

## PENERAPAN ALGORITMA *RANDOM FOREST* UNTUK KLASIFIKASI PREDIKAT KELULUSAN MAHASISWA UNIVERSITAS KHAIRUN

Kamila Aprilia<sup>1</sup>, Rosihan<sup>2</sup>, Saiful Do Abdullah<sup>3</sup>, She Turuy<sup>4</sup>, Alfanugrah A. Hi. Usmanullah<sup>5</sup>

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Khairun  
Jl. Jati Metro, Kota Ternate Selatan<sup>12345</sup>

Email: kamilaaprilia@gmail.com<sup>1</sup>, rosihan@unkhair.ac.id<sup>2</sup>, saiful.abdullah@unkhair.ac.id<sup>3</sup>  
seh.turuy@unkhair.ac.<sup>4</sup>, nugrahalfa@gmail.com<sup>5</sup>

### Abstrak

Penelitian ini bertujuan untuk menerapkan algoritma *Random Forest* dalam mengklasifikasikan predikat kelulusan mahasiswa Universitas Khairun ke dalam tiga kategori, yaitu pujian, sangat memuaskan, dan memuaskan. Metode penelitian yang digunakan meliputi pengumpulan data kelulusan mahasiswa sebanyak 1.950 data yang diperoleh dari BAAKAPSI dan UPT TIK Universitas Khairun, tahap *data preprocessing* yang mencakup *data cleaning* dan *data transformation* serta implementasi algoritma *Random Forest* menggunakan bahasa pemrograman *Python*. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* untuk mengukur nilai akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* mampu mengklasifikasikan predikat kelulusan mahasiswa dengan sangat baik, dengan nilai akurasi sebesar 100%, *precision* 100%, *recall* 100%, dan *F1-score* 100%. Selain itu, hasil pengujian *black box* pada sistem aplikasi menunjukkan bahwa seluruh fitur utama berjalan dengan baik sesuai dengan kebutuhan pengguna. Dengan demikian, algoritma *Random Forest* terbukti efektif dan dapat digunakan sebagai solusi dalam mendukung pengambilan keputusan terkait predikat kelulusan mahasiswa di Universitas Khairun.

**Kata kunci:** Klasifikasi, Predikat Kelulusan, *Random Forest*, *Confusion Matrix*

## IMPLEMENTATION OF *RANDOM FOREST* ALGORITHM FOR CLASSIFICATION OF STUDENT GRADUATION PREDICTIONS AT KHAIRUN UNIVERSITY

### Abstract

This study aims to apply the *Random Forest* algorithm in classifying the graduation predicate of Khairun University students into three categories, namely praise, very satisfactory, and satisfactory. The research method used includes collecting 1,950 student graduation data obtained from BAAKAPSI and UPT TIK Khairun University, the data preprocessing stage which includes data cleaning and data transformation and the implementation of the *Random Forest* algorithm using the *Python* programming language. Model performance evaluation was carried out using a *confusion matrix* to measure the accuracy, *precision*, *recall*, and *F1-score* values. The results of the study show that the *Random Forest* algorithm is able to classify student graduation predicates very well, with an accuracy value of 100%, *precision* 100%, *recall* 100%, and *F1-score* 100%. In addition, the results of *black box* testing on the application system show that all main features run well according to user needs. Thus, the *Random Forest* algorithm is proven effective and can be used as a solution to support decision making regarding student graduation predicates at Khairun University.

**Keywords:** Classification, Graduation Predicate, *Random Forest*, *Confusion Matrix*.

### 1. PENDAHULUAN

Lembaga Pendidikan tertinggi adalah perguruan tinggi yang merupakan penyelenggara pendidikan akademik bagi mahasiswa [1]. Peningkatan kualitas pendidikan merupakan hal yang paling penting bagi semua lembaga Pendidikan[2]. Salah satu tolak ukur

dari perguruan tinggi yang berkualitas didasarkan pada catatan prestasi akademik bagi mahasiswa, sangat baik selama menempuh studi di perguruan tinggi [3] seperti halnya Universitas Khairun.

