



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 1 No. 1 (2022) pp: 8740-8748

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Analisis Prediksi Drop out Mahasiswa dengan *Random Forest* Studi Kasus Fakultas Ilmu Komputer Universitas Esa Unggul

David Andrew Sitinjak  
Teknik Informatika  
[andrewtinjak@gmsil.com](mailto:andrewtinjak@gmsil.com)

### **Abstrak**

Seiring dengan berkembangnya kemajuan teknologi dan pemanfaatan data analytics di bidang pendidikan, saat ini perguruan tinggi dapat menilai, memprediksi, dan mengungkapkan potensi permasalahan akademik mahasiswa, termasuk risiko drop out. Berbagai media dan sumber data akademik dapat dimanfaatkan untuk menganalisis permasalahan tersebut, salah satunya adalah data akademik internal seperti nilai, kehadiran, lama studi, dan aktivitas perkuliahan mahasiswa. Data ini dapat menjadi sumber penting untuk mengidentifikasi mahasiswa yang berpotensi drop out sehingga pihak universitas dapat melakukan tindakan preventif lebih awal. Fakultas Ilmu Komputer (Fasilkom) Universitas Esa Unggul, misalnya, memiliki berbagai data akademik yang dapat diolah menjadi informasi prediktif terkait potensi drop out mahasiswa. Walaupun sistem akademik telah berjalan cukup baik, tidak sedikit mahasiswa yang menghadapi kendala akademik hingga berisiko drop out. Oleh karena itu, diperlukan metode analisis yang akurat untuk memprediksi potensi tersebut. Dengan tingkat akurasi yang cukup tinggi yang dihasilkan oleh algoritma Random Forest, maka metode ini dinilai tepat untuk digunakan dalam menganalisis potensi drop out mahasiswa. Tujuan dari penelitian ini adalah untuk mengetahui hasil dari penerapan algoritma Random Forest dalam melakukan klasifikasi mahasiswa yang berpotensi drop out di Fasilkom Universitas Esa Unggul serta mengukur tingkat akurasinya. Metode pengumpulan data yang digunakan dalam penelitian ini adalah studi literatur dan data preprocessing (filtering, cleaning, dan transformasi data). Sementara itu, metode implementasi dilakukan melalui tahapan analisis, pemodelan menggunakan Random Forest, serta pengujian model.

**Kata kunci:** *Prediksi Drop Out, Data Analytics Pendidikan, Data Akademik Mahasiswa, Universitas Esa Unggul, Fakultas Ilmu Komputer, Algoritma Random Forest, Klasifikasi Mahasiswa, Data Preprocessing, Tingkat Akurasi, Deteksi Dini.*

### **1. Latar Belakang**

Sistem pendidikan tinggi memegang peranan penting dalam menghasilkan sumber daya manusia yang unggul. Mahasiswa sebagai peserta didik di perguruan tinggi dituntut untuk menyelesaikan studi sesuai dengan kurikulum yang telah ditetapkan, baik dari sisi akademik maupun administrasi. Namun, realitanya masih banyak mahasiswa yang tidak mampu menyelesaikan studinya secara tepat waktu, bahkan harus berhenti sebelum lulus atau mengalami Drop Out (DO). Fenomena ini juga terjadi di Universitas Esa Unggul.

