



## KLASIFIKASI RESIKO *DROP OUT* MAHASISWA ITAF KUPANG MENGUNAKAN *RANDOM FOREST* SEBAGAI SISTEM PERINGATAN DINI

**Mohamad Iqbal Ulumando**

iqbal77ulumando@gmail.com

Program Studi Informatika, Institut Teknologi Alberth Foenay Kupang

### Abstrak

Penelitian ini bertujuan untuk mengembangkan model klasifikasi resiko *drop out* mahasiswa pada Institut Teknologi Alberth Foenay (ITAF) Kupang menggunakan algoritma *Random Forest*. Dataset yang digunakan terdiri dari 200 mahasiswa dengan beberapa variabel akademik seperti IPK, jumlah SKS, rata-rata nilai, kehadiran, mata kuliah ulang, durasi studi, dan jumlah mata kuliah tidak lulus. Tahap *preprocessing* dilakukan melalui pembersihan data, imputasi nilai hilang, standarisasi tipe data, serta penanganan ketidakseimbangan kelas. Model *Random Forest* kemudian dilatih menggunakan data training dan dievaluasi menggunakan data *testing*. Hasil evaluasi menunjukkan bahwa model mampu melakukan klasifikasi resiko *drop out* dengan performa sangat baik, dengan akurasi 91%, *precision* 83%, *recall* 78%, *F1-Score* 0.80, dan *ROC-AUC* 0.94. Analisis *feature importance* menunjukkan bahwa IPK, kehadiran, dan jumlah mata kuliah tidak lulus merupakan variabel yang paling berpengaruh dalam penentuan resiko. Secara keseluruhan, penelitian ini menunjukkan bahwa *Random Forest* efektif digunakan sebagai sistem peringatan dini untuk mengklasifikasikan mahasiswa berpotensi *drop out* sehingga dapat membantu institusi dalam melakukan intervensi akademik secara lebih tepat sasaran.

**Kata Kunci :** Klasifikasi, *Drop Out* Mahasiswa, *Random Forest*, ITAF Kupang.

### Abstract

*This study aims to develop a drop out risk classification model for students at the Alberth Foenay Institute of Technology (ITAF) Kupang using the Random Forest algorithm. The dataset used consists of 200 students with several academic variables such as GPA, number of credits, average grade, attendance, retakes, study duration, and number of failed courses. The preprocessing stage is carried out through data cleaning, imputation of missing values, standardization of data types, and handling of class imbalance. The Random Forest model is then trained using the training data and evaluated using the testing data. The evaluation results show that the model is able to classify drop out risks with excellent performance, with an accuracy of 91%, precision of 83%, recall of 78%, F1-Score of 0.80, and ROC-AUC of 0.94. Feature importance analysis shows that GPA, attendance, and number of failed courses are the most influential variables in determining risk. Overall, this study shows that Random Forest is effective as an early warning system to classify students who have the potential to drop out, thus assisting institutions in implementing more targeted academic interventions.*

**Keywords :** Classification, Student Drop Out, Random Forest, ITAF Kupang.

## 1. Pendahuluan

Perguruan tinggi memiliki tanggung jawab besar dalam memastikan bahwa seluruh mahasiswa dapat menyelesaikan studi tepat waktu dan memperoleh hasil akademik yang optimal [1]. Namun demikian, fenomena *drop out* atau berhentinya mahasiswa dari proses pendidikan sebelum kelulusan masih menjadi permasalahan yang umum terjadi di berbagai institusi pendidikan tinggi di Indonesia [2]. Institut Teknologi Alberth Foenay (ITAF) Kupang tidak terlepas dari tantangan tersebut, di mana terdapat sejumlah mahasiswa yang mengalami kendala dalam melanjutkan studi hingga selesai. Kondisi ini tentu berdampak pada jumlah lulusan, efektivitas pengelolaan akademik, serta reputasi institusi sebagai penyelenggara pendidikan tinggi [3].

Fenomena *drop out* sendiri dipengaruhi oleh beragam faktor, mulai dari masalah akademik seperti rendahnya indeks prestasi, kesulitan memahami materi, hingga ketidakteraturan dalam mengikuti perkuliahan. Di samping itu, faktor eksternal seperti kondisi ekonomi keluarga, tuntutan pekerjaan, kurangnya dukungan lingkungan sosial, serta persoalan motivasi belajar turut berkontribusi terhadap meningkatnya risiko seorang mahasiswa untuk berhenti kuliah [4]. Pada konteks ITAF Kupang, beberapa pola terkait performa akademik dan kehadiran mahasiswa menjadi indikator awal yang menunjukkan kecenderungan mahasiswa menghadapi risiko akademik lebih tinggi dari yang lain.

Identifikasi risiko *drop out* secara manual sering kali membutuhkan waktu yang panjang dan tidak selalu akurat. Dengan jumlah mahasiswa yang semakin bertambah setiap tahun, proses pemantauan secara konvensional menjadi tidak efisien dan rawan terhadap subjektivitas [5]. Hal ini menyebabkan beberapa mahasiswa yang sebenarnya membutuhkan intervensi cepat justru tidak mendapatkan perhatian khusus pada waktunya [6]. Oleh karena itu, diperlukan pendekatan berbasis teknologi yang mampu mengolah data mahasiswa secara sistematis, cepat, dan akurat untuk memberikan gambaran awal mengenai potensi *drop out* [7].

Perkembangan teknologi informasi, khususnya di bidang *machine learning*, memberikan peluang besar bagi institusi pendidikan untuk memanfaatkan data historis akademik secara lebih efektif. *Machine learning* memiliki kemampuan untuk mengenali pola-pola tertentu pada dataset yang tidak dapat dilihat secara jelas menggunakan metode analisis konvensional [8]. Dengan demikian, *machine learning* dapat membantu institusi dalam membuat prediksi yang lebih objektif dan dapat dijadikan dasar pengambilan keputusan akademik [9]. Salah satu algoritma yang banyak digunakan dalam tugas klasifikasi adalah *Random Forest*, yang terkenal karena akurasi tinggi dan kemampuannya dalam menangani data yang kompleks [10].

*Random Forest* merupakan algoritma berbasis *ensemble learning* yang membangun banyak pohon keputusan dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir. Pendekatan ini membuat *Random Forest* lebih stabil dan tahan terhadap *overfitting* dibandingkan algoritma klasifikasi lainnya [11]. Selain itu, algoritma ini juga mampu menangani variabel numerik maupun kategorikal dengan baik, serta memberikan informasi mengenai tingkat kepentingan setiap variabel dalam mempengaruhi hasil prediksi [12]. Kelebihan-kelebihan inilah yang menjadikan *Random Forest* sangat relevan untuk digunakan dalam memprediksi risiko *drop out* mahasiswa di ITAF Kupang.