Universitas Khairun, sebagai sebuah institusi pendidikan tinggi tentunya menghadapi tantangan dalam memastikan bahwa mahasiswanya dapat

menyelesaikan studi tepat waktu.

Predikat kelulusan adalah kemampuan aktual seorang mahasiswa yang berwujud penguasaan ilm pengetahuan, sikap dan keterampilan untuk mencapai tujuan akhir dari proses belajar [4]. yang diharapkan menghasilkan peserta didiknya yang lebih cerdas, inovatif, dan kompetitif [5].

Sebaliknya prestasi akademik mahasiswa yang rendah dilihat sebagai ketidakmampuan pihak perguruan tinggi dalam melaksanakan proses belajar dengan baik, Kegagalan atau prestasi akademik mahasiswa yang rendah perlu menjadi perhatian bagi Lembaga perguruan tinggi, guna menghasilkan lulusan berkualitas [3]. Bentuk keberhasilan studi yang didapatkan mahasiswa diukur berdasarkan lama studi dan indeks prestasi kumulatif (IPK) [2].

Random Forest merupakan metode yang dikembangkan dari metode CART (Classification and Regression Trees), yaitu dengan menerapkan metode bootstrap aggregating (bagging) dan random feature selection [6] CART merupakan metode eksplorasi data berdasarkan teknik pohon keputusan. Regresi pohon dihasilkan ketika respons peubah memiliki data numerik, klasifikasi pohon dihasilkan ketika respons peubah berisi data kategorikal [7].

Berdasarkan permasalahan diatas, penulis melakukan penelitian yaitu “Penerapan algoritma Random Forest untuk klasifikasi predikat kelulusan mahasiswa universitas khairun”.

## 2. TINJAUAN PUSTAKA

### 2.1 Klasifikasi

Klasifikasi berasal dari bahasa latin *classis* yang berarti mengelompokkan benda yang sejenis dan memisahkan benda yang berbeda [9] Klasifikasi juga sering disebut sebagai metode untuk menentukan kategori kelas data yang telah tersedia. Metode ini cocok untuk mendeskripsikan *Dataset* yang memiliki tipe data biner atau nominal, kekurangan dari metode ini adalah tidak efektif untuk himpunan data ordinal karena pendekatan-pendekatan yang digunakan secara implisit dalam kategori data [10].

### 2.2 Pendidikan Karakter

Perguruan tinggi mempunyai kewajiban dalam mengontrol mahasiswanya untuk menghasilkan lulusan yang berkualitas. Lulusan yang berkualitas dapat dilihat dari indeks prestasi dan lamanya studi yang menghasilkan predikat kelulusan [11]. Predikat kelulusan mahasiswa pun terpengaruh ditentukan oleh beberapa faktor [12].

Adapun Peraturan Rektor Universitas Khairun Nomor 4 tahun 2020 tentang peraturan akademik, pada BAB V tentang penyelesaian studi bagian

keempat mengenai predikat kelulusan, pasal 21 ayat 2 sebagai Berikut:

Kelulusan mahasiswa dari jenjang program sarjana (S1) dapat diberikan predikat memuaskan, sangat memuaskan, atau pujian yang dinyatakan dalam transkrip akademik dengan kriteria:

1. Mahasiswa dinyatakan lulus dengan predikat memuaskan apabila mencapai IPK 2,76 (dua koma tujuh enam) sampai dengan 3,00 (tiga koma nol).
2. Mahasiswa dinyatakan lulus dengan predikat sangat memuaskan apabila mencapai IPK 3,01 (Tiga koma nol satu) sampai dengan 3,50 (tiga koma lima nol) dan masa studi tidak lebih dari 10 semester efektif.
3. Mahasiswa dinyatakan lulus dengan predikat pujian (cumlaude) apabila mencapai IPK lebih dari 3,50 (tiga koma lima nol) sampai dengan 4,00 (empat koma nol) dengan masa studi tidak lebih dari 8 semester efektif tanpa nilai terendah C.

