

KLASIFIKASI MAHASISWA BERPOTENSI DROP OUT (DO) MENGUNAKAN ALGORITMA RANDOM FOREST

Mitha Zalina Ramadhany¹, Lidya Wati²

¹ Rekayasa Perangkat Lunak, Politeknik Negeri Bengkalis

Jalan Sungai Alam, Kec. Bengkalis, Kabupaten Bengkalis, Riau, Indonesia

² Departement Elektronika Fakultas Teknik, Universitas Negeri Padang

Jalan Prof. Dr. Hamka, Kampus Air Tawar, Padang, Sumatera Barat, Indonesia

mitazalina974@gmail.com

ABSTRAK

Drop Out mahasiswa merupakan masalah serius yang berdampak pada kualitas lulusan dan reputasi perguruan tinggi. Politeknik Negeri Bengkalis mencatat 198 kasus *Drop Out* sejak 2023–2024. Penelitian ini bertujuan membangun sistem klasifikasi mahasiswa berpotensi *Drop Out* ke dalam dua kelas yaitu *Drop Out* (DO) dan Tidak *Drop Out* (DO) menggunakan algoritma *Random Forest* berbasis data akademik dan non-akademik. Metode *Rapid Application Development* (RAD) digunakan dalam pengembangan sistem dengan dataset 398 mahasiswa dan sembilan atribut: jenis kelamin, lama studi, pelanggaran, IPK, ketidakhadiran, penghasilan orang tua, riwayat cuti dan registrasi, status KIP, serta asal tempat tinggal. Evaluasi model menggunakan 20% data uji (80 data) menghasilkan akurasi 97%, presisi 98%, dan *Recall* 97%. Fitur *Early Warning* berhasil mengidentifikasi 49 mahasiswa dari 200 mahasiswa aktif (24%) sebagai berpotensi *Drop Out*. Hasil ini membuktikan bahwa algoritma *Random Forest* efektif sebagai sistem klasifikasi berbasis data untuk deteksi dini mahasiswa berpotensi *drop out*.

Kata kunci : *Random Forest, Drop Out, Early Warning, Klasifikasi Mahasiswa, Machine Learning*

1. PENDAHULUAN

Perguruan tinggi vokasi memiliki peran strategis dalam mencetak sumber daya manusia kompeten melalui pendidikan terapan yang relevan dengan kebutuhan industri. Namun, tantangan serius yang dihadapi institusi pendidikan tinggi adalah tingginya angka mahasiswa yang tidak menyelesaikan studi (*Drop Out*). Fenomena ini tidak hanya berdampak pada penurunan jumlah lulusan berkualitas, tetapi juga merusak reputasi institusi serta menghambat capaian akreditasi [1]. Di tingkat nasional, Kementerian Pendidikan melaporkan rata-rata *Drop Out* rate perguruan tinggi mencapai 8–12% per tahun, dengan variasi signifikan antar institusi dan program studi [2].

Politeknik Negeri Bengkalis sebagai perguruan tinggi vokasi di Provinsi Riau menghadapi tantangan serupa. Berdasarkan data administratif periode 2023–2024, tercatat 198 mahasiswa dinyatakan *Drop Out* melalui Surat Keputusan resmi. Angka ini mewakili sekitar 7,2% dari total populasi mahasiswa aktif, melebihi rata-rata nasional untuk institusi vokasi. Analisis awal menunjukkan bahwa faktor akademik (IPK rendah, ketidakhadiran tinggi) dan non-akademik (kondisi ekonomi, riwayat cuti, domisili) berkontribusi signifikan terhadap risiko *Drop Out*. Proses identifikasi sebelumnya dilakukan secara manual oleh bagian akademik melalui evaluasi data seperti IPK, ketidakhadiran dan pelanggaran. Proses ini seringkali terlambat dalam mendeteksi mahasiswa berisiko karena keterbatasan sumber daya manusia dan volume data yang besar.