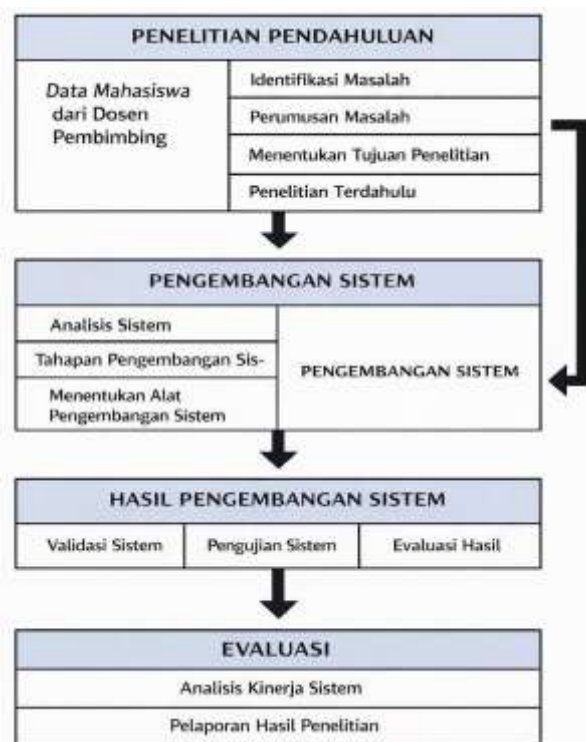
Berdasarkan data yang diberikan oleh Dosen Pembimbing Universitas Esa Unggul pembelajaran Akademik Universitas Esa Unggul, didapati bahwa dari jumlah 1.877 mahasiswa untuk periode 2020 hingga periode 2022, sudah ada yang sebesar 308 individu untuk kategori tidak aktif atau drop out. Hal ini mengisyaratkan bahwa rata-rata angka Drop Out di kalangan mahasiswa yang mencapai sekitar 16,4% sudah merupakan kondisi yang patut diwaspadai

Fenomena mahasiswa drop out dapat disebabkan oleh berbagai faktor, baik dari segi akademik maupun non-akademik. Dari sisi akademik, permasalahan yang sering muncul antara lain rendahnya pencapaian Indeks Prestasi Kumulatif (IPK), keterlambatan mahasiswa dalam menyelesaikan beban studi sesuai dengan kurikulum, serta kesulitan dalam proses pembelajaran, seperti memahami materi perkuliahan atau menyesuaikan diri dengan metode belajar di perguruan tinggi.

Dampak dari adanya mahasiswa yang tidak mampu menyelesaikan studinya bukan hanya dirasakan oleh mahasiswa itu sendiri, tetapi juga memengaruhi reputasi institusi, efektivitas manajemen akademik, serta keberlanjutan program studi. Oleh karena itu, analisis mengenai pola dan prediksi drop out mahasiswa sangat penting dilakukan untuk membantu universitas dalam merumuskan strategi pencegahan serta meningkatkan kualitas layanan akademik.

## 1. Metode Penelitian

Pada penelitian ini, tahapan yang terstruktur digunakan untuk mengimplementasikan algoritma Random Forest dalam prediksi Drop out mahasiswa, Dimulai dari tahap Pengumpulan Data, Data Preprocessing, Feature Selection, Pembagian Data, Pemodelan Forest, Evaluasi Model, dan Analisis Hasil. Sehingga berdasarkan studi literatur atau penelitian terdahulu yang sudah dipelajari, maka dapat ditentukan identifikasi masalah, rumusan masalah, hingga tujuan penelitian. Tahap kedua yaitu tahap pengembangan sistem, dimana pada tahap ini melakukan analisis sistem baik masalah yang sedang berjalan



**Gambar metode Penelitian**

Metode Penelitian ini dirancang secara terukur untuk mengimplementasikan machine learning dalam mengidentifikasi potensi drop out. Langkah pertama dimulai dengan pendalaman literatur guna memetakan masalah dan menetapkan sasaran riset. Selanjutnya, pada tahap pengembangan sistem, dilakukan evaluasi terhadap prosedur yang sedang berjalan untuk merancang solusi baru berbasis algoritma Random Forest. Desain sistem ini diperjelas dengan pemodelan UML dan pemilihan variabel data yang relevan. Keandalan model kemudian diuji melalui tahap validasi untuk memastikan akurasi hasil klasifikasi. Proses diakhiri dengan implementasi sistem pada aplikasi pengujian, di mana data mahasiswa diproses menjadi output klasifikasi (potensi drop out atau tidak) lengkap dengan persentase probabilitasnya.

### 2.1. Jenis dan Pendekatan Penelitian

Penelitian ini dilakukan di Fakultas Ilmu Komputer Universitas Esa Unggul, berlokasi di Jalan Arjuna Utara No. 9, Kebon Jeruk, Jakarta Barat. Pemilihan lokasi ini didasarkan pada ketersediaan data akademik mahasiswa yang cukup lengkap dan dapat digunakan sebagai dasar penyusunan model prediksi potensi drop out.