### 2.3 *Random Forest*

Algoritma Random Forest dirancang oleh J.Ros Quinlan, disebut Random Forest karena adalah turunan dari metode ID3 untuk membangun pohon keputusan. Random Forest adalah algoritma yang dapat digunakan untuk masalah klasifikasi dalam pembelajaran mesin dan Penambangan data [6].

Random Forest merupakan metode yang dikembangkan dari metode CART (Classification and Regression Trees), yaitu dengan menerapkan metode bootstrap aggregating (bagging) dan random feature selection [15] CART merupakan metode eksplorasi data berdasarkan teknik pohon keputusan. Regresi pohon dihasilkan ketika respons peubah memiliki data numerik, klasifikasi pohon dihasilkan ketika respons peubah berisi data kategorikal [7]. Random Forest pertama kali dikenalkan oleh Breiman pada tahun 2001. Penelitiannya menunjukkan bahwa kelebihan dari random forest antara lain menghasilkan error yang lebih rendah, memberikan hasil klasifikasi yang baik, dapat mengatasi data training dalam jumlah sangat besar secara efisien dan metode yang efektif untuk mengestimasi missing data [15].

1. Tahapan persiapan dan estimasi menggunakan Random Forest antara lain:
2. (Tahapan bootstrap) Ambil sampel acak ukuran pemulihan  $n$  dari cluster data pelatihan.
3. (Tahapan pengelompokan acak) Susun pohon sesuai data, tetapi pilih secara acak jumlah variabel prediktor ( $m$ ) <  $d$  variabel penjelas pada setiap pemisahan dan lakukan pemisahan terbaik.
4. Ulangi langkah 1 dan 2  $k$  kali untuk mendapatkan  $k$  pohon acak.

5. Mulailah percobaan berdasarkan k pohon yang disebutkan di atas (misalnya, menggunakan majority vote untuk kasus klasifikasi atau rata rata untuk kasus regresi).

Perhatikan bahwa setiap kali pohon dibangun, variabel penjelas yang dapat digunakan untuk pemisahan hanyalah sebagian hasil yang dipilih secara acak, bukan seluruh variabel yang terlibat.

Pohon keputusan dimulai dengan cara menghitung nilai entropy sebagai penentu tingkat ketidakmurnian atribut dan nilai information gain. Untuk menghitung nilai entropy digunakan rumus rumus seperti pada persamaan 1, sedangkan nilai information gain menggunakan persamaan 2 [17].

Berikut ini merupakan rumus dari Random Forest dapat dilihat pada persamaan 1 dan persamaan 2.

$$Entropy(Y) = -\sum_i p(c|Y) \log_2 p(c|Y).....(1)$$

Keterangan: P =Himpunan Kasus.  
 P(c|Y) =Proporsi nilai Y terhadap kelas

$$information\ gain(Y,a)=Entropy(Y)-$$

$$\sum_{v \in Values(a)} |Yv|/|Ya| Entropy(Yv).....(2)$$

Keterangan:

- Values(a) = Nilai yang mungkin dalam himpunan kasus a.
- Yv = Subkelas dari Y dengan kelas v yang berhubungan dengan kelas a.
- Ya = semua nilai dari sesuai dengan a.

### 2.4 Confusion Matrix

*Confusion matrix* adalah sebuah tabel atau matriks yang diberisikan empat nilai yang merupakan pengukuran performa dari masalah klasifikasi yang telah dilakukan. Ada empat nilai atau point yang ada pada *confusion matrix* yaitu *true positive* , *true negative*, *false positive* dan *false negative* [18].

Gambaran confusion matrix dapat dilihat pada tabel 1.

Tabel 1. Gambaran *Confusion matrix*  
*True Values*

		Positive	Negative
Predictions	Positive	True Positive	False Negative
	Negative	False Negative	True Positive

Keterangan:

TP = Prediksi yang bernilai positif dan benar sesuai target.

TN = Prediksi yang bernilai negatif dan benar sesuai target.

FP = rediksi yang bernilai positif dan salah tidak sesuai target.