Deteksi dini mahasiswa berisiko *Drop Out* menjadi krusial untuk memungkinkan intervensi tepat waktu seperti bimbingan akademik intensif, konseling

psikososial, atau bantuan keuangan. Pendekatan berbasis *Machine Learning* menawarkan solusi otomatisasi yang dapat memproses data multidimensi secara real-time dengan akurasi tinggi [3]. Berbagai algoritma telah diuji dalam penelitian terdahulu: *K-Nearest Neighbor* (K-NN) [4], *Naive Bayes* [5], *Support Vector Machine* (SVM) [6], *Decision Tree* C4.5 [7], hingga *ensemble methods* seperti *Random Forest* [8]. *Random Forest* menunjukkan keunggulan konsisten dalam menangani data tidak seimbang (imbalanced data) dan menghindari overfitting melalui mekanisme *bootstrap aggregating* dan *feature randomness* [9].

Penelitian ini membangun sistem klasifikasi berbasis web ke dalam dua kelas, yaitu *Drop Out* (DO) dan Tidak *Drop Out* (DO) menggunakan algoritma *Random Forest* dengan pendekatan *Rapid Application Development* (RAD) untuk mendukung pengambilan keputusan berbasis data di lingkungan akademik Politeknik Negeri Bengkalis. Dengan pendekatan ini, sistem diharapkan memberikan solusi teknologi yang praktis dan terukur untuk membantu proses deteksi dini mahasiswa berisiko secara lebih cepat dan akurat.

2. TINJAUAN PUSTAKA

2.1. Studi Komparatif Algoritma Klasifikasi *Drop Out*

Penelitian klasifikasi *Drop Out* mahasiswa telah menggunakan berbagai algoritma *Machine Learning* dengan performa bervariasi. Rofi dkk. [1] membandingkan K-NN dan *Random Forest* pada dataset 210 mahasiswa dengan enam atribut (IPK, SKS, penghasilan orang tua, jalur bayar, status bayar, persentase keaktifan). *Random Forest* menghasilkan

akurasi 99,05% dibanding K-NN 98,10%, dengan atribut persentase keaktifan sebagai *feature importance* tertinggi.

Kamsurya [4] menerapkan *Naive Bayes* pada 320 mahasiswa Fakultas Teknik Unismuh Makassar dengan validasi 10-*fold cross-validation*. Akurasi rata-rata mencapai 95,38%, dengan IPK dan kehadiran sebagai faktor dominan. Hidayat et al. [6] menggunakan SVM pada 150 mahasiswa Politeknik Negeri Lhokseumawe dengan atribut IPS, IPK, kehadiran, dan jumlah cuti. Akurasi 86,92% diperoleh dengan rasio pembagian data 80:20, menunjukkan SVM kurang optimal untuk dataset kecil dengan dimensi rendah.

Sinaga dkk. [6] menguji *Decision Tree* C4.5 pada 97 mahasiswa Universitas Advent Indonesia dengan atribut jenis kelamin, umur, agama, tempat tinggal, IPS, disiplin, dan hutang. Akurasi 90% dicapai dengan *Recall* 100%, namun *Precision* hanya 87,50% mengindikasikan adanya *false positive* yang signifikan. Penelitian oleh Salmawati et al. [10] membandingkan *Naive Bayes* (akurasi 97,83%) dan *Decision Tree* (99,13%) pada 1.155 mahasiswa, menemukan bahwa IPK merupakan atribut paling berpengaruh dalam klasifikasi.

Random Forest konsisten menunjukkan performa superior dalam studi terkini. Pradana dkk. [10] mencapai akurasi 98% pada klasifikasi performa karyawan menggunakan *Random Forest*, sementara Dhea dkk. [9] memperoleh akurasi 93,65% pada klasifikasi tingkat stres tidur. Keunggulan *Random Forest* terletak pada kemampuannya menggabungkan prediksi dari *ensemble* pohon keputusan melalui mekanisme *voting*, sehingga mengurangi varians dan meningkatkan stabilitas model [9].

