Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia memanfaatkan algoritma Random Forest untuk mengolah data akademik mahasiswa, khususnya IPK dan kehadiran, guna memetakan potensi kelulusan tepat waktu. Proses penelitian ini mencakup pembersihan data, rekayasa fitur, serta validasi model lintas program studi dan semester, yang menghasilkan pemahaman mendalam tentang faktor akademik yang memengaruhi keberhasilan studi. Temuan tersebut memberikan dasar kuat bagi pengembangan sistem peringatan dini dalam manajemen pendidikan tinggi.

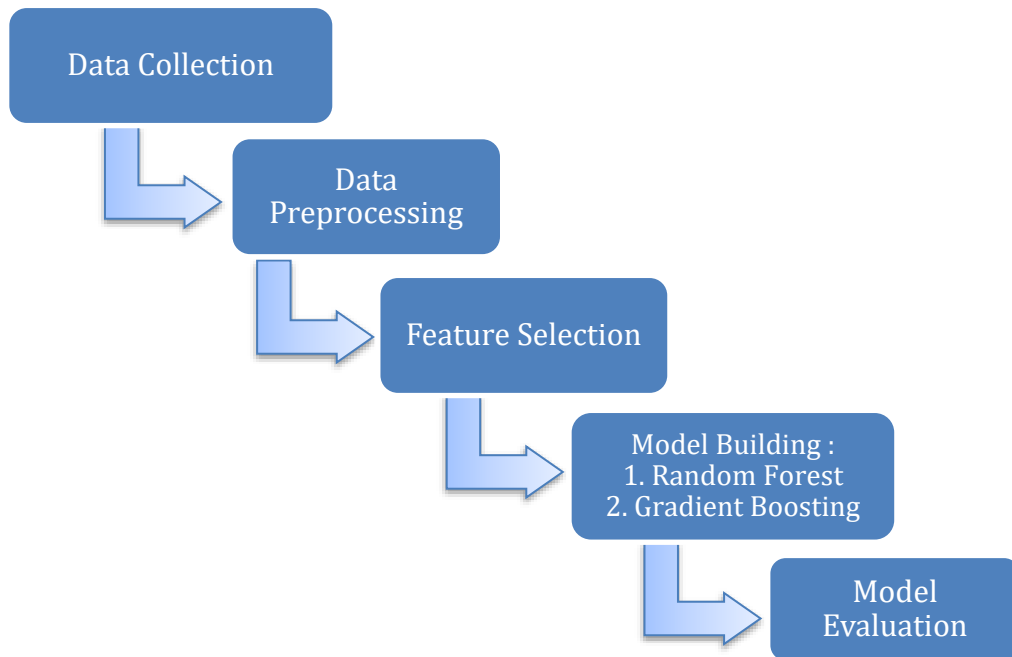
Sementara itu, Mulyo & Khanif Zyen (2025) dalam penelitiannya berjudul Pengaruh Hyperparameter Tuning Gradient Boosting terhadap Prediksi Pemilihan Program Studi Mahasiswa Baru menekankan peran tuning hyperparameter dalam meningkatkan kinerja model Gradient Boosting. Data historis selama sepuluh tahun dianalisis untuk menemukan konfigurasi optimal, serta mengidentifikasi fitur penting seperti jenis sekolah asal, asal sekolah, dan jenis kelamin. Penelitian ini menyoroti pentingnya pendekatan prediktif berbasis data dalam pengambilan keputusan strategis pada proses seleksi program studi.

Sudarman & Budi (2023) melalui riset Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting untuk Prediksi Keberhasilan Studi Mahasiswa menggunakan pendekatan XGBoost untuk mengevaluasi kontribusi variabel akademik dan sosial-ekonomi terhadap keberhasilan studi mahasiswa. Dataset yang kompleks dan heterogen diproses melalui teknik klasifikasi serta validasi K-Fold, menghasilkan model yang mampu menyesuaikan pada ketidakseimbangan kelas.

Kajian bertajuk Prediksi Siswa Putus Sekolah dan Keberhasilan Akademik Menggunakan Machine Learning dari Fitriana et al. (2024) memanfaatkan algoritma klasifikasi untuk menganalisis risiko putus sekolah berdasarkan indikator akademik, latar belakang sosial, serta keterlibatan siswa. Model ini dikembangkan sebagai alat bantu untuk mendeteksi siswa berisiko dan mendukung kebijakan intervensi yang lebih efektif. Keempat penelitian tersebut memiliki keterkaitan erat dengan studi ini, yang mengembangkan pendekatan prediktif berbasis Random Forest dan Gradient Boosting untuk memetakan risiko dropout mahasiswa sejak semester awal. Penelitian ini memperluas cakupan dengan fokus pada analisis performa akademik tahap awal sebagai dasar pengambilan keputusan intervensi, serta mendukung transformasi institusi ke arah sistem edukasi berbasis data.