FN = Prediksi yang bernilai negatif dan salah tidak sesuai target.

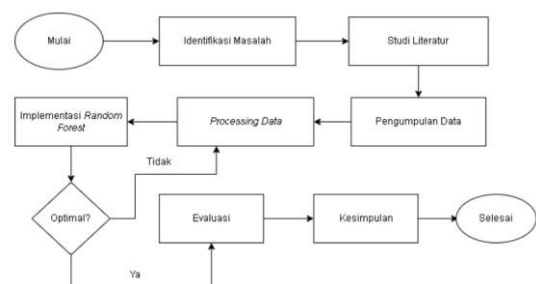
Akurasi menggambarkan seberapa akurat data yang diminta sesuai dengan prediksi yang diberikan oleh sistem [19] Sedangkan *F1-score* merupakan perhitungan yang menggambarkan perbandingan antara presisi dan *recall* [18].

### 2.5 Phytion

Bahasa pemrograman yang diciptakan oleh Guido van Rossum di belanda pada tahun 1990 inilah menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak sintuitif dan memiliki pustaka yang luas Python merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaan kambesar maupun para *developer* untuk mengembangkan berbagai macam aplikasi berbasis desktop, web dan mobile [20].

## 3. METODE PENELITIAN

### 3.1. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

### 3.2. Data Pre-Processing

#### 1. Data Cleaning

*Data cleaning* adalah proses menganalisis kualitas data dengan mengubah, mengoreksi, atau menghapus data yang salah, tidak lengkap, tidak akurat, atau cacat dalam database untuk menghasilkan data berkualitas tinggi. data sebelum *Cleaning* dapat dilihat pada Tabel 2. Dan data sesudah di *Cleaning* pada Tabel

3.

Tabel 2. Data Sebelum *Cleaning*

NPM	Nama	TM	TK	IPK	Lama Studi
02041511037	Ansari Syahafin	1 Sep 2016	31Jan 2020	3,51	3 Thn 4 bln
02041411105	Dzulkifli Kala	1 Sep 2016	-	-	3 Thn 4 bln
02041511024	Sriwiyanti Junus	1 Sep 2015	27Des 2019	3,45	4Thn 3 Bln
02041411032	Sugiono L.	1 Sep 2013	30Jan 2020	3,19	6Thn 4 Bln
02041511077	Lufti A.	1 Sep 2014	29Jan 2020	3,26	5Thn 4 Bln

Tabel 3. Data Sesudah *Cleaning*

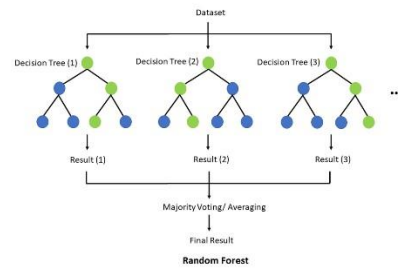
NPM	Nama	TM	TK	IPK	Lama Studi
02041511037	Ansari Syahafin	1 Sep 2016	31Jan 2020	3,51	3 Thn 4 bln
02041511024	Sriwiyanti Junus	1 Sep 2015	27Des 2019	3,45	4Thn 3 Bln
02041411032	Sugiono L.	1 Sep 2013	30Jan 2020	3,19	6Thn 4 Bln
02041511077	Lufti A.	1 Sep 2014	29Jan 2020	3,26	5Thn 4 Bln

### 2. Data Transformation

Data *transformation* merupakan proses mengubah data ke dalam bentuk yang sesuai agar dapat digunakan oleh algoritma *Random Forest*. Pada tahap ini, data kategorikal diubah menjadi data numerik melalui proses *encoding*, serta dilakukan normalisasi data untuk menyeragamkan skala nilai antar variabel. *Transformation* ini bertujuan agar proses pembelajaran model berjalan lebih optimal dan menghasilkan klasifikasi yang akurat.