2.2. Random Forest

Random Forest merupakan algoritma *ensemble learning* yang menggabungkan prediksi dari banyak *Decision Tree* untuk meningkatkan akurasi dan mengurangi *overfitting* [9]. Setiap pohon dibangun dari subset data acak (*bootstrap sampling*) dan subset fitur acak (*feature bagging*). Prediksi akhir ditentukan melalui *majority voting* untuk klasifikasi atau rata-rata untuk regresi.

Perhitungan *split* pada setiap *node* menggunakan dua metrik utama:

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Untuk $\log_2 p_i$ didapat dari persamaan:

$$\log(X) = \frac{\ln(X)}{\ln(2)}$$

Keterangan:

S: Himpunan kasus

n: Jumlah partisi S

pi: Proporsi dari Si terhadap S

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2)$$

Keterangan:

S: Himpunan kasus

A: Atribut

n: Jumlah partisi atribut A

|Si|: Jumlah kasus pada partisi ke-i

|S|: Jumlah kasus dalam S [11].

2.3. Machine Learning dan Data mining

Machine Learning merupakan cabang kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa diprogram eksplisit [12]. Dalam konteks klasifikasi *Drop Out*, pendekatan *supervised learning* digunakan karena tersedia label kelas (*Drop Out/Tidak Drop Out*) [13]. *Data mining* berperan sebagai proses ekstraksi pola dari dataset besar melalui teknik klasifikasi, klusterisasi, atau asosiasi [11].

2.4. Data Preprocessing

Data preprocessing merupakan tahapan kritis sebelum pelatihan model, meliputi: *data cleaning* untuk menghapus *missing values* dan *outliers*; *data integration* untuk menggabungkan sumber data berbeda; *data transformation* untuk normalisasi dan *encoding* kategorikal; *data reduction* untuk mengurangi dimensi tanpa kehilangan informasi signifikan [14]. Pada penelitian ini, *Label Encoding* diterapkan pada variabel kategorikal seperti jenis kelamin dan status KIP.

2.5. Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi performa model klasifikasi melalui empat komponen: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False negative (FN)* [11]. Metrik evaluasi yang dihitung meliputi:

Accuracy adalah metode evaluasi yang mengukur seberapa dekat hasil prediksi terhadap nilai aktual. Dengan menghitung jumlah data yang berhasil diklasifikasikan dengan benar, akurasi prediksi dapat diketahui.

$$\text{Rumus Accuracy} : \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

Precision adalah metode evaluasi yang membandingkan jumlah informasi relevan yang berhasil diambil oleh sistem dengan total seluruh informasi yang diambil, baik relevan maupun tidak.

$$\text{Rumus Precision} : \frac{TP}{TP+FP} \quad (4)$$

Recall adalah metode evaluasi yang membandingkan jumlah informasi relevan yang berhasil diambil oleh sistem dengan total seluruh informasi relevan yang tersedia dalam koleksi, baik yang ditemukan maupun yang tidak.

$$\text{Rumus Recall} : \frac{TP}{TP+FN} \quad (5)$$

2.6. Rapid Application Development (RAD)

Rapid Application Development merupakan metodologi pengembangan perangkat lunak iteratif

dengan siklus cepat yang terdiri dari empat fase: perencanaan kebutuhan melalui wawancara dengan *stakeholder*; desain *workshop* untuk *prototype* antarmuka; pengembangan komponen sistem; implementasi dan pengujian [15]. Pendekatan RAD dipilih karena memungkinkan validasi cepat terhadap kebutuhan pengguna akademik.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Rapid Application Development (RAD)* sebagai metodologi pengembangan sistem yang terdiri dari empat fase: perencanaan kebutuhan, desain *workshop* RAD, pengembangan, dan implementasi [15]. RAD dipilih karena memungkinkan iterasi cepat dan validasi langsung dengan *stakeholder* (bagian akademik Politeknik Negeri Bengkalis) untuk memastikan sistem sesuai kebutuhan operasional.