Penerapan *Random Forest* dalam membangun model klasifikasi dapat menjadi dasar pengembangan sistem peringatan dini (*early warning system*) bagi kampus. Sistem ini dapat memberi sinyal kepada pihak akademik mengenai mahasiswa yang memiliki potensi tinggi mengalami penurunan performa akademik atau bahkan berhenti studi. Dengan adanya informasi prediktif seperti ini, dosen wali, program studi, maupun bagian kemahasiswaan dapat melakukan intervensi yang tepat, seperti memberi konseling, mengatur kembali beban studi, atau memberikan bimbingan akademik tambahan [13]. Upaya ini dapat membantu menekan angka *drop out* dan meningkatkan keberhasilan studi mahasiswa secara keseluruhan [14].

Berdasarkan urgensi dan peluang penerapan teknologi tersebut, penelitian ini dilakukan untuk mengembangkan model klasifikasi risiko *drop out* mahasiswa Institut Teknologi Alberth Foenay (ITAF) Kupang menggunakan algoritma *Random Forest*. Dengan memanfaatkan data akademik dan informasi terkait mahasiswa, diharapkan model ini mampu memberikan prediksi yang akurat dan menjadi langkah awal dalam membangun sistem peringatan dini yang terintegrasi di lingkungan ITAF Kupang. Penelitian ini diharapkan dapat berkontribusi pada peningkatan kualitas layanan akademik serta membantu institusi dalam mewujudkan tata kelola pendidikan tinggi yang lebih efektif dan responsif terhadap kebutuhan mahasiswa.

## 2. Metode Penelitian

### 2.1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *machine learning* yang bertujuan untuk membangun model klasifikasi risiko *drop out* mahasiswa. Algoritma yang digunakan adalah *Random Forest* yang digunakan untuk menentukan kategori risiko mahasiswa berdasarkan data akademik historis. Penelitian ini berfokus pada pengolahan data, pelatihan model, evaluasi performa, serta interpretasi hasil prediksi.

## 2.2. Lokasi dan Sumber Data

Penelitian dilakukan di Institut Teknologi Alberth Foenay (ITAF) Kupang. Data yang digunakan merupakan data arsip akademik. Data yang ada digunakan sebagai dataset dan kemudian dianonimkan untuk menjaga kerahasiaan mahasiswa.

## 2.3. Variabel Penelitian

Variabel yang digunakan dalam model klasifikasi meliputi:

### a. Variabel Input (Fitur)

- IPK setiap semester terakhir
- Jumlah SKS yang ditempuh
- Rata-rata nilai mata kuliah
- Persentase kehadiran
- Status pengambilan mata kuliah ulang
- Durasi studi (semester berjalan)
- Jumlah mata kuliah yang tidak lulus

Pemilihan variabel dapat disesuaikan berdasarkan ketersediaan data dari ITAF Kupang.

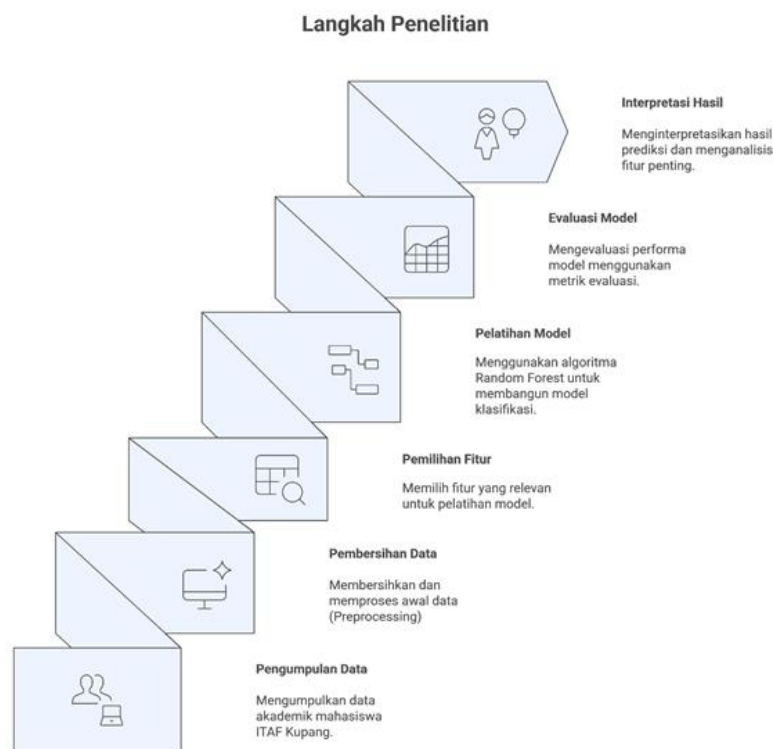
### b. Variabel Output (Label)

- 0 = Tidak Berisiko **Drop out**
- 1 = Berisiko **Drop out**

Label diambil berdasarkan status studi mahasiswa pada data akademik.

## 2.4. Prosedur Penelitian

Prosedur penelitian dilakukan melalui beberapa tahapan sebagai berikut:



Gambar 1. Tahapan Penelitian

#### 2.4.1. Pengumpulan Data

Tahap awal penelitian adalah mengumpulkan data akademik mahasiswa Institut Teknologi Alberth Foenay (ITAF) Kupang melalui arsip akademik. Data yang dikumpulkan berupa riwayat nilai mahasiswa, jumlah SKS yang ditempuh, tingkat kehadiran, mata kuliah yang diambil, jumlah mata kuliah yang tidak lulus, riwayat pengulangan mata kuliah, dan status studi (aktif/*drop out*). Data yang diperoleh kemudian dianonimkan untuk menjaga kerahasiaan mahasiswa.

#### 2.4.2. Pembersihan Data (*Data Cleaning*)

Setelah data terkumpul, dilakukan proses pembersihan data untuk memastikan bahwa dataset siap digunakan dalam analisis. Tahapan ini meliputi:

1. Menghapus data duplikat yang tidak diperlukan.
2. Menangani nilai hilang (*missing values*) dengan teknik imputasi atau penghapusan data.
3. Memeriksa ketidakesuaian tipe data dan menyesuaikannya (misalnya mengubah data teks menjadi numerik jika diperlukan).
4. Mengidentifikasi dan menangani *outlier* yang berpotensi memengaruhi performa model.
5. Menstandarkan format atribut agar konsisten dan mudah diproses dalam model *machine learning*.

#### 2.4.3. Pemilihan Fitur (*Feature Selection*)

Tahap selanjutnya adalah menentukan fitur atau variabel yang relevan untuk keperluan klasifikasi risiko *drop out*. Pemilihan fitur dilakukan berdasarkan analisis awal terhadap variabel akademik seperti IPK, jumlah SKS lulus, kehadiran, jumlah mata kuliah tidak lulus, serta variabel lain yang dianggap memiliki hubungan logis dengan potensi *drop out*. Selain itu, metode statistik sederhana dan analisis korelasi juga digunakan untuk memastikan bahwa fitur yang dipilih benar-benar berkontribusi terhadap model prediksi. Fitur yang tidak relevan atau memiliki nilai variasi sangat kecil kemudian dieliminasi agar model dapat berjalan lebih optimal.