### 3.3. Random Forest

*Random Forest* bekerja dalam dua tahap. Tahap pertama adalah menggabungkan N pohon keputusan untuk membuat *Random Forest*. Kemudian tahap kedua adalah melakukan prediksi pada setiap pohon yang dibuat pada tahap pertama, Cara kerja *Random Forest* dapat dilihat pada gambar 2.



Gambar 2. Cara Kerja *Random Forest*

Cara kerja algoritma *Random Forest* dapat dijabarkan dalam langkah-langkah berikut:

1. Algoritma memilih sampel acak dari *Dataset* yang disediakan, dapat dilihat pada Tabel 4.

Tabel 4. Data Sesudah *Cleaning*

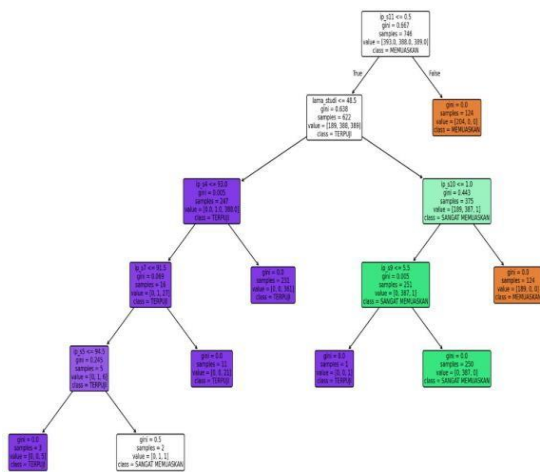
No	NPM	Lama Studi	IPK	Predikat
1.	02041511037	3Tahun 4 Bulan	3,51	Pujian
2.	02041411105	3 Tahun 4 Bulan	3,50	Pujian
3.	02041511049	4 Tahun 4 Bulan	3,30	Sangat Memuaskan
4.	02041511077	5 Tahun 4 Bulan	3,26	Memuaskan

2. Membuat *decision tree* untuk setiap sampel kelas terpuji, kemudian akan didapatkan hasil prediksi dari setiap *decision tree* yang telah dibuat.
3. Dilakukan proses *voting* untuk setiap hasil prediksi. Untuk masalah klasifikasi menggunakan modus (nilai yang sering muncul), sedangkan untuk masalah regresi akan menggunakan mean (nilai rata-rata).
  - a. Pohon Keputusan 1: lama studi 3 tahun 4 bulan dengan IPK 3,51 (Hasil prediksi: Pujian).
  - b. Pohon Keputusan 2: Lama studi 3 tahun 4 bulan dengan IPK 3,50 (Hasil prediksi: Pujian).
  - c. Pohon Keputusan 3: Lama studi 4 tahun 4 bulan dengan IPK 3,30 (Hasil prediksi: Sangat Memuaskan).
  - d. Pohon Keputusan 4: Lama studi 5 tahun 4 bulan dengan IPK 3,26 (Hasil prediksi: Memuaskan).
4. Algoritma *random forest* akan memilih prediksi dengan pilihan terbanyak (*Majority Voting*) sebagai prediksi akhir. Pada penerapannya, jumlah pohon keputusan yang digunakan pada metode ini bisa lebih banyak dan hasil prediksi akan mempertimbangkan *voting* dari semua pohon tersebut (*Majority Voting*).

## 4 HASIL DAN PEMBAHASAN

### 4.1. Implementasi *Random Forest*

Berikut ini, ditampilkan Plot Visualisasi salah satu struktur pohon keputusan (*decision tree*) dari hasil proses pelatihan model *Random Forest* dalam mengklasifikasikan terhadap tiga kelas predikat kelulusan mahasiswa yaitu Terpuji, Sangat Memuaskan dan Memuaskan. Pohon keputusan ini memperlihatkan bagaimana model melakukan proses pelatihan data berdasarkan fitur-fitur penting yang memiliki pengaruh signifikan terhadap hasil klasifikasi menggunakan metode *Random Forest*. dapat dilihat pada gambar 10.