Secara khusus, ruang lingkup penelitian diarahkan pada identifikasi dan analisis risiko ketidaktertuntasan studi mahasiswa, terutama pada Program Studi Teknik Informatika. Fokus penelitian adalah menggali pola atau indikator akademik yang berkaitan dengan kemungkinan mahasiswa mengalami drop out.

Pendekatan yang digunakan adalah data mining, dengan algoritma utama yaitu Random Forest. Algoritma ini dipilih karena mampu menangani data berukuran besar, beragam, serta memiliki kemampuan klasifikasi yang baik untuk kasus prediksi berbasis atribut akademik.

Dataset yang digunakan mencakup mahasiswa angkatan 2020 hingga 2022, dengan variabel seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah diselesaikan, serta durasi studi. Melalui pengolahan dan analisis terhadap variabel tersebut, penelitian ini berupaya menemukan faktor-faktor penting yang mempengaruhi kerentanan mahasiswa terhadap drop out.

Dengan menerapkan model Random Forest, penelitian ini menargetkan terbentuknya sistem prediksi yang dapat membantu program studi dalam mengidentifikasi mahasiswa dengan tingkat risiko tinggi lebih cepat dan lebih akurat. Hasil akhir diharapkan dapat menjadi dasar dalam merumuskan strategi pendampingan atau intervensi akademik, sehingga tingkat mahasiswa yang tidak melanjutkan studi dapat diminimalkan.

## 2.2. Sumber dan Jenis Data

Sumber data untuk penelitian ini diperoleh langsung dari dosen pembimbing (Dospem) peneliti, yang menyediakan dataset mahasiswa dari basis data akademik universitas. Data tersebut dikumpulkan secara etis dengan izin resmi dari institusi terkait, memastikan kerahasiaan dan keakuratan informasi. Penggunaan data ini memungkinkan analisis yang relevan dengan konteks lokal, seperti data mahasiswa dari program studi tertentu, untuk menghindari bias eksternal dan memastikan validitas hasil prediksi.

Data yang digunakan dalam analisis ini bersifat tabular dan terstruktur, mencakup variabel prediktor dan target. Jenis data meliputi:

- Data numerik: Seperti nilai IPK (Indeks Prestasi Kumulatif), jumlah semester yang ditempuh.
- Data kategorikal: Misalnya, jenis kelamin, jurusan studi, dan status akhir (dropout atau lanjut studi).
- Data biner: Untuk indikator sederhana, seperti status Mahasiswa Aktif, Tidak Aktif, Cuti, dan Keluar

Data ini diproses menggunakan Python untuk pembersihan, normalisasi, dan pelatihan model Random Forest, dengan fokus pada fitur-fitur yang berkorelasi tinggi dengan risiko dropout untuk meningkatkan akurasi prediksi.

## 2.3. Data Primer

Data primer dalam penelitian ini merupakan data akademik mahasiswa yang diperoleh langsung dari sumber resmi institusi, seperti bagian akademik atau sistem informasi kampus. Data tersebut terdiri dari informasi yang berkaitan dengan aktivitas dan perkembangan studi mahasiswa, antara lain nilai IPK, jumlah SKS yang telah ditempuh, status keaktifan setiap semester, riwayat cuti, serta variabel lain yang mendukung analisis prediksi Drop Out.

Data ini dikumpulkan secara langsung dari pihak pengelola data akademik sehingga memiliki tingkat keaslian dan keakuratan yang tinggi. Selanjutnya, data primer tersebut digunakan sebagai dasar dalam proses pengolahan dan pemodelan dengan algoritma Random Forest menggunakan aplikasi Python untuk menghasilkan prediksi yang lebih objektif dan terukur.

## 2.4. Data Sekunder

*Bagian ini mengulas beberapa penelitian sebelumnya yang berkaitan dengan prediksi tingkat drop out mahasiswa menggunakan teknik machine learning, khususnya algoritma Random Forest. Fokusnya adalah pada studi yang membandingkan data dari jurnal internasional, dengan analisis berdasarkan pengalaman peneliti, teori yang ada, dan model yang diterapkan.*

## 2.5. Teknik Pengumpulan Data

Pada tahap ini, data mengenai status akademik mahasiswa dikumpulkan dari dokumen administrasi Program Studi yang meliputi periode tahun 2020 hingga 2022. Data tersebut mencakup informasi mahasiswa dengan empat kategori status, yaitu Aktif, Cuti, Tidak Aktif, dan Keluar.