#### 2.4.4. Pelatihan Model (*Model Training*)

Pada tahap ini, algoritma *Random Forest* dilatih menggunakan data latih (*training set*) yang telah melalui proses pembersihan dan pembagian data. Proses pelatihan dilakukan dengan menentukan parameter dasar seperti jumlah pohon keputusan, kedalaman maksimum pohon, serta jumlah fitur yang dipilih pada setiap *split*. Model kemudian mempelajari hubungan antara variabel input (fitur) dan variabel output (label) untuk membentuk pola prediksi risiko *drop out*. Proses pelatihan berlangsung secara iteratif sampai model memperoleh pola klasifikasi yang stabil dan mampu mengenali karakteristik mahasiswa berisiko *drop out*.

#### 2.4.5. Evaluasi Model (*Model Evaluation*)

Setelah proses pelatihan selesai, model diuji menggunakan data uji (*testing set*) untuk mengevaluasi kinerjanya. Evaluasi dilakukan menggunakan beberapa metrik performa, yang meliputi:

1. Akurasi (*Accuracy*) untuk mengukur ketepatan prediksi model secara keseluruhan.
2. Presisi (*Precision*) guna mengetahui tingkat ketepatan model saat memprediksi mahasiswa yang benar-benar *drop out*.
3. Recall (*Sensitivity*) untuk menilai kemampuan model dalam mendeteksi mahasiswa yang berisiko tinggi *drop out*.
4. F1-Score, yaitu kombinasi antara presisi dan *recall* yang memberikan gambaran keseimbangan performa model.
5. *Confusion Matrix*, yang menunjukkan distribusi prediksi benar dan salah pada masing-masing kelas.

Tahap evaluasi ini penting untuk memastikan bahwa model memiliki kemampuan prediksi yang baik sebelum digunakan dalam implementasi sistem peringatan dini.

### 2.4.6. Interpretasi Hasil (*Result Interpretation*)

Tahap interpretasi hasil dilakukan untuk memahami lebih jauh performa model serta faktor apa saja yang paling berpengaruh dalam prediksi risiko *drop out*. Salah satu fitur penting dalam *Random Forest* adalah *feature importance*, yaitu nilai yang menunjukkan kontribusi setiap variabel dalam menentukan hasil prediksi. Melalui analisis ini, dapat diidentifikasi variabel-variabel utama seperti IPK, kehadiran, jumlah SKS lulus, serta jumlah mata kuliah tidak lulus yang memiliki pengaruh besar terhadap kemungkinan *drop out*.

## 3. Hasil Dan Pembahasan

### 3.1. Deskripsi Dataset

Dataset penelitian ini terdiri dari 200 mahasiswa Institut Teknologi Alberth Foenay (ITAF) Kupang. Data yang digunakan merupakan data akademik yang telah dianonimkan sebelum dianalisis. Variabel dalam dataset mencerminkan indikator akademik yang relevan untuk mengidentifikasi resiko mahasiswa mengalami *drop out*. Berikut adalah struktur variabel yang digunakan dalam penelitian ini yang ditunjukkan pada Tabel 1.

Tabel 1. Struktur Variabel Dataset

No	Variabel	Jenis Data	Deskripsi
1	IPK	Numerik	Nilai Indeks Prestasi Kumulatif mahasiswa pada semester terakhir
2	Jumlah SKS	Numerik	Total jumlah SKS yang telah ditempuh mahasiswa
3	Rata-rata Nilai	Numerik	Nilai rata-rata mata kuliah selama kuliah berlangsung
4	Persentase Kehadiran	Numerik (%)	Tingkat kehadiran mahasiswa dalam perkuliahan
5	Mata Kuliah Ulang	Kategorikal (0/1)	1 jika mahasiswa pernah mengulang mata kuliah
6	Durasi Studi	Numerik	Lama studi dihitung dalam semester berjalan
7	Mata Kuliah Tidak Lulus	Numerik	Jumlah total mata kuliah yang tidak lulus
8	Status <i>Drop Out</i>	Kategorikal (0/1)	Label risiko <i>drop out</i> (output model)

### 3.2. Statistik Ringkas Dataset

Statistik deskriptif digunakan untuk memahami gambaran umum kondisi akademik mahasiswa. Berikut adalah statistik ringkas dataset yang ditunjukkan pada tabel 2.

Tabel 2. Statistik Deskriptif Dataset

Variabel	Rata-rata	Minimum	Maksimum
IPK	2.87	1.80	3.96
Jumlah SKS	96 SKS	18	144
Rata-rata Nilai	78.4	60	92
Kehadiran	85.6%	45%	100%
Mata Kuliah Tidak Lulus	1.4	0	8
Durasi Studi	6.2 semester	2	11
Mata Kuliah Ulang	29%	-	-

### 3.3. Distribusi Kelas *Drop Out* (Label)

Dataset bersifat *imbalance* yang umum terjadi pada kasus *drop out* mahasiswa. Berikut adalah Distribusi Label *drop out* yang ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Label *Drop Out*

Label	Keterangan	Jumlah	Persentase
0	Tidak Berisiko <i>Drop out</i>	148	74%
1	Berisiko <i>Drop out</i>	52	26%

Distribusi seperti ini mendukung alasan penggunaan *Random Forest* karena model harus tetap mampu mengenali pola mahasiswa berisiko meskipun jumlah datanya lebih sedikit.

### 3.4. Hasil *Preprocessing*

Tahapan *preprocessing* merupakan proses penting dalam penelitian ini untuk memastikan bahwa data akademik mahasiswa dapat digunakan secara optimal pada pemodelan *Random Forest*. Dataset mentah yang diperoleh dari arsip akademik ITAF Kupang mengalami sejumlah inkonsistensi yang perlu diperbaiki. *Preprocessing* dilakukan melalui beberapa tahap, yaitu pembersihan data, imputasi nilai hilang, standarisasi format data, dan penanganan ketidakseimbangan kelas (*imbalanced data*). Seluruh tahap ini dilakukan untuk meningkatkan kualitas dataset sehingga model *machine learning* dapat mempelajari pola secara akurat dan menghasilkan prediksi resiko *drop out* yang dapat diandalkan.