3.1. Dataset dan Atribut Penelitian

Dataset diperoleh dari bagian akademik Politeknik Negeri Bengkalis, terdiri dari 398 data mahasiswa yaitu 200 data mahasiswa aktif (kelas "Tidak Drop Out") dan 198 data mahasiswa *Drop Out* angkatan 2013-2024 berdasarkan Surat Keputusan (kelas "Drop Out"). Sembilan atribut penelitian ditetapkan (Tabel 1), mencakup atribut akademik dan non-akademik sebagai berikut:

Tabel 1. Atribut Penelitian

Atribut	Keterangan
Jenis Kelamin	Jenis kelamin mahasiswa.
Lama Studi	Jenjang gelar yang ditempuh berdasarkan batas waktu.
Pelanggaran	Sanksi ketika melakukan tindakan yang melanggar peraturan akademik.
IPK	Nilai mutu penilaian akhir semester hasil belajar mahasiswa.
Ketidakhadiran	Sanksi kehadiran dengan batas tertentu.
Penghasilan Orang Tua	Jumlah penghasilan orang tua mahasiswa.
Riwayat Cuti & Registrasi	Status cuti akademik dan proses registrasi ulang mahasiswa selama masa studi.
Status KIP	Status penerima Kartu Indonesia Pintar (bantuan pendidikan).
Asal Tempat Tinggal	Lokasi atau daerah asal tempat tinggal mahasiswa sebelum atau selama studi.

Data dibagi secara stratified menjadi 80% data latih (318 data) dan 20% data uji (80 data) untuk menjaga proporsi kelas dalam kedua subset.

3.2. Preprocessing Data

Tahapan data preprocessing meliputi:

- Data cleaning*: Menghapus *missing values* dan *outliers* menggunakan pendekatan *interquartile range (IQR)* sesuai rekomendasi [16].

- Encoding*: Variabel kategorikal dikonversi ke numerik menggunakan *Label Encoding* (misal: "Laki-laki" → 0, "Perempuan" → 1).
- Transformasi IPK: Mengganti koma menjadi titik desimal untuk kompatibilitas numerik (misal: "3,54" → 3.54).
- Validasi Integritas: Memastikan tidak ada duplikasi data dan konsistensi label kelas.

3.3. Implementasi Random Forest

Model *Random Forest* diimplementasikan menggunakan *library Scikit-learn* dengan parameter:

- $n_estimators = 100$ (jumlah pohon keputusan)
- $criterion = 'entropy'$
- $random_state = 42$ (konsistensi model saat dijalankan ulang)

Algoritma bekerja dengan membangun *ensemble* dari 100 pohon keputusan melalui *bootstrap sampling*. Setiap pohon melakukan klasifikasi berdasarkan *Information Gain* pada subset fitur acak, kemudian hasil akhir ditentukan melalui *majority voting*. Perhitungan *Entropy* dan *Information Gain* mengikuti Persamaan (1) dan (2):

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (6)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (7)$$

3.4. Evaluasi Model Random Forest

Performa model dievaluasi menggunakan *Confusion Matrix* dengan metrik:

$$\text{Rumus Accuracy} : \frac{TP+TN}{TP+FP+FN+TN} \quad (8)$$

$$\text{Rumus Precision} : \frac{TP}{TP+FP} \quad (9)$$

$$\text{Rumus Recall} : \frac{TP}{TP+FN} \quad (10)$$

Analisis *feature importance* dilakukan untuk mengidentifikasi atribut dominan yang berkontribusi terhadap prediksi *Drop Out*.

4. HASIL DAN PEMBAHASAN

4.1. Dataset dan Atribut Penelitian

Dataset yang digunakan dalam penelitian diperoleh dari bagian akademik Politeknik Negeri Bengkalis dengan total data 398 data mahasiswa yang terbagi ke dalam dua kelas, yaitu 200 mahasiswa Tidak *Drop Out* (DO) dan 198 mahasiswa *Drop Out* (DO). Penelitian menggunakan sembilan atribut yang mencakup aspek akademik dan non-akademik, yaitu jenis kelamin, lama studi, pelanggaran, IPK, ketidakhadiran, penghasilan orang tua, riwayat cuti dan registrasi, status KIP serta asal tempat tinggal. Dataset kemudian dibagi menggunakan *stratified sampling* dengan rasio perbandingan 80% data latih dan 20% data uji untuk menjaga keseimbangan distribusi kelas.