### **METODE PENELITIAN**

Penelitian ini bertujuan membangun model prediksi dropout mahasiswa berdasarkan performa akademik semester awal menggunakan pendekatan *supervised machine learning*. Fokus utama diarahkan pada eksplorasi pola akademik mahasiswa yang berhubungan dengan risiko tidak menyelesaikan pendidikan tinggi. Penelitian ini menggunakan pendekatan kuantitatif dengan teknik *data mining*, melibatkan serangkaian tahapan sistematis mulai dari pengumpulan data hingga evaluasi performa model prediktif.



Gambar 1. Alur Penelitian

1. *Data Collection*

Dataset bersumber dari UCI Machine Learning Repository, berisi 4424 *instances* dan 36 *features*. Data mencakup atribut akademik, demografis, sosial ekonomi, serta capaian akademik mahasiswa pada semester pertama dan kedua. Setiap *record* merepresentasikan satu mahasiswa dari berbagai program studi.

2. *Data Preprocessing*

*Data preprocessing* dilakukan untuk memastikan kualitas data sebelum analisis. Langkah yang diambil meliputi:

- Penggantian delimiter dari ; menjadi , untuk kompatibilitas pembacaan oleh pandas.
- Pemeriksaan missing values dan verifikasi bahwa tidak terdapat data kosong.
- Encoding terhadap fitur kategorikal menggunakan label encoding untuk mengubah data kategorikal menjadi numerik.
- Standardisasi fitur numerik untuk memperkecil skala yang tidak seragam antar fitur, khususnya untuk model Gradient Boosting yang sensitif terhadap skala data.

3. *Feature Selection*

*Feature selection* difokuskan pada atribut yang mencerminkan performa akademik mahasiswa di awal masa studi, yaitu:

- Curricular Units 1st Sem Enrolled
- Curricular Units 1st Sem Evaluations
- Curricular Units 1st Sem Approved
- Curricular Units 1st Sem Grade
- Admission Grade
- Age at Enrollment

4. *Model Building*

Pengembangan model klasifikasi dilakukan menggunakan dua algoritma utama:

- Random Forest Classifier: Algoritma *ensemble* berbasis *decision tree* yang membangun banyak pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi serta mengurangi risiko *overfitting*.
- Gradient Boosting Classifier: Algoritma *boosting* yang membangun model secara sekuensial dengan menekankan pada kesalahan prediksi dari model sebelumnya, efektif dalam menangkap pola kompleks pada data.

Model *baseline* dikembangkan terlebih dahulu tanpa tuning, kemudian dilakukan optimasi hyperparameter untuk mendapatkan konfigurasi terbaik.

## 5. Model Evaluation

Evaluasi model dilakukan menggunakan beberapa metrik untuk mengukur performa klasifikasi:

- a. Accuracy: Persentase prediksi yang benar terhadap total data.
- b. Confusion Matrix: Matriks perbandingan antara prediksi dan label aktual.
- c. Classification Report: Metrik precision, recall, dan F1-score untuk masing-masing kelas.

Perbandingan hasil antara model *baseline* dan model *tuning* menjadi bagian penting dalam menentukan efektivitas optimasi parameter.

## HASIL DAN PEMBAHASAN

Hasil dari klasifikasi *dropout* mahasiswa berdasarkan performa akademik awal menghasilkan temuan yang signifikan terhadap upaya institusi pendidikan dalam mendeteksi risiko putus studi secara dini. Dua algoritma utama digunakan dalam proses ini yaitu Random Forest dan Gradient Boosting, yang keduanya dikembangkan dalam kerangka *supervised learning* dengan tujuan memprediksi status akhir mahasiswa berdasarkan atribut akademik semester pertama. Fokus utama pada fitur meliputi jumlah mata kuliah yang diambil, jumlah evaluasi yang dilakukan, jumlah mata kuliah yang diluluskan, rata-rata nilai semester pertama, nilai masuk awal (*admission grade*), serta usia saat pendaftaran.