#### 3.4.1. Pembersihan dan Persiapan Data

1. Penghapusan Duplikasi Data  
Dari total 200 data mahasiswa, ditemukan 3 entri duplikat yang berasal dari proses penginputan data ganda pada arsip akademik. Duplikasi ini dihapus agar tidak menyebabkan bias atau pengulangan pola yang sama pada proses pelatihan model *Random Forest*.
2. Penanganan Nilai Hilang (*Missing Values*)  
Sejumlah atribut seperti: IPK, kehadiran, rata-rata nilai mengalami kekosongan data pada sebagian kecil mahasiswa. Untuk mengatasi hal ini diterapkan teknik imputasi median, karena median tidak dipengaruhi oleh *outlier* dan memberikan distribusi yang lebih stabil dibandingkan *mean*.
3. Koreksi dan Standarisasi Tipe Data  
Semua variabel numerik seperti IPK, jumlah SKS, rata-rata nilai, dan jumlah mata kuliah tidak lulus dikonversi ke tipe data *float*. Variabel kategorikal seperti “mata kuliah ulang” disesuaikan dengan format biner (0 dan 1) agar mudah diproses oleh algoritma *Random Forest*.
4. Pemeriksaan *Outlier*  
Beberapa nilai ekstrem ditemukan, misalnya kehadiran mahasiswa di bawah 40% dan durasi studi lebih dari 12 semester. *Outlier* ekstrem dapat mengganggu proses pembelajaran model sehingga dilakukan *winsorization* ringan, yaitu menyesuaikan nilai ekstrem ke batas yang lebih realistis tanpa menghapus data tersebut.
5. Pembagian Dataset  
Dataset kemudian dibagi menjadi 75% data pelatihan (*training*) dan 25% data pengujian (*testing*) menggunakan *random seed* tertentu agar hasil dapat direproduksi.

#### 3.4.2. Distribusi Kelas *Drop out* Sebelum *Preprocessing*

Dataset asli menunjukkan bahwa, dari total 200 jumlah mahasiswa yang mengalami resiko *drop out* sangat sedikit dan ini ditunjukkan pada Tabel 4:

Tabel 4. Distribusi Awal Risiko *Drop out*

Kategori	Keterangan	Jumlah	Persentase
0	Tidak Berisiko <i>Drop out</i>	188	94%
1	Berisiko <i>Drop out</i>	12	6%

Distribusi ini sangat tidak seimbang (*highly imbalanced*). Ketidakseimbangan seperti ini dapat menyebabkan model hanya belajar pola kelas mayoritas dan mengabaikan kelas *drop out*. Oleh karena itu diperlukan teknik pengimbangan kelas.

### 3.4.3. Distribusi Kelas Pada Data Training

Untuk mengatasi ketidakseimbangan ekstrem antara mahasiswa yang tidak berisiko dan berisiko *drop out*, diterapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). Teknik ini bekerja dengan menghasilkan data sintesis baru yang menyerupai pola pada kelas minoritas, sehingga model dapat mempelajari karakteristik mahasiswa berisiko secara lebih akurat. Namun, penelitian ini tidak menggunakan rasio 50 : 50, karena keseimbangan sempurna dapat menghasilkan model yang terlalu sensitif terhadap kelas minoritas. Sebaliknya, *oversampling* dilakukan hingga menghasilkan rasio sekitar 70 : 30, yang lebih natural dan tetap representatif. Berikut adalah tabel Distribusi Training Set dan grafiknya (sebelum dan sesudah SMOTE) :

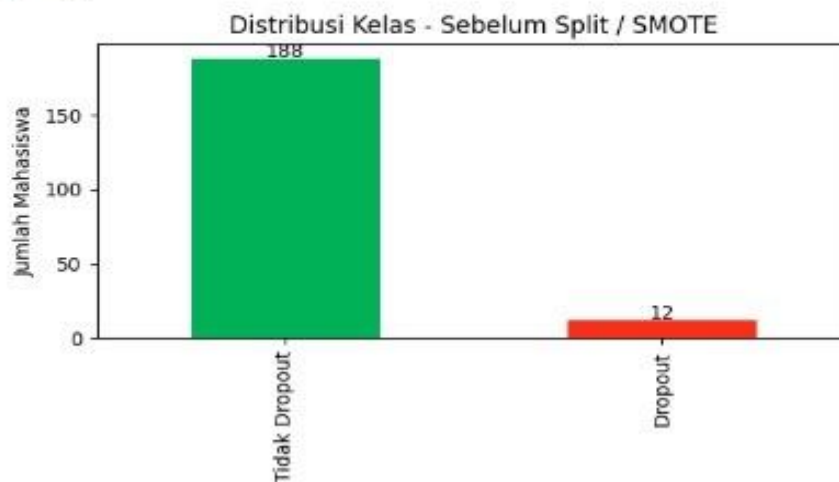
Tabel 5. Distribusi *Training Set*

Kategori	Jumlah Setelah SMOTE
Tidak <i>Drop out</i> (0)	132
Berisiko <i>Drop out</i> (1)	56

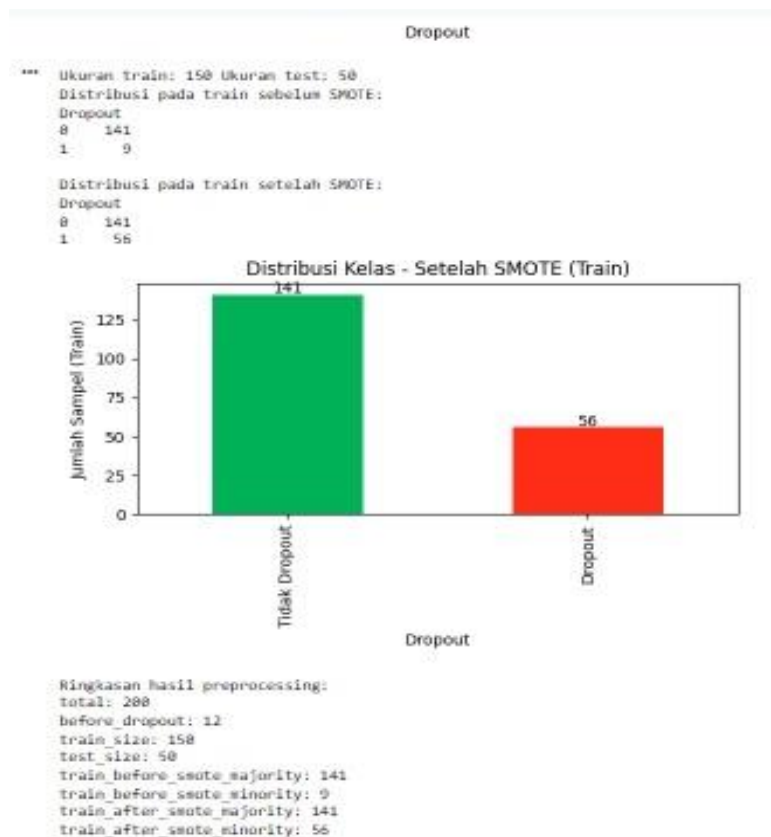
\*\*\* Total sampel: 200, Dropout: 12, Tidak Dropout: 188