4.2. Preprocessing Data

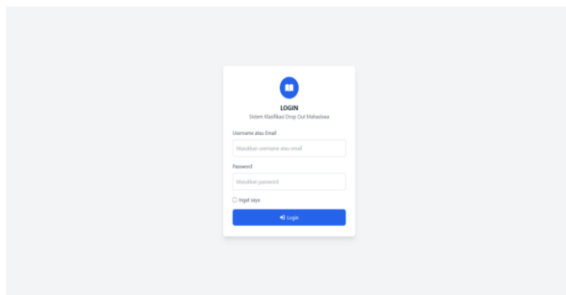
Data mahasiswa dibersihkan dari *missing values* dan *outlier* untuk memastikan kualitas data. Variabel kategorikal dikonversi menjadi numerik menggunakan label *Encoding*, serta dilakukan transformasi pada atribut IPK dengan mengubah format desimal dari

koma menjadi titik. Validasi integritas data untuk memastikan tidak terdapat duplikasi serta konsistensi label kelas antara *Drop Out* dan *Tidak Drop Out*. Hasil *preprocessing* menghasilkan dataset yang bersih dan siap digunakan pada tahap pelatihan model.

4.3. Implementasi *Random Forest*

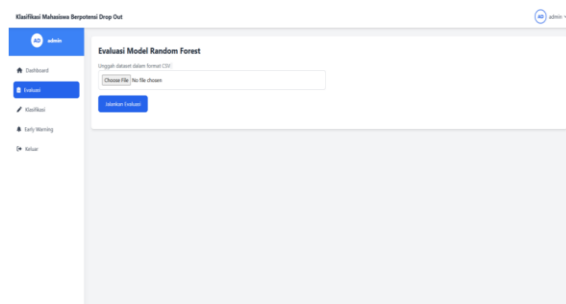
Sistem klasifikasi mahasiswa berpotensi *drop out* dikembangkan berbasis web berbasis *Python Flask* dengan arsitektur tiga lapis (*three-tier architecture*) yang terdiri dari lapisan presentasi (*frontend*), lapisan logika bisnis (*backend*), dan lapisan basis data. Lapisan *frontend* dibangun menggunakan *HTML*, *CSS* dengan *framework Tailwind CSS*, dan *JavaScript* untuk menciptakan antarmuka yang responsif dan intuitif pada berbagai ukuran layar perangkat. Lapisan *backend* mengimplementasikan logika klasifikasi melalui *library Scikit-learn* untuk pembelajaran mesin algoritma *Random Forest*, *Pandas* untuk analisis data dan beberapa *library* pendukung lainnya. Lapisan basis data menggunakan *MySQL* untuk menyimpan data mahasiswa dan historis hasil klasifikasi.

Model *Random Forest* diimplementasikan menggunakan *library Scikit-learn* dengan parameter $n_estimators=100$, $criterion=entropy$ dan $random_state=42$. Sehingga menghasilkan sistem klasifikasi seperti di bawah ini:



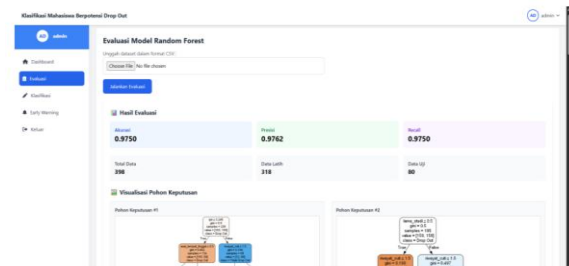
Gambar 1. *Login Admin*

Halaman *login* berfungsi sebagai gerbang autentikasi pengguna sebelum mengakses fitur sistem. Antarmuka dirancang dengan prinsip *user-centered design* yang menampilkan form input untuk *username* dan *password*, dilengkapi *checkbox* "Ingat Saya" untuk kenyamanan akses berulang.