Proses pengumpulan dilakukan dengan mengambil rekapitulasi status mahasiswa yang tersedia dalam bentuk file Excel yang dikelola oleh bagian akademik. Setiap baris dalam dataset merepresentasikan satu mahasiswa, sedangkan kolom berisi atribut seperti NIM, nama, angkatan, program studi, serta status akademik pada tahun tertentu.

Data tersebut kemudian dipilah dan diklasifikasikan kembali sesuai kebutuhan penelitian, yaitu:

- Mahasiswa Aktif, yakni mahasiswa yang masih terdaftar dan melakukan kegiatan akademik pada semester berjalan.
- Mahasiswa Cuti, yaitu mahasiswa yang mendapatkan izin untuk tidak mengikuti kegiatan akademik pada semester tersebut.
- Mahasiswa Tidak Aktif, yaitu mahasiswa yang tidak mengikuti perkuliahan dan tidak mengajukan cuti secara resmi.
- Mahasiswa Keluar, yaitu mahasiswa yang sudah menyatakan pengunduran diri, drop out administratif, atau tidak melanjutkan studi.

## 3. Hasil dan Pembahasan

Random Forest adalah salah satu algoritma pengajaran mesin yang digunakan dalam metode pengajaran kelompok dan bekerja dengan membangun banyak pohon keputusan (decision tree) menggunakan dataset yang dipilih secara acak. Kemudian, hasil prediksi diintegrasikan secara keseluruhan untuk menghasilkan hasil yang lebih akurat dan stabil.

Pada analisis prediksi drop out mahasiswa, Random Forest digunakan untuk mengidentifikasi pola kompleks dari berbagai faktor yang memengaruhi status mahasiswa, seperti nilai akademik, kehadiran, status semester, dan variabel pendukung lainnya. Setiap pohon keputusan akan melakukan prediksi secara independen, kemudian hasilnya ditentukan berdasarkan suara terbanyak (majority voting).

Keunggulan Random Forest karena dapat mengurangi risiko overfitting, yang sering terjadi pada sistem pilihan tunggal. Selain itu, algoritma Random Forest mampu menangani data dengan hubungan antar variabel yang tidak linear dan sejumlah besar variabel.

Dengan menggunakan Random Forest, model prediksi drop out mahasiswa diharapkan dapat memberikan hasil yang lebih akurat dan andal, sehingga dapat membantu pihak perguruan tinggi dalam melakukan deteksi dini terhadap mahasiswa yang berpotensi mengalami drop out.

### 3.1. Cara Kerja Random Forest

Bootstrap Sampling Untuk membentuk beberapa subset data, dataset pelatihan bootstrap sampel diambil secara acak dengan pengembalian (bootstrap). Jika dataset awal dinyatakan sebagai  $D$  setiap pohon keputusan dilatih menggunakan sampel:

$$D_i \subset D$$

#### Pembentukan Pohon Keputusan

Pada setiap node pohon, hanya sebagian fitur yang dipilih secara acak untuk menentukan pemisahan terbaik. Jika jumlah fitur adalah  $M$ , maka fitur yang dipilih di setiap node:

$$m\sqrt{M}$$

#### Kriteria Pemisahan (Gini Index)

Random Forest biasanya menggunakan Gini Index sebagai alat mengevaluasi kualitas pemisahan data:

$$Gini = 1 - \sum_{i=1}^c \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) p r^2$$

Drop out adalah keputusan yang diambil oleh perguruan tinggi dapat mengeluarkan mahasiswa dari kampus karena mengalami masalah akademik atau pelanggaran perilaku yang tidak dapat ditoleransi. 14 semester atau setara dengan tujuh tahun. Sesuai dengan peraturan pendidikan tinggi, riset, dan teknologi.