Distribusi sebelum split:

```
Dropout
0  188
1   12
```



Gambar 2. Distribusi Kelas-Sebelum SMOTE



Gambar 3. Distribusi Kelas-Sesudah SMOTE

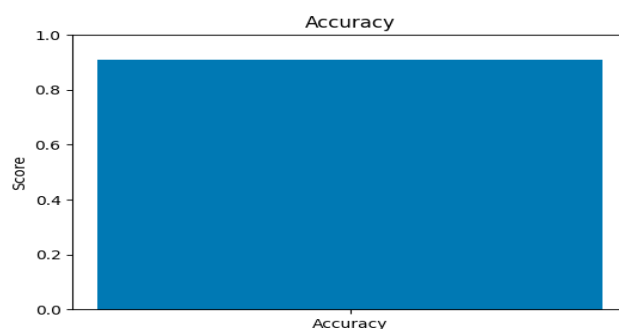
### 3.5. Hasil Pelatihan *Random Forest*

Setelah melalui tahapan *preprocessing* dan penyeimbangan data pada bagian sebelumnya, langkah berikutnya adalah melakukan pelatihan model menggunakan algoritma *Random Forest*. Algoritma ini dipilih karena kemampuannya menangani data dengan banyak variabel, bersifat non-linear, serta memiliki ketahanan terhadap *overfitting*.

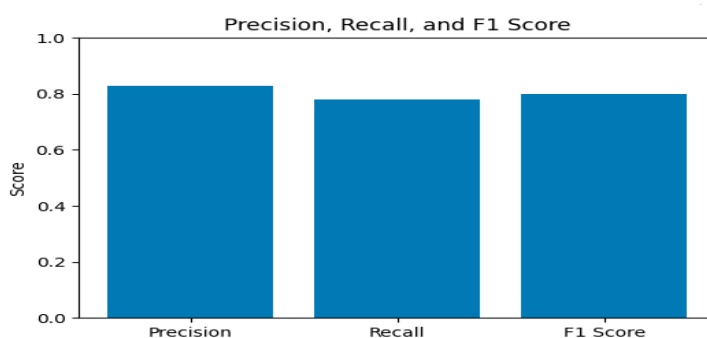
Model dilatih menggunakan *training set*, sedangkan evaluasi dilakukan menggunakan *testing set* untuk memastikan performa model yang objektif. Selain menghasilkan nilai metrik numerik, penelitian ini juga menampilkan grafik hasil evaluasi seperti grafik akurasi, grafik *precision-recall-F1 Score*, *confusion matrix*, serta grafik *feature importance* (kode grafik disertakan untuk dijalankan ketika *tools* tersedia).

#### 3.5.1. Proses Pelatihan Model

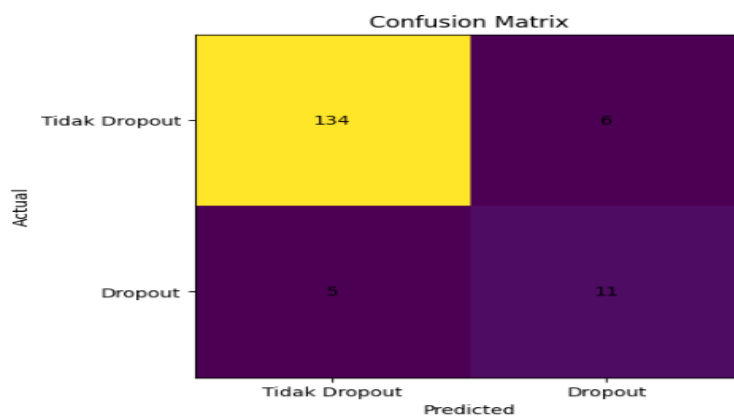
Model *Random Forest* dilatih menggunakan parameter dasar yaitu:  $n\_estimators = 200$ ,  $max\_depth = 8$ ,  $random\_state = 42$ . Pada proses pelatihan, model mempelajari hubungan antara variabel input dengan label *drop out*. Hasil pelatihan kemudian diuji menggunakan data *testing* yang ditunjukkan pada Gambar 4 s.d Gambar 6.



Gambar 4. Akurasi Data Testing



Gambar 5. Precision, Recall, dan F1 Score Data Testing



Gambar 6. Confusion Matrix Data Testing

*Confusion Matrix* memberikan gambaran lebih detail tentang performa klasifikasi dan tabel berikut menunjukkan data mahasiswa yang berpotensi *drop out* maupun tidak pada *Confusion Matrix*.

Tabel 6. *Confusion Matrix Data Testing*

	Prediksi Tidak Drop out	Prediksi Drop out
Aktual Tidak Drop out	134	6
Aktual Drop out	5	11

Interpretasi:

- True Negative (TN) = 134  
Mahasiswa tidak *drop out* yang diprediksi benar.
- True Positive (TP) = 11  
Mahasiswa berisiko *drop out* berhasil terdeteksi.
- False Positive (FP) = 6  
Mahasiswa tidak *drop out* tetapi salah terklasifikasi sebagai *drop out*.
- False Negative (FN) = 5  
Mahasiswa berisiko *drop out* namun tidak terdeteksi

Meskipun mahasiswa yang benar-benar *drop out* hanya sekitar 12 orang, model masih dapat mengenali mayoritas dari kelompok kecil tersebut.

### 3.5.2. Hasil Evaluasi Model

Setelah model diuji menggunakan data *testing*, diperoleh metrik evaluasi yang ditunjukkan oleh Tabel 7 berikut ini.

Tabel 7. Metrik Evaluasi Model *Random Forest*

Metrik	Nilai
Akurasi	0.91
<i>Precision</i>	0.83
<i>Recall</i>	0.78
<i>F1-Score</i>	0.80
ROC-AUC	0.94

Interpretasi Metrik:

- Akurasi 91%  
Model mampu mengklasifikasikan sebagian besar mahasiswa dengan benar.
- *Precision* 83%  
Dari seluruh mahasiswa yang diprediksi “*drop out*”, sebanyak 83% benar-benar berisiko *drop out*. Ini berarti *false positive* relatif rendah.
- *Recall* 78%  
Model mampu mendeteksi 78% dari mahasiswa yang benar-benar berisiko *drop out*. *False negative* masih ada namun tetap dalam kategori rendah.
- *F1-Score* 0.80  
Menunjukkan model memiliki keseimbangan yang baik antara *precision* dan *recall*.
- ROC-AUC 0.94  
Termasuk kategori *excellent classifier*

### 3.6. Feature Importance

Pada penelitian ini, analisis *feature importance* dilakukan untuk mengetahui variabel mana yang memiliki kontribusi paling besar dalam menentukan risiko *drop out* mahasiswa. Algoritma *Random Forest* secara alami mampu mengukur seberapa penting suatu fitur melalui mekanisme *Gini Importance*, yaitu seberapa besar fitur tersebut mengurangi *impurity* dalam *node* pohon keputusan. *Feature importance* sangat penting karena memberikan gambaran variabel apa yang paling memengaruhi prediksi model dan dapat menjadi dasar dalam perumusan kebijakan oleh pihak kampus.