Gambar 2. *Evaluasi Model Random Forest*

Setelah pengunggahan, sistem secara otomatis melakukan *preprocessing* data termasuk konversi format IPK (koma ke titik desimal), label *encoding* untuk variabel kategorikal, serta penanganan *missing values*. Proses ini memastikan kompatibilitas data input dengan struktur model yang telah dilatih pada dataset 398 mahasiswa Politeknik Negeri Bengkalis.



Gambar 3. *Hasil Evaluasi Model Random Forest*

Setelah *preprocessing*, sistem menampilkan hasil evaluasi kuantitatif berupa metrik akurasi (97%), presisi (98%), dan *recall* (97%) yang dihitung dari *confusion matrix*. Selain itu, sistem menyajikan visualisasi pohon keputusan representatif dari *ensemble Random Forest* (dari total 100 pohon) menggunakan *library Graphviz* dengan parameter $n_estimator=100$, $criterion=entropy$ dan $random_state=42$.



Gambar 4. *Halaman Early Warning System*

Fitur *early warning* merupakan fitur utama sistem yang menganalisis 200 mahasiswa aktif (kelas "Tidak Drop Out") untuk mengidentifikasi mahasiswa yang akan berpotensi *drop out* sehingga dihasilkan 49 mahasiswa (24%) berpotensi *drop out* dengan indikator risiko yaitu IPK < 2.5 , ketidakhadiran > 48 jam dan kombinasi domisili luar pulau serta tidak penerima KIP. Indikator risiko seperti IPK < 2.5 dan ketidakhadiran > 48 jam merupakan kombinasi antara kebijakan akademik yang berlaku di Politeknik Negeri Bengkalis serta hasil analisis model *Random Forest* melalui *feature importance*.

4.4. Evaluasi Kinerja Model *Random Forest*

Pada gambar 5, Model *Random Forest* dievaluasi menggunakan 20% data uji (80 data) dari total dataset 398 mahasiswa (200 Tidak *Drop Out* dan 198 *Drop Out*). Hasil evaluasi ditampilkan dalam bentuk *Confusion Matrix*.

```

=====
METRIK EVALUASI MODEL
=====
Akurasi : 97.50%
Presisi : 97.62%
Recall : 97.50%
F1-Score : 97.50%

Confusion Matrix:
[[38  2]
 [ 0 40]]

Laporan Klasifikasi:
      precision    recall  f1-score   support

 Drop Out         1.00      0.95      0.97         40
 Tidak Drop Out    0.95      1.00      0.98         40

 accuracy          0.97         80
 macro avg         0.98      0.97      0.97         80
 weighted avg      0.98      0.97      0.97         80
=====

```

Gambar 5. Metrik Evaluasi Model

Dari hasil *Confusion Matrix* diperoleh TP=38, TN=40, FP=2, FN=0. Berdasarkan *Confusion Matrix*, metrik evaluasi dihitung sebagai berikut:

$$Accuracy = \frac{38+40}{38+2+0+40} = 97\% \quad (11)$$

$$Precision = \frac{38}{38+2} = 98\% \quad (12)$$

$$Recall = \frac{38}{38+0} = 97\% \quad (13)$$

Akurasi 97% menunjukkan bahwa model mampu mengklasifikasikan 78 dari 80 kasus dengan benar. Nilai *Recall* sebesar 97% menunjukkan bahwa model mampu mendeteksi hampir seluruh mahasiswa yang benar-benar berpotensi *Drop Out*, yang sangat penting dalam konteks deteksi dini. Selain itu, nilai *Precision* sebesar 98% menunjukkan bahwa sebagian besar prediksi mahasiswa berpotensi *Drop Out* adalah benar, dengan tingkat kesalahan prediksi relatif rendah. Hasil evaluasi dengan *confusion matrix* menunjukkan bahwa secara keseluruhan, model menunjukkan performa stabil dan dapat diandalkan untuk diimplementasikan ke dalam sistem.