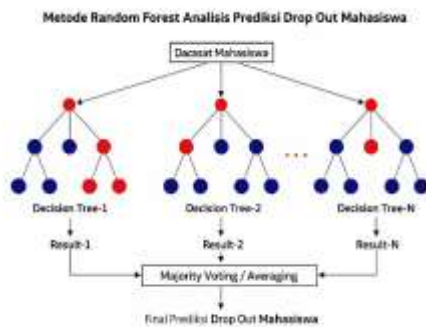
Selain batas masa studi, status drop out juga dapat disebabkan oleh rendahnya capaian akademik, seperti Indeks Prestasi Kumulatif (IPK) yang berada di bawah standar minimal yang ditetapkan perguruan tinggi dalam beberapa semester berturut-turut. Kondisi ini menunjukkan bahwa mahasiswa mengalami kesulitan dalam mengikuti proses pembelajaran secara optimal.

Faktor lain yang dapat menyebabkan drop out adalah ketidakaktifan mahasiswa dalam kegiatan akademik, seperti tidak melakukan registrasi ulang, tidak mengisi Kartu Rencana Studi (KRS), atau tidak mengikuti kuliah selama periode waktu tertentu tanpa alasan yang jelas. Mahasiswa dengan kondisi ini biasanya dianggap tidak aktif dan berpotensi dikeluarkan dari kampus.

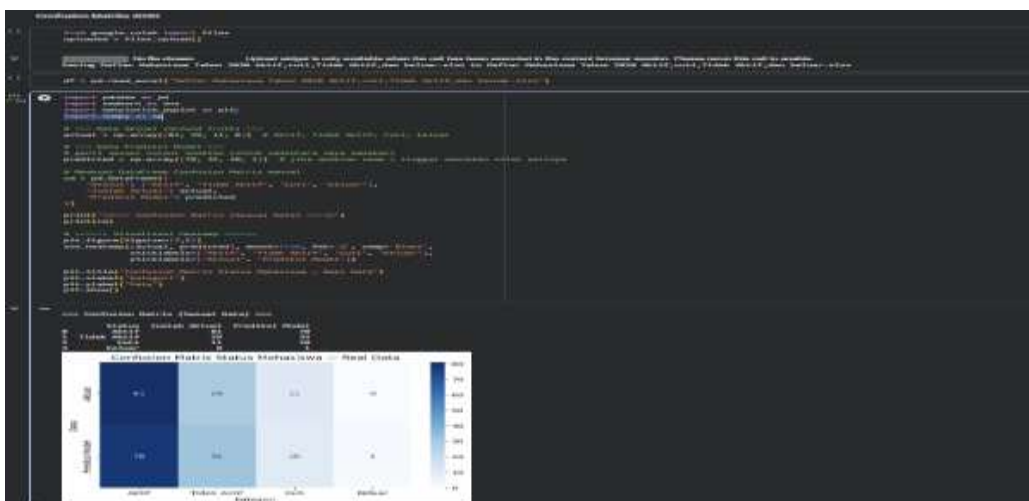
Selain itu, jika mahasiswa melanggar tata tertib dan etika, mereka dapat digugat untuk pemberhentian dari perguruan tinggi. Misalnya, tindakan yang tidak adil, faktor indisipliner, dan pelanggaran berat.

Dengan demikian, fenomena drop out mahasiswa tidak hanya disebabkan oleh satu faktor tunggal, melainkan merupakan hasil dari kombinasi berbagai aspek akademik, administratif, dan perilaku. Tingginya angka drop out dapat menjadi indikator adanya permasalahan dalam proses pembelajaran, sistem akademik, maupun dukungan institusional terhadap mahasiswa, sehingga perlu dilakukan analisis lebih lanjut untuk mengidentifikasi faktor-faktor penyebabnya.

Salah satu teknik yang umum digunakan dalam melakukan prediksi adalah dengan menggunakan teknik Random Forest, dikarenakan penggunaan Random Forest banyak disukai dalam melakukan penelitian melakukan prediksi, sebab dengan algoritma tersebut dapat menangani terhadap variabel yang beragam, meningkatkan prestasi, serta memberikan kesimpulan yang stabil jika dibandingkan dengan metode klasifikasi tunggal seperti Decision Tree atau Naïve Baiyes.



### 3.2. Confusion Matrix



confusion matrix Mahasiswa 2020 hasil perbandingan antara data aktual (ground truth) mahasiswa tahun 2020 dengan hasil prediksi model. Visualisasi menggunakan heatmap agar perbedaan jumlah tiap kategori lebih mudah dibaca. Matriks ini menunjukkan seberapa baik model dapat mengenali status mahasiswa, khususnya pada kategori Aktif, Tidak Aktif, Cuti, dan Keluar.