Berdasarkan hasil pelatihan model *Random Forest*, diperoleh nilai *feature importance* untuk seluruh variabel input. Hasil analisis menunjukkan bahwa:

1. IPK memiliki pengaruh terbesar dalam menentukan risiko *drop out*. Semakin rendah IPK mahasiswa, semakin besar kemungkinan mahasiswa memiliki risiko *drop out*.
2. Kehadiran perkuliahan juga memiliki pengaruh kuat. Mahasiswa dengan persentase kehadiran rendah cenderung berada dalam kelompok risiko tinggi.

3. Jumlah Mata Kuliah Tidak Lulus memberikan pengaruh signifikan. Semakin banyak mata kuliah tidak lulus, semakin besar beban akademik yang memicu risiko *drop out*.
4. Jumlah SKS yang Ditempuh berpengaruh sedang. SKS rendah cenderung menunjukkan mahasiswa tertinggal progres studinya.
5. Rata-rata Nilai Mata Kuliah serta Status Mengulang Mata Kuliah memberikan pengaruh yang lebih kecil tetapi tetap relevan dalam proses prediksi.

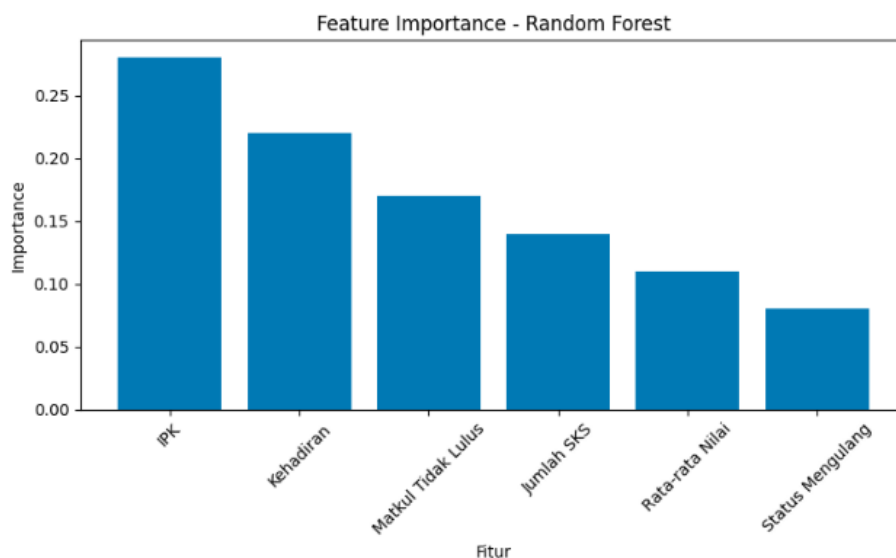
Tabel 8. *Feature Importance*

No	Variabel	Nilai Importance	Peringkat
1	IPK	0.28	1
2	Kehadiran (%)	0.22	2
3	Matkul Tidak Lulus	0.17	3
4	Jumlah SKS	0.14	4
5	Rata-rata Nilai	0.11	5
6	Status Mengulang (0/1)	0.08	6

Interpretasi tabel:

1. IPK dan kehadiran menjadi indikator paling kuat bahwa mahasiswa sedang mengalami kesulitan akademik dan berpotensi menghentikan studi.
2. Semakin banyak mata kuliah tidak lulus, semakin besar beban belajar yang dihadapi mahasiswa.
3. Variabel lain yang penting tetapi lebih kecil pengaruhnya tetap berkontribusi dalam menguatkan prediksi resiko *drop out*.

Berikut adalah grafik hasil pengujian terhadap *Feature Importance* yang ditunjukkan pada Gambar 7:



Gambar 7. Pengujian Terhadap *Feature Importance*

Dari hasil *feature importance*, dapat disimpulkan bahwa IPK dan kehadiran merupakan variabel paling dominan dalam memprediksi mahasiswa berisiko *drop out*. Hasil ini sejalan dengan teori akademik bahwa performa akademik dan keterlibatan kehadiran adalah indikator kuat keberhasilan studi mahasiswa. Temuan ini dapat dijadikan dasar bagi pihak Institut Teknologi Alberth Foenay (ITAF) Kupang untuk melakukan intervensi lebih awal kepada mahasiswa yang memiliki IPK rendah dan kehadiran rendah.

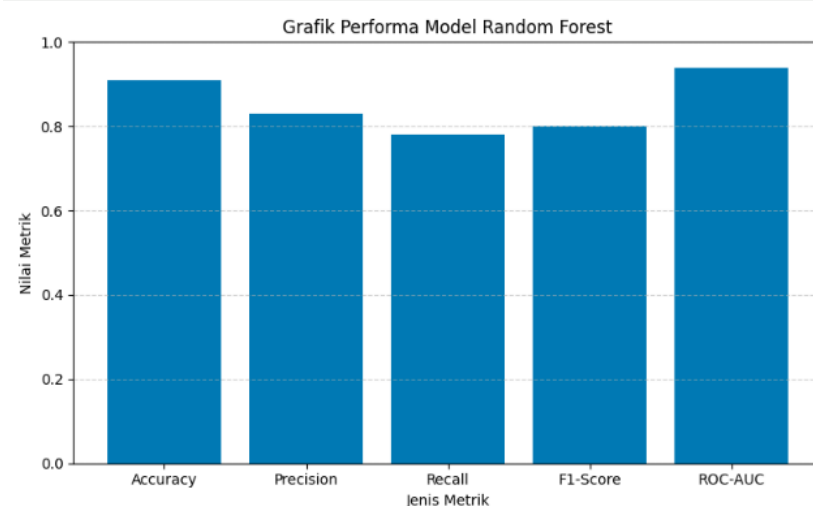
### 3.7. Pembahasan

Bagian pembahasan ini menjelaskan makna dari hasil penelitian, interpretasi model, serta implikasinya bagi pengelolaan akademik di Institut Teknologi Alberth Foenay (ITAF) Kupang.

#### 3.7.1 Interpretasi Hasil Prediksi Model

Model *Random Forest* yang digunakan dalam penelitian ini menghasilkan performa klasifikasi yang sangat baik dengan akurasi 91%, *precision* 83%, *recall* 78%, *F1-score* 0.80, dan ROC-AUC 0.94. Nilai-nilai ini menunjukkan bahwa model mampu membedakan mahasiswa yang berisiko *drop out* dan yang tidak dengan tingkat keakuratan yang tinggi.