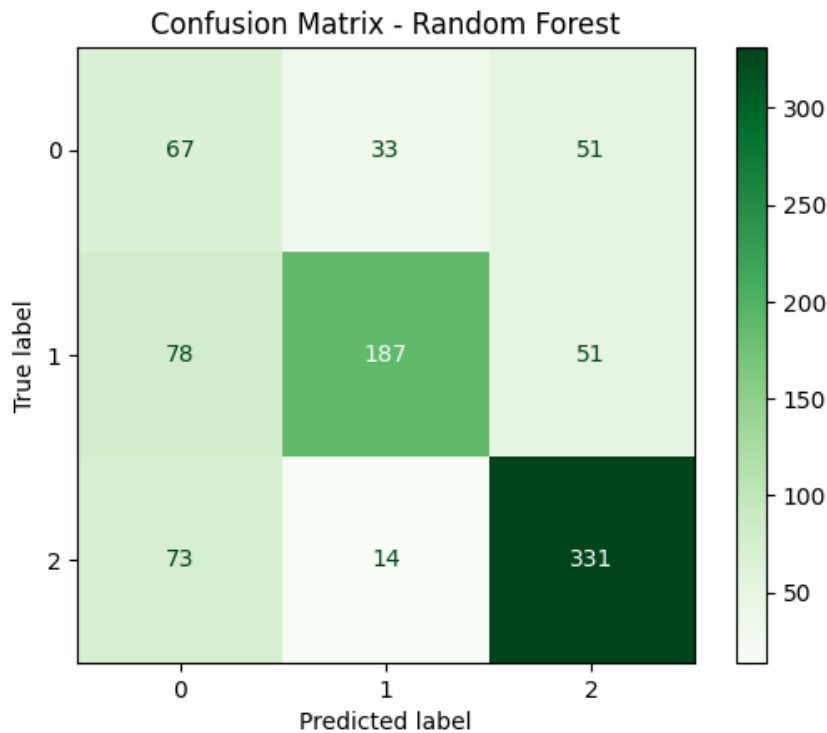
Langkah awal dalam proses ini adalah pemilihan fitur (*feature selection*) yang ditentukan berdasarkan atribut performa akademik awal. Setelah dilakukan pembersihan dan penyesuaian dataset, proses standarisasi diterapkan agar semua fitur berada dalam skala yang seragam. Data kemudian dibagi menjadi data latih dan uji dengan komposisi 80:20 secara stratifikasi agar proporsi setiap kelas tetap seimbang.

Proses *tuning* dilakukan untuk meningkatkan performa masing-masing model. Pada Random Forest, digunakan *GridSearchCV* untuk menjelajahi kombinasi parameter seperti kedalaman pohon maksimum (*max\_depth*), jumlah minimum sampel untuk split (*min\_samples\_split*), dan fungsi pengukuran gini sebagai kriteria pembentukan pohon keputusan. Konfigurasi terbaik diperoleh pada parameter berikut:

```
param_grid_rf = {
    'n_estimators': [100],
    'max_depth': [7],
    'min_samples_split': [5],
    'min_samples_leaf': [2],
    'criterion': ['gini']
}
grid_rf = GridSearchCV(RandomForestClassifier(), param_grid_rf,
cv=5)
grid_rf.fit(X_train, y_train)
y_pred_rf = grid_rf.predict(X_test)
```

Hasil dari model Random Forest menunjukkan akurasi sebesar 0.7016, dengan *precision* tertinggi pada kelas *Dropout* sebesar 0.76 dan *recall* tertinggi pada kelas *Graduate* sebesar 0.92. Hasil ini menunjukkan bahwa model cukup baik dalam mengenali mahasiswa yang berhasil lulus, namun masih mengalami kendala dalam mengklasifikasikan mahasiswa dengan status *Enrolled*, terlihat dari

nilai *recall* kelas tersebut yang hanya 0.14. Berikut merupakan Confusion Matrix dari model Random Forest yang memberikan visualisasi distribusi prediksi dibanding label aktual:



Gambar 2. Confusion Matrix Algoritma Random Forest

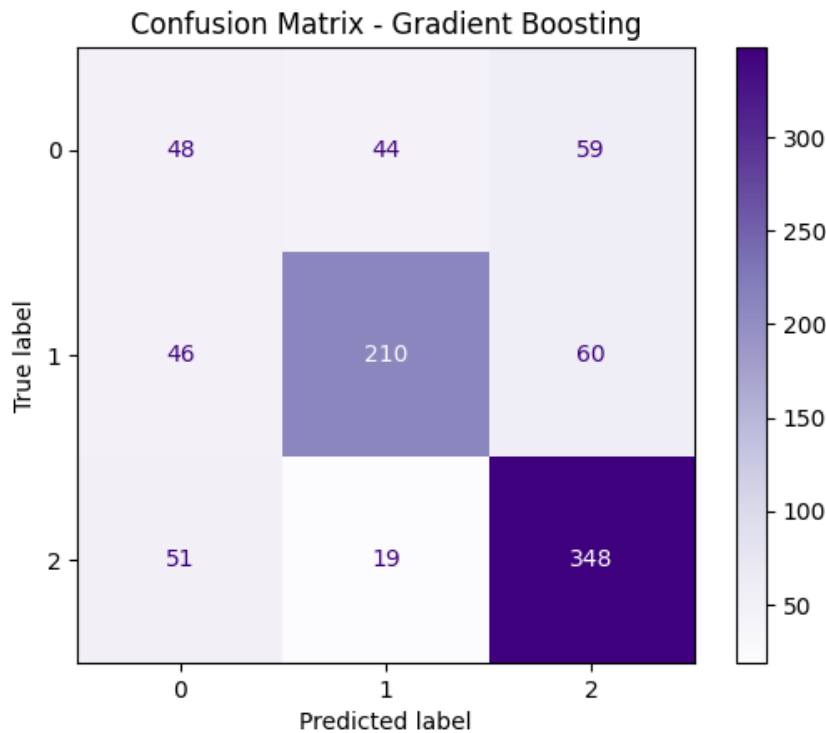
Pada matrix ini terlihat bahwa model cukup konsisten dalam mengenali mahasiswa *Graduate*, dengan angka prediksi benar sebanyak 331 dari 418. Sebaliknya, kelas *Enrolled* memiliki tingkat kesalahan yang signifikan, yang menunjukkan bahwa mahasiswa dalam fase ini memiliki karakteristik yang tumpang tindih dengan kategori *Dropout* maupun *Graduate*.

Model kedua yang digunakan adalah Gradient Boosting. Model ini dibangun melalui kombinasi estimator lemah yang diperbaiki secara iteratif. Proses *tuning* dilakukan pada parameter *learning\_rate*, *max\_depth*, dan *n\_estimators*, yang hasil terbaiknya ditunjukkan sebagai berikut:

```
param_grid_gb = {  
    'n_estimators': [100],  
    'learning_rate': [0.1],  
    'max_depth': [3]  
}  
grid_gb = GridSearchCV(GradientBoostingClassifier(), param_grid_gb, cv=5)  
grid_gb.fit(X_train, y_train)  
y_pred_gb = grid_gb.predict(X_test)
```

Akurasi model Gradient Boosting mencapai 0.7005, dengan *precision* sebesar 0.76 untuk kelas *Dropout* dan *recall* 0.89 untuk kelas *Graduate*. Meski nilai akurasinya hampir setara dengan Random Forest, model ini memberikan nilai f1-score yang lebih merata antar kelas. Berikut adalah Confusion

Matrix untuk Gradient Boosting yang menunjukkan distribusi klasifikasi prediksi terhadap label sebenarnya:



Gambar 3. Confusion Matrix Algoritma Gradient Boosting

Tampak bahwa model ini berhasil mengklasifikasikan 348 dari 418 mahasiswa *Graduate* dengan tepat, dan performanya terhadap kelas *Dropout* maupun *Enrolled* sedikit lebih stabil dibanding Random Forest. Meski demikian, kelas *Enrolled* tetap menjadi tantangan utama karena kurangnya perbedaan karakteristik yang tajam dalam dataset.

### SIMPULAN

Hasil penelitian ini menunjukkan bahwa kinerja akademik mahasiswa pada semester awal berperan penting dalam memprediksi kemungkinan *dropout*. Atribut-atribut seperti jumlah mata kuliah yang diambil, jumlah evaluasi yang diikuti, jumlah mata kuliah yang lulus, nilai rata-rata semester pertama, nilai masuk awal, serta usia saat pendaftaran memiliki kontribusi yang signifikan terhadap klasifikasi status akhir mahasiswa. Penerapan algoritma Random Forest dan Gradient Boosting menunjukkan performa yang cukup baik dalam membedakan tiga kelas target, yaitu *Dropout*, *Enrolled*, dan *Graduate*. Model Gradient Boosting menghasilkan akurasi sebesar 70%, sementara Random Forest memiliki akurasi yang hampir setara. Keduanya menunjukkan kemampuan klasifikasi tertinggi pada kategori *Graduate*, tetapi masih menghadapi tantangan dalam mengenali mahasiswa yang berstatus *Enrolled*. Proses *tuning hiperparameter* pada kedua model berdampak positif dalam meningkatkan akurasi prediksi dibanding model *baseline*. Secara keseluruhan, penelitian ini berhasil mengembangkan pendekatan prediktif berbasis *machine learning* yang dapat dimanfaatkan oleh institusi pendidikan tinggi sebagai sistem peringatan dini. Lembaga dapat mengidentifikasi mahasiswa yang berisiko mengalami *dropout* lebih cepat dengan menerapkan model ini dan merancang intervensi yang lebih tepat sasaran. Upaya ini berpotensi meningkatkan angka kelulusan dan kualitas pembelajaran secara menyeluruh.