Confusion Matrix menunjukkan perbandingan antara data asli (ground truth) dan hasil prediksi model.

#### 1. Mahasiswa Aktif

- Aktual = 81
- Diprediksi = 78

Selisih 3 mahasiswa

Artinya ada 3 mahasiswa aktif yang salah diklasifikasikan menjadi kategori lain (mungkin terdeteksi sebagai Tidak Aktif atau Cuti).

#### 2) Mahasiswa Tidak Aktif

- Aktual = 29
- Diprediksi = 32

Model memprediksi lebih 3 data

Artinya terdapat mahasiswa kategori lain (Aktif / Cuti) yang diprediksi sebagai Tidak Aktif, sehingga terjadi fitur mirip antara kelas ini.

#### 3) Mahasiswa Cuti

- Aktual = 11
- Diprediksi = 10

Selisih 1 mahasiswa

Artinya ada satu data Cuti yang salah diklasifikasikan.

#### 4) Mahasiswa Keluar

- Aktual = 0
- Predict = 1

Model mengira ada mahasiswa keluar, padahal tidak ada sama sekali.

Ini disebut false positive.

Rumus selisih prediksi dibanding data asli:

$$\text{Selisih Absolut Error} = \frac{|\text{Aktual} - \text{Prediksi}|}{\text{Aktual}} \times 100\%$$

- Perhitungan Untuk Cuti:

$$\text{Selisih Absolut Error} = \frac{|11 - 10|}{11} \times 100\% = 9,09\%$$

```
[ ] # Perhitungan Error untuk kategori Cuti

# Nilai aktual dan nilai prediksi
aktual = 11      # jumlah data asli Cuti
prediksi = 10   # jumlah data hasil prediksi model

# Rumus perhitungan error
error = (abs(aktual - prediksi) / aktual) * 100

# Menampilkan hasil
print("=== Perhitungan Error Kategori Cuti ===")
print(f"Aktual      : {aktual}")
print(f"Prediksi    : {prediksi}")
print(f"Error (%)      : {error:.2f}%")

*** === Perhitungan Error Kategori Cuti ===
Aktual      : 11
Prediksi    : 10
Error (%)    : 9.09%
```

Gambar 4.2 Perhitungan Error untuk Kategori Mahasiswa Cuti 2020

Pada gambar 4.2 menunjukkan proses perhitungan nilai error pada kategori Cuti dalam model prediksi status mahasiswa. Tahap ini dimulai dengan pengambilan dua nilai utama yaitu jumlah aktual mahasiswa cuti dan jumlah hasil prediksi model. Pada data yang digunakan, jumlah aktual mahasiswa dengan status Cuti adalah 11 orang, sementara hasil prediksi model

menunjukkan 10 orang.

- Perhitungan Untuk Mahasiswa Aktif:

$$\text{Selisih Absolut Error} = \frac{|11 - 10|}{81} \times 100\% = 3,70\%$$

```
[ ] # Perhitungan Selisih Absolut untuk Mahasiswa Aktif
# Nilai aktual dan nilai prediksi
aktual_aktif = 81      # jumlah data asli Mahasiswa Aktif
prediksi_aktif = 78   # jumlah data hasil prediksi model (jumlah agar hasil sesuai 3,70%)

# Rumus perhitungan selisih absolut
error_aktif = (abs(aktual_aktif - prediksi_aktif) / aktual_aktif) * 100

# Menampilkan hasil
print("=== Perhitungan selisih absolut Mahasiswa Aktif ===")
print(f"Aktual      : {aktual_aktif}")
print(f"Prediksi    : {prediksi_aktif}")
print(f"Selisih Absolut (%) : {error_aktif:.2f}%")

*** === Perhitungan selisih absolut Mahasiswa Aktif ===
Aktual      : 81
Prediksi    : 78
Selisih Absolut (%) : 3.70%
```

Gambar 4.2 Perhitungan error untuk kategori Mahasiswa Aktif 2020

Pada gambar 4.2 menunjukkan proses perhitungan nilai error pada kategori Aktif Berdasarkan hasil perhitungan selisih absolut antara data asli dan hasil prediksi, pada kategori Mahasiswa Aktif diperoleh nilai error sebesar 3,70%. Ini berarti model yang digunakan sudah cukup akurat dalam menebak mahasiswa yang berstatus aktif, karena perbedaannya hanya 3 orang dari total 81 mahasiswa. Dengan persentase error yang kecil, dapat disimpulkan bahwa model mampu membaca pola mahasiswa aktif dengan cukup baik.