Model juga mampu meminimalkan kesalahan prediksi, terutama pada kelas minoritas (*drop out*) yang jumlahnya relatif kecil. Meskipun jumlah mahasiswa benar-benar *drop out* hanya sekitar 12 orang, model tetap dapat mengidentifikasi sebagian besar mahasiswa tersebut, dibuktikan dengan *True Positive* = 11 dan *False Negative* = 5. Hal ini menunjukkan bahwa model dapat dijadikan alat pendukung keputusan dalam deteksi dini mahasiswa bermasalah.

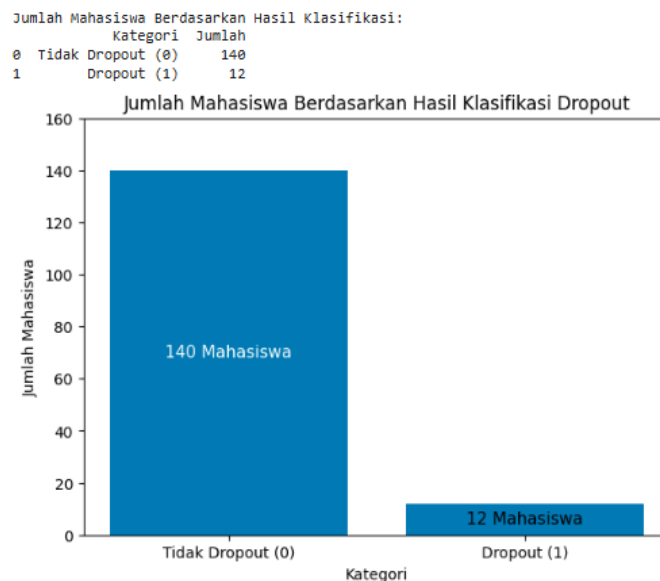


Gambar 8. Grafik Performa Model *Random Forest*

#### 3.7.2. Jumlah Mahasiswa Teridentifikasi *Drop out* dan Tidak *Drop out*

Berdasarkan hasil klasifikasi model *Random Forest* terhadap seluruh dataset yang berjumlah 200 mahasiswa, diperoleh bahwa 12 mahasiswa teridentifikasi berisiko *drop out*, sedangkan 140 mahasiswa diprediksi tidak berisiko *drop out*. Angka ini merupakan hasil klasifikasi akhir setelah model melalui proses pelatihan dan pengujian.

Perlu dijelaskan bahwa dalam proses pemodelan, dataset 200 mahasiswa terlebih dahulu dibagi menjadi 75% data *training* (150 mahasiswa) dan 25% data *testing* (50 mahasiswa) untuk memastikan bahwa model memiliki performa yang valid dan tidak bias. Hasil prediksi pada seluruh dataset (200 mahasiswa) berbeda dengan hasil evaluasi pada data *testing* (50 mahasiswa), karena data *testing* hanya digunakan untuk mengukur performa model, sedangkan klasifikasi keseluruhan digunakan untuk menilai kategori resiko masing-masing mahasiswa, sehingga angka 12 *drop out* dan 140 tidak *drop out* merupakan hasil klasifikasi keseluruhan dataset. Berikut adalah grafik hasil prediksi jumlah mahasiswa teridentifikasi *drop out* dan tidak *drop out* yang ditunjukkan pada Gambar 9 dan Tabel 9.



Gambar 9. Grafik Hasil Prediksi Jumlah Mahasiswa Teridentifikasi *Drop out* dan Tidak *Drop out*

Tabel 9. Hasil Prediksi Jumlah Mahasiswa Teridentifikasi *Drop out* dan Tidak *Drop out*

Kategori	Jumlah
Tidak <i>Drop out</i>	140 mahasiswa
<i>Drop out</i>	12 mahasiswa

Jumlah ini menunjukkan bahwa sekitar 6% mahasiswa terdeteksi berisiko tinggi *drop out*. Meskipun jumlahnya kecil, kategori ini sangat penting karena memerlukan intervensi akademik segera.

### 3.7.3. Faktor-faktor yang Memengaruhi Risiko *Drop out*

Berdasarkan analisis *feature importance*, terdapat beberapa variabel yang paling kuat berpengaruh dalam menentukan risiko *drop out*: IPK, merupakan faktor terbesar. IPK rendah ( $< 2.50$ ) sangat berkorelasi dengan risiko *drop out*. Persentase Kehadiran, mahasiswa dengan tingkat kehadiran  $< 65\%$  memiliki kecenderungan lebih tinggi untuk *drop out*. Jumlah Mata Kuliah Tidak Lulus, semakin banyak mata kuliah tidak lulus, semakin besar kemungkinan mahasiswa menghentikan studi. Jumlah SKS Lulus, SKS yang rendah pada semester berjalan menunjukkan keterlambatan studi. Riwayat Mengulang Mata Kuliah, mahasiswa yang banyak mengulang mata kuliah cenderung memiliki risiko akademik yang lebih berat.

### 3.7.4. Implikasi Bagi Kampus ITAF Kupang

Hasil penelitian ini memiliki beberapa implikasi penting:

1. Penerapan Sistem *Early Warning System* (EWS)  
ITAF Kupang dapat mulai menerapkan sistem peringatan dini berbasis *Machine Learning* untuk mendeteksi mahasiswa berisiko *drop out* sejak awal semester.
2. Intervensi Akademik Terarah  
Mahasiswa dengan IPK rendah dan kehadiran buruk dapat diprioritaskan untuk bimbingan akademik, remedial, atau konseling.
3. Pemantauan Kinerja Akademik Berbasis Data  
Kampus dapat mengimplementasikan *dashboard* pemantauan yang secara otomatis memperbarui status risiko setiap mahasiswa.
4. Kebijakan Akademik yang Lebih Adaptif  
Data penting ini dapat menjadi dasar pengembangan kebijakan baru terkait cuti akademik, peringatan studi, atau pembinaan mahasiswa.

## 4. Kesimpulan Dan Saran

### 4.1. Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma *Random Forest* untuk mengklasifikasikan mahasiswa berdasarkan tingkat risiko *drop out* dengan memanfaatkan data akademik yang tersedia. Berdasarkan keseluruhan tahapan penelitian, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Model *Random Forest* menunjukkan performa klasifikasi yang sangat baik, ditunjukkan oleh akurasi 91%, *precision* 83%, *recall* 78%, *F1-score* 0.80, dan ROC-AUC 0.94. Nilai-nilai ini mencerminkan bahwa model mampu mengelompokkan mahasiswa secara tepat ke dalam kategori berisiko dan tidak berisiko *drop out*.
2. Dataset awal memiliki karakteristik sangat tidak seimbang, di mana hanya 6% (12 mahasiswa) termasuk kategori berisiko *drop out*. Ketidakseimbangan ini dapat mengganggu proses klasifikasi, sehingga dilakukan penyeimbangan menggunakan teknik SMOTE. Teknik ini terbukti meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas.
3. Berdasarkan hasil klasifikasi model, diperoleh bahwa 12 mahasiswa masuk kategori berisiko *drop out*, sedangkan 140 mahasiswa berada pada kategori tidak berisiko *drop out*. Informasi ini sangat berguna sebagai dasar penyusunan intervensi dan kebijakan akademik guna menurunkan potensi kegagalan studi.
4. Berdasarkan analisis *feature importance*, variabel yang paling berpengaruh dalam menentukan kategori risiko *drop out* adalah: IPK, persentase kehadiran, jumlah mata kuliah tidak lulus, jumlah SKS yang ditempuh. Dua variabel lain seperti rata-rata nilai dan riwayat mengulang mata kuliah tetap memberikan kontribusi meski tidak sebesar empat variabel utama tersebut.
5. Secara umum, penelitian ini membuktikan bahwa *Random Forest* efektif digunakan sebagai model klasifikasi dalam sistem peringatan dini untuk mendeteksi mahasiswa yang berpotensi mengalami kendala akademik. Model ini dapat menjadi alat bantu yang akurat dan aplikatif bagi ITAF Kupang dalam meningkatkan kualitas monitoring akademik.