Gambar 10. Visualisasi Pohon Keputusan

Berikut adalah penjelasan terperinci tentang makna warna dalam pohon keputusan ini:

1. Warna Ungu mengindikasikan prediksi kelas "Terpuji"
2. Warna Hijau Mengindikasikan prediksi kelas "Sangat Memuaskan"
3. Warna Orange mengindikasikan prediksi kelas "Memuaskan"
4. Warna Putih mengindikasikan prediksi campuran kelas dan belum bisa diklasifikasikan sepenuhnya ke salah satu kelas (gini impurity masih tinggi).

**4.2. Evaluasi Model**

Evaluasi performa model dilakukan menggunakan *Confusion Matrix* yang dapat dilihat pada gambar 11.

Jumlah Prediksi Benar (Confusion Matrix)		
254	0	0
0	269	0
0	0	257

Gambar 11. *Confusion Matrix* Evaluasi Model

Berdasarkan hasil evaluasi, model menunjukkan performa yang sangat baik dalam mengklasifikasikan data ke dalam tiga kelas, yaitu

Terpuji, Sangat Memuaskan, dan Memuaskan. Dominasi nilai pada diagonal utama *Confusion Matrix* menandakan kesesuaian yang tinggi antara label aktual dan label prediksi. Model berhasil mengklasifikasikan masing-masing 254 data pada kelas Terpuji, 269 data pada kelas Sangat Memuaskan, dan 257 data pada kelas Memuaskan secara tepat tanpa kesalahan klasifikasi. Hasil ini menunjukkan bahwa pola dan karakteristik fitur pada setiap kelas dapat dikenali dengan sangat baik oleh model, sehingga menghasilkan akurasi sebesar 100% serta menunjukkan stabilitas dan konsistensi model dalam proses klasifikasi. Untuk mengetahui akurasi, Presisi, *Recall* dan *F1-Score* dari gambar 13 diuraikan pada perhitungan berikut:

$$\begin{aligned}
 \text{Presisi} &= \frac{254}{254 + 0} = \frac{254}{254} = 1,00 \\
 \text{Recall} &= \frac{254}{254 + 0} = \frac{254}{254} = 1,00 \\
 \text{F1 - Score} &= \frac{2 \times 1 \times 1}{1 + 1} = 1
 \end{aligned}$$

**Kelas Sangat Memuaskan**

$$\begin{aligned}
 \text{Presisi} &= \frac{269}{269 + 0} = \frac{269}{269} = 1,00 \\
 \text{Recall} &= \frac{269}{269 + 0} = \frac{269}{269} = 1,00 \\
 \text{F1 - Score} &= \frac{2 \times 1 \times 1}{1 + 1} = 1
 \end{aligned}$$

**Kelas Memuaskan**

$$\begin{aligned}
 \text{Presisi} &= \frac{257}{257 + 0} = \frac{257}{257} = 1,00 \\
 \text{Recall} &= \frac{257}{257 + 0} = \frac{257}{257} = 1,00 \\
 \text{F1 - Score} &= \frac{2 \times 1 \times 1}{1 + 1} = 1
 \end{aligned}$$

Untuk perhitungan *macro Precision*:

$$\text{Macro Precision} = \frac{1 + 1 + 1}{3} = \frac{3}{3} = 1,00$$

Untuk perhitungan *macro Recall*:

$$\text{Macro Recall} = \frac{1 + 1 + 1}{3} = \frac{3}{3} = 1,00$$

Untuk perhitungan *macro F1-Score*:

$$\frac{1 + 1 + 1}{3} = 1$$

$$\text{Macro F1 - Score} = \frac{\frac{254}{3} + \frac{269}{3} + \frac{257}{3}}{3} = 1,00$$