4.5. Validasi Pengujian *Blackbox*

Pengujian *blackbox* dilakukan terhadap lima modul utama sistem untuk memastikan kesesuaian dengan kebutuhan fungsional. Ringkasan hasil pengujian disajikan sebagai berikut:

Tabel 2. Pengujian *Blackbox*

Modul	Kasus Uji	Berhasil	Gagal	Hasil
Login	3	3	0	100%
Evaluasi Model	2	2	0	100%
Klasifikasi	4	4	0	100%
Early Warning	1	1	0	100%
Profile	6	6	0	100%

Seluruh fitur sistem lolos pengujian *blackbox* dengan menunjukkan kinerja yang sangat baik. Pengujian validasi pengguna oleh bagian akademik

Politeknik Negeri Bengkalis juga memberikan umpan balik positif bahwa sistem telah berjalan sesuai kebutuhan operasional untuk mendukung pengambilan keputusan berbasis data dalam menangani mahasiswa berpotensi *Drop Out*.

Berdasarkan hasil evaluasi menyeluruh, dapat disimpulkan bahwa algoritma *Random Forest* terbukti efektif sebagai sistem klasifikasi mahasiswa berpotensi *Drop Out* dengan akurasi tinggi (97%) dan kemampuan deteksi dini melalui fitur *Early Warning*. Integrasi atribut akademik dan non-akademik dapat memberikan pandangan dalam mengidentifikasi faktor risiko *Drop Out* di konteks perguruan tinggi vokasi.

5. KESIMPULAN DAN SARAN

Penelitian berhasil mengimplementasikan sistem klasifikasi mahasiswa berpotensi *Drop Out* menggunakan *Random Forest* dengan akurasi 97%, presisi 98%, dan *Recall* 97% berdasarkan hasil *Confusion Matrix* dengan TP=38, TN=40, FP=2 dan FN=0 dari total 80 data uji. Sistem telah melalui pengujian *Blackbox* pada seluruh fitur dan berhasil berjalan dengan baik sesuai dengan kebutuhan fungsional yang telah ditentukan. Sistem membangun model berdasarkan atribut akademik maupun non-akademik yaitu Jenis Kelamin, Lama Studi, Pelanggaran, IPK, Penghasilan Orang Tua, Riwayat Cuti dan Registrasi, Status KIP dan Asal Tempat Tinggal. Sistem mampu mengidentifikasi 24% mahasiswa aktif sebagai berisiko tinggi melalui fitur *Early Warning*. *Random Forest* terbukti efektif menangani kombinasi data akademik maupun non-akademik untuk prediksi *Drop Out*. Untuk penelitian lanjutan disarankan: penambahan dataset lebih besar untuk meningkatkan generalisasi model, meskipun pembagian kelas relatif seimbang variasi dalam atribut belum sepenuhnya merata untuk keseimbangan representasi model dalam menangkap pola yang lebih luas di populasi mahasiswa.

DAFTAR PUSTAKA

- [1] M. M. Rofi, F. A. Setiawan, Dan F. Riana, "Perbandingan Metode K-Nn Dan Random Forest Pada Klasifikasi Mahasiswa Berpotensi Drop Out," *Infotech Journal*, Vol. 10, No. 1, Hlm. 84–89, Mar 2024, Doi: 10.31949/Infotech.V10i1.8856.
- [2] K. R. Dan T. R. Kementerian Pendidikan, *Laporan Statistik Pendidikan Tinggi 2024*. Jakarta: Kemendikbudristek, 2024.
- [3] M. Marzuqi, T. Ahmad, Dan E. Kristiani, "Prediksi Mahasiswa Drop-Out Di Universitas Xyz," *Urnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 11, No. 6, Hlm. 1345–1350, 2024.
- [4] R. Kamsurya, M. A. M. Hayat, Dan R. Y. Bakti, "Klasifikasi Mahasiswa Berpotensi Putus Studi Menggunakan Algoritma Naive Bayes Pada Fakultas Teknik Unismuh Makassar," *Arus Jurnal Sains Dan Teknologi*, Vol. 2, No. 2, Hlm.