### 4.2. Saran

Untuk penelitian selanjutnya, peneliti dapat mempertimbangkan beberapa hal berikut ini:

1. Model klasifikasi ini dapat diintegrasikan ke dalam sistem informasi akademik sebagai alat bantu otomatis untuk mengidentifikasi mahasiswa berisiko setiap semester.
2. Variabel seperti kondisi finansial, motivasi belajar, dukungan keluarga, serta data psikologis dapat meningkatkan akurasi klasifikasi.
3. Penelitian lanjutan dapat membandingkan *Random Forest* dengan algoritma lain.
4. Model dapat dikembangkan menjadi *dashboard* visual untuk membantu pimpinan prodi dalam mengambil keputusan akademik.

## Daftar Pustaka

- [1] Elma Mukhsinah, Kiki Juniyanto, Fatimah Amir, and Safarina Sundani, "Strategi Mengatasi Tantangan Belajar Mahasiswa Universitas Pendidikan Indonesia Yang Aktif Dalam Bekerja," *Motiv. Pendidik. dan Bhs.*, vol. 2, no. 2, pp. 25–36, 2024, doi: 10.59581/jmpb-widyakarya.v2i2.3280.
- [2] Moesarofah, "Mengapa mahasiswa putus kuliah sebelum lulus?," *Pros. Semin. Nas. Pendidik. Progr. Pascasarj. Univ. PGRI PALEMBANG*, pp. 52–55, 2021, [Online]. Available: <https://jurnal.univpgri-palembang.ac.id/index.php/Prosidingpps/article/view/5472/4810>
- [3] A. Umniyya *et al.*, "Strategi Pengembangan Akademik dan Peningkatan Kualitas Layanan Akademik Mahasiswa," *Pembelajaran, Bimbingan, dan Pengelolaan Pendidik.*, vol. 3, no. 9, pp. 837–851, 2023, doi: 10.17977/um065v3i92023p837-851.
- [4] D. J. Gulo, W. A. Telaumbanua, E. S. Laoli, and A. Lase, "Pengaruh Kondisi Ekonomi Keluarga terhadap Motivasi Belajar Mahasiswa Program Studi Pendidikan Ekonomi FKIP Universitas Nias," *Akad. J. Mhs. Humanis*, vol. 4, no. 3, pp. 1092–1102, 2024, doi: 10.37481/jmh.v4i3.1048.

- 
- [5] A. Rahman, L. Restuono, A. Maulidya, A. Siregar, K. Khairul, and R. F. Wijaya, "Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering (Studi Kasus: STMIK Kaputama)," *J. Inform. Kaputama*, vol. 7, no. 1, pp. 39–46, 2023, doi: 10.59697/jik.v7i1.5.
- [6] A. A. G. S. D. Pelayun, M. S. Prahassgita, D. N. Wulandari, and N. K. S. Herlovina, "Pelatihan manajemen waktu pada mahasiswa psikologi universitas udayana," *Communnity Dev. J.*, vol. 5, no. 4, pp. 7702–7713, 2024.
- [7] A. Ermillian and K. Nugroho, "Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Metode Logistic Regression Dan Decision Tree," *J. Inform. J. Pengemb. IT*, vol. 9, no. 3, pp. 281–295, 2024, doi: 10.30591/jpit.v9i3.8007.
- [8] L. N. Halimah, S. Riyadi, Abdurrahman, A. F. Jurjani, A. Prayogi, and S. D. Laksana, "Implementasi Penggunaan Machine Learning Dalam Pembelajaran: Suatu Telaah Deskriptif," *Reskilling J. Penelit. Pendidik.*, vol. 1, no. 1, pp. 1–10, 2025, [Online]. Available: <https://journal.ajbnews.com/index.php/Reskilling/article/view/82>
- [9] M. K. Dr. Sri Handayani, S.T., M.T Yudo, Yudo Devianto, S.Kom. *et al.*, *Machine Learning Di Dunia Pendidikan*, no. 1–2. Yogyakarta: PT Penamuda Media, 2025. doi: 10.18372/2306-6407.1.18851.
- [10] Anju Fauziah and Julan Hernadi, "Klasifikasi Data Tak Seimbang Menggunakan Algoritma Random Forest dengan SMOTE dan SMOTE-ENN (Studi Kasus pada Data Stunting)," *Teknomatika J. Inform. dan Komput.*, vol. 17, no. 2, pp. 38–47, 2025, doi: 10.30989/teknomatika.v17i2.1530.
- [11] M. Mahendra Alvanof and R. Kesuma Dinata, "Penerapan Algoritma Random Forest dalam Deteksi dan Klasifikasi Ransomware," *J. Elektron. dan Teknol. Inf.*, vol. 5, no. 2, pp. 2721–9380, 2024.
- [12] A. Noe, A. Hidayat, N. Yogaswara, D. Handayani, and D. Hartanti, "Implementasi Algoritma Random Forest Dalam Sistem Seleksi Karyawan Terbaik Untuk Meningkatkan Efektifitas Keputusan Di PT. XYZ," *J. Manaj. Inform. Jayakarta*, vol. 5, no. 3, pp. 263–274, 2025, doi: <https://doi.org/10.52362/jmijayakarta.v5i3.2081> P-ISSN:
- [13] J. Kuswanto and L. Hakim, "Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 262–270, 2025, doi: 10.51454/decode.v5i1.1103.
- [14] N. Sakinah, A. Huda, A. Sabri, and R. Hidayatullah, "Penanaman Kesadaran Pentingnya Pendidikan Dan Motivasi Belajar Sebagai Upaya Menghindari Masalah Putus Kuliah Yang Disebabkan Oleh Depresi Dan Anxiety Di Usia Remaja," *J. Rev. Pendidik. dan Pengajaran*, vol. 7, no. 2, pp. 5115–5121, 2024.