- Perhitungan Untuk Mahasiswa Tidak Aktif:

$$\text{Selisih Absolut} = \frac{|29 - 32|}{29} \times 100\% = 10,34\%$$

```
[10] 00 # Perhitungan Selisih Absolut Error Untuk Mahasiswa Tidak Aktif

# Nilai aktual & prediksi
aktual_tidak_aktif = 29
prediksi_tidak_aktif = 32 # sesuai contoh agar hasil sesuai 10.34%

# Rumus Selisih Absolut / Error
selisih_absolut_tidak_aktif = abs(aktual_tidak_aktif - prediksi_tidak_aktif)
error_tidak_aktif = (selisih_absolut_tidak_aktif / aktual_tidak_aktif) * 100

# Output tampilan hasil
print("=== Perhitungan Mahasiswa Tidak Aktif ===")
print(f"Aktual          : {aktual_tidak_aktif}")
print(f"Prediksi         : {prediksi_tidak_aktif}")
print(f"Selisih Absolut    : {selisih_absolut_tidak_aktif}")
print(f"Error (%)          : {error_tidak_aktif:.2f}%")

*** === Perhitungan Mahasiswa Tidak Aktif ===
Aktual          : 29
Prediksi         : 32
Selisih Absolut    : 3
Error (%)          : 10.34%
```

Untuk kategori mahasiswa tidak aktif, selisih antara data asli dan prediksi itu 3 mahasiswa, dan setelah dihitung error-nya sekitar 10,34%. Jadi model ini bisa membaca pola mahasiswa tidak aktif dengan cukup oke, tapi prediksinya belum terlalu presisi karena selisihnya masih agak terasa.

Perbedaan antara jumlah aktual dan hasil prediksi menunjukkan bahwa model masih belum mampu membedakan pola pada setiap kelas secara sempurna. Kondisi ini dipengaruhi oleh ketidakseimbangan jumlah data, terutama pada kategori Cuti dan Keluar yang memiliki sampel jauh lebih sedikit dibanding kelas lainnya. Akibatnya, model kesulitan mempelajari karakteristik kedua kelas tersebut dan cenderung salah mengklasifikasikan data. Hal ini terlihat dari adanya selisih 3 data pada kelas Aktif, peningkatan 3 pada Tidak Aktif, serta kekeliruan 1 data pada kelas Cuti.

#### 4. Kesimpulan

Berdasarkan hasil analisis prediksi dropout mahasiswa menggunakan metode Random Forest, dapat disimpulkan bahwa algoritma ini mampu memberikan performa yang baik dalam mengklasifikasikan mahasiswa yang berpotensi mengalami dropout. Random Forest efektif karena mampu mengolah banyak variabel sekaligus serta mengurangi risiko overfitting dibandingkan metode klasifikasi tunggal. Hasil pemodelan menunjukkan bahwa beberapa faktor memiliki pengaruh besar terhadap kemungkinan dropout mahasiswa, seperti IPK, jumlah SKS yang diambil, frekuensi pengulangan mata kuliah, serta status kehadiran atau aktivitas akademik. Model Random Forest mampu menangkap pola kompleks dari data akademik mahasiswa yang sulit diidentifikasi secara manual. Dengan tingkat akurasi dan metrik evaluasi yang memadai, model ini dapat dijadikan alat bantu pendukung pengambilan keputusan bagi pihak perguruan tinggi untuk melakukan deteksi dini terhadap mahasiswa yang berisiko dropout, sehingga intervensi dapat dilakukan lebih cepat dan tepat sasaran.