### DAFTAR PUSTAKA

Addison, L., & Williams, D. (2023). Predicting student retention in higher education institutions (HEIs). *Higher Education, Skills and Work-Based Learning*, 13(5), 865–885. <https://doi.org/10.1108/HESWBL-12-2022-0257>

- Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction. *Sustainability*, *14*(11), 6651. <https://doi.org/10.3390/su14116651>
- Andrianof, H., Gusman, A. P., & Putra, O. A. (2025). Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia. *Jurnal Sains Informatika Terapan (JSIT) E-ISSN*, *4*(1), 2025.
- Colpo, M. P., Thompsen Primo, T., Aguiar, M. S. de, & Cechinel, C. (2024). Educational Data Mining for Dropout Prediction: Trends, Opportunities, and Challenges. *Revista Brasileira de Informática Na Educação*, *32*, 220–256. <https://doi.org/10.5753/rbie.2024.3559>
- Crowther, P., & Briant, S. (2021). Predicting Academic Success: A Longitudinal Study of University Design Students. *International Journal of Art & Design Education*, *40*(1), 20–34. <https://doi.org/10.1111/jade.12329>
- Feng, G., & Fan, M. (2024). Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization. *Expert Systems with Applications*, *237*, 121555. <https://doi.org/10.1016/j.eswa.2023.121555>
- Fitriana, S., Rinianty, & Pratama, S. A. (2024). Prediksi Siswa Putus Sekolah dan Keberhasilan Akademik Menggunakan Machine Learning. *The Indonesian Journal of Computer Science*, *13*(6).
- Gabriel, K. F., & Flake, S. M. (2023). *Teaching Unprepared Students*. Routledge. <https://doi.org/10.4324/9781003447450>
- Herbaut, E. (2021). Overcoming failure in higher education: Social inequalities and compensatory advantage in dropout patterns. *Acta Sociologica*, *64*(4), 383–402. <https://doi.org/10.1177/0001699320920916>
- Kabathova, J., & Drlik, M. (2021). Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. *Applied Sciences*, *11*(7), 3130. <https://doi.org/10.3390/app11073130>
- Laksita, A. L., & Sasi, K. (2024). Studi Komparasi Kurikulum Pendidikan Tingkat Menengah di Finlandia dan Norwegia. *Jurnal Multidisiplin West Science*, *3*(10), 1592–1606. <https://doi.org/10.58812/jmws.v3i10.1619>
- Mulyo, H., & Khanif Zyen, A. (2025). Pengaruh Hyperparameter Tuning Gradient Boosting Terhadap Prediksi Pemilihan Program Studi Mahasiswa Baru. *BULLETIN OF COMPUTER SCIENCE RESEARCH*, *5*(2), 131–137. <https://doi.org/10.47065/bulletinsr.v5i2.454>
- Nurmalitasari, Awang Long, Z., & Faizuddin Mohd Noor, M. (2023). Factors Influencing Dropout Students in Higher Education. *Education Research International*, *2023*, 1–13. <https://doi.org/10.1155/2023/7704142>
- Ouadah, A., Zemmouchi-Ghomari, L., & Salhi, N. (2022). Selecting an appropriate supervised machine learning algorithm for predictive maintenance. *The International Journal of Advanced Manufacturing Technology*, *119*(7–8), 4277–4301. <https://doi.org/10.1007/s00170-021-08551-9>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. <https://doi.org/10.58496/BJML/2024/007>
- Sudarman, E. J., & Budi, S. (2023). Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting untuk Prediksi Keberhasilan Studi Mahasiswa. *Jurnal Strategi*, *5*(2), 297–314.
- Syafii, A., Bahar, B., Shobichah, S., & Muharam, A. (2023). Pengukuran Indeks Mutu Pendidikan Berbasis Standar Nasional. *Jurnal Multidisiplin Indonesia*, *2*(7), 1697–1701. <https://doi.org/10.58344/jmi.v2i7.332>