Untuk Perhitungan *Weighted Precision*:

$$\text{Weighted Precision} = \frac{(254 \times 1,0) + (269 \times 1,0) + (257 \times 1,0)}{780}$$

$$= \frac{780}{780} = 1,00$$

Untuk Perhitungan *Weighted Recall*:

$$\text{Weighted Recall} = \frac{(254 \times 1,0) + (269 \times 1,0) + (257 \times 1,0)}{780}$$

$$= \frac{780}{780} = 1,00$$

Untuk Perhitungan *Weighted F1-Score*:

$$\text{Weighted F1 - Score} = \frac{(254 \times 1,0) + (269 \times 1,0) + (257 \times 1,0)}{780}$$

$$= \frac{780}{780} = 1,00$$

Untuk menghitung akurasi model:

$$\text{Akurasi} = \frac{254 + 269 + 257}{780} = \frac{780}{780} = 100\%$$

### 4.3. Implementasi Interface

Penelitian ini terdapat beberapa *Interface* yang diantaranya terdiri :

#### 4.3.1. Halaman *Login*

Tampilan ini merupakan tampilan pertama pengguna dalam menggunakan sistem yang mewajibkan pengguna harus login dengan memasukan *Username* dan *Password*. Dapat dilihat pada gambar 12.



Gambar 12. Tampilan Halaman *Login*

#### 4.3.2. Tampilan Menu *Dataset Mahasiswa*

Pada tampilan halaman *dataset* mahasiswa terdapat informasi mengenai *dataset* yang digunakan. Dapat dilihat pada gambar 13 dan gambar 14.

No	NPM	Nama Mahasiswa	Jenis Kelamin	Action
1	01011811009	FANESA A SANGAJI	P	✖
2	01011811020	FEBILAHER	P	✖
3	01011811035	PATRY ZULFIKAR	L	✖
4	01011811045	BUDWANG LATARNA	P	✖
5	01011811057	PATROL TOHMY LOEWIK SABONO	L	✖
6	01011811086	HINNA WATI RUSTAM	P	✖
7	01011811092	RAHMAWATI HUD	P	✖
8	01011811094	SRI DEWI YULUF	P	✖
9	01011811098	SYAMBAK DUNWILA	L	✖
10	01011811103	M RYAN MURDIN	L	✖
11	01011811104	ASRI SHAMAHANG	L	✖
12	01011811109	M RIDKY LASTORY	L	✖

Gambar 13. *Dataset* Mahasiswa

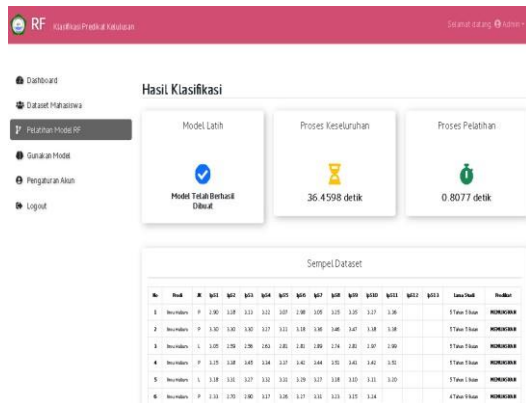
Data FANESA A SANGAJI	
Nomor Induk Mahasiswa 01011811009	Jenis Kelamin P
Nama Lengkap FANESA A SANGAJI	Tahun Masuk 2018
Program Studi Ilmu Hukum	Lama Studi 5 Tahun 5 Bulan
IP Semester 1 2.90	IP Semester 8 3.25
IP Semester 2 3.18	IP Semester 9 3.35
IP Semester 3 3.13	IP Semester 10 3.27
IP Semester 4 3.22	IP Semester 11 3.36
IP Semester 5 3.07	IP Semester 12 3.13
IP Semester 6 2.98	
IP Semester 7 3.05	

Gambar 14. Tampilan Informasi *Dataset*

#### 4.3.3. Tampilan Menu *Pelatihan Model RF*

Pada tampilan menu ini melibatkan pemberian sejumlah besar data (disebut data latih atau *training data*) kepada algoritma *Machine Learning (Random Forest)*. Pada tampilan menu ini dapat mengunggah data berlabel sejumlah 1950 *dataset* dengan 1170 data latih dan 780 data uji, kemudian sistem akan menggunakan 1170 data latih tersebut untuk mengajari algoritma *Random Forest*, setelah pelatihan selesai akan mengevaluasi akurasi model menggunakan 780 data uji untuk memastikan model