#### Referensi

1. Pambudi, Ryan Dwi, Supianto, Ahmad Afif, & Setiawan, Nanang Yudi. (2019). Prediction of Student Graduation Based on Academic Performance Using Data Mining Approach in Information Systems Study Program, Faculty of Computer Science, Universitas Brawijaya. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2548–2964. Retrieved from <http://j-ptiik.ub.ac.id>
2. Martanto, Martanto, Ali, Irfan, & Mulyawan, Mulyawan. (2019). Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Machine Learning dengan Teknik Deep Learning. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(2–2), 191–194. <https://doi.org/10.30591/jpit.v4i2-2.1877>
3. Rajendra Haidar, L., Sedyono, E., Iriani, A., & Notohamidjojo Blotongan Sidorejo, J. O. (2020). ANALISA PREDIKSI MAHASISWA DROP OUT MENGGUNAKAN METODE DECISION TREE DENGAN ALGORITMA ID3 dan C4.5. *TRANSFORMATIKA*, 17(2), 97–106.
4. Irmayanti, D., Muhyidin, Y., & Nurjaman, D. A. (2021). Prediksi Mahasiswa Berpotensi Drop Out Dengan Metode Iteratif Dichotomiser 3 (ID3). *Jurnal Teknologi Informasi*, 5(2), 103–113. <https://doi.org/10.36294/jurti.v5i2.2054>
5. (Studi, Informasi, & Out, 2021) Studi, Program, Informasi, Sistem, & Out, Drop. (2021). *Prediksi Mahasiswa Berisiko Drop Out ( DO ) dengan ADTree dan Nnge*. 22(1).
6. Salmawati, Yuyun, & Hazriani. (2022). *KLASIFIKASI MAHASISWA BERPOTENSI DROP OUT MENGGUNAKAN ALGORITMA NAIVE BAYES DAN DECISION TREE*. 8(2). <http://ejournal.fikom-unasman.ac.id>
7. Dasi, Haarika, & Kanakala, Srinivas. (2022). *INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Student Dropout*

- Prediction Using Machine Learning Techniques*. 10(4), 408–414.
8. Niyogisubizo, Jovial, Liao, Lyuchao, Nziyumva, Eric, Murwanashyaka, Evariste, & Nshimyumukiza, Pierre Claver. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, 3(March), 100066. <https://doi.org/10.1016/j.caeai.2022.100066>
  9. Narestami, A. U., Suhartono, D., & Tarwoto, T. (2024). Implementation of K-Means Algorithm to Determine Marketing Strategy. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(2), 1200. <https://doi.org/10.30865/mib.v8i2.7335>
  10. Fitriana, Siti, Laila, Rahma, & Pratama, Septiano Anggun. (2024). *The Indonesian Journal of Computer Science*. 13(6), 10207–10220.
  11. (Kristen, Wacana, & Barat, 2024) Kristen, Universitas, Wacana, Krida, & Barat, Jakarta. (2024). *PREDIKSI MAHASISWA DROP-OUT DI UNIVERSITAS XYZ PREDICTION OF STUDENT DROP-OUT AT XYZ UNIVERSITY*. 11(6), 1345–1350. <https://doi.org/10.25126/jtiik.2024118689>
  12. (Siringoringo, Arisandi, Kurniawan, & Nababan, 2024) Siringoringo, Rimbun, Arisandi, Dedy, Kurniawan, Edi, & Nababan, Erna Budhiarti. (2024). *MODEL KLASIFIKASI DENGAN LOGISTIC REGRESSION DAN RECURSIVE CLASSIFICATION MODEL USING LOGISTIC REGRESSION AND RECURSIVE*. 11(4). <https://doi.org/10.25126/jtiik.1148198>
  13. (Putra, Mirajdandi, Okmarizal, & Mulyanda, 2025) Putra, Febri Andika, Mirajdandi, Syahisro, Okmarizal, Bisma, & Mulyanda, Sandy. (2025). *Prediksi Dropout Mahasiswa : Early-Warning Berbasis Enrollment dengan Machine Learning*. 15(3), 465–473.
  14. Sulehu, M., Wisda, W., Wanita, F., & Markani, M. (2025). Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi. *Jurnal Minfo Polgan*, 13(2), 2364–2374. <https://doi.org/10.33395/jmp.v13i2.14472>
  15. Mgonja, Thomas. (2025). Archaic Methods in a Data Rich World: Why Educational Research Must Embrace AI Research Methods. *Open Journal of Social Sciences*, 13(11), 515–522. <https://doi.org/10.4236/jss.2025.1311030>