- 511–517, Okt 2024, Doi: 10.57250/Ajst.V2i2.673.
- [5] R. Hidayat, M. Haris, Dan Z. K. Simbolon, “Implementasi Metode Support Vector Machine (Svm) Pada Klasifikasi Drop Out (Do) Mahasiswa,” *Jurnal Infomedia: Teknik Informatika, Multimedia, Dan Jaringan*, Vol. 9, No. 2, Hlm. 51–59, 2024.
- [6] D. Sinaga, E. J. Solaiman, Dan F. J. Kaunang, “Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out Di Universitas Advent Indonesia,” *Teika*, Vol. 11, No. 2, Hlm. 167–173, Okt 2021, Doi: 10.36342/Teika.V11i2.2613.
- [7] H. Sulistiani Dan A. A. Aldino, “Decision Tree C4.5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia),” *Eductic - Scientific Journal Of Informatics Education*, Vol. 7, No. 1, Nov 2020, Doi: 10.21107/Eductic.V7i1.8849.
- [8] S. Samasil, Y. Yuyun, Dan H. Hazriani, “Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Naive Bayes Dan Decision Tree,” *Jurnal Ilmiah Ilmu Komputer*, Vol. 8, No. 2, Hlm. 108–114, Okt 2022, Doi: 10.35329/Jiik.V8i2.242.
- [9] D. Fuji Astari, Y. Herry Chrisnanto, Dan M. Melina, “Klasifikasi Tingkat Stres Saat Tidur Menggunakan Algoritma Random Forest,” *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 7, No. 5, Hlm. 3676–3684, Jan 2024, Doi: 10.36040/Jati.V7i5.7750.
- [10] R. Y. Pradana, F. E. Nastiti, Dan I. Oktaviani, “Machine Learning Pengklasifikasikan Performa Karyawan Direct Sales Force Kartu Prabayar Menggunakan Metode Random Forest Classifier,” *Jekin - Jurnal Teknik Informatika*, Vol. 4, No. 3, Hlm. 590–599, Agu 2024, Doi: 10.58794/Jekin.V4i3.864.
- [11] “Confusion Matrix: Pengertian, Cara Kerja Dan Contoh Soal.” Diakses: 26 Mei 2025. [Daring]. Tersedia Pada: <https://Kantinit.Com/Kecerdasan-Buatan/Confusion-Matrix-Pengertian-Cara-Kerja-Dan-Contoh-Soal/>
- [12] B. Sutedjo Dan D. Oetomo, *Perencanaan & Pembangunan Sistem Informasi*, 1 Ed. Yogyakarta, 2022.
- [13] I. Saputra Dan K. Dinar Ajeng, *Machine Learning Untuk Pemula*, 1 Ed. Informatika, 2022.
- [14] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, Dan S. Ipnuwati, *Implementasi Data Mining (Clustering, Asociation, Prediction, Estimation, Classification)*. Cv. Adanu Abimata, 2023. Diakses: 3 Mei 2025. [Daring]. Tersedia Pada: https://Www.Google.Co.Id/Books/Edition/Implementasi_Data_Mining_Clustering_Assoc/Lsoqaaaqbaj?Hl=Id&Gbpv=1&Dq=Data%20preprocessing%20data%20mining%20bahasa%20indonesia&Pg=Pr2&Printsec=Frontcover
- [15] S. Wijayanto, R. Adrian Putra, Darmasyah, A. Wahyu Aranski, Dan S. Astiti, *Buku Ajar Analisa Perancangan Sistem Informasi*. Pt. Sonpedia Publishing Indonesia, 2024. Diakses: 3 Mei 2025. [Daring]. Tersedia Pada: https://Www.Google.Co.Id/Books/Edition/Buku_Ajar_Analisa_Perancangan_Sistem_Inf/Enl8eaaaqbaj?Hl=Id&Gbpv=1&Dq=Model%20pengembangan%20perangkat%20lunak%20rad&Pg=Pp3&Printsec=Frontcover
- [16] M. Riziq Sirfatullah Alfarizi, M. Zidan Al-Farish, M. Taufiqurrahman, G. Ardiansah, Dan M. Elgar, “Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning Dan Deep Learning,” Bogor, 2023. Doi: <https://Doi.Org/10.30997/Karimahtauhid.V2i1.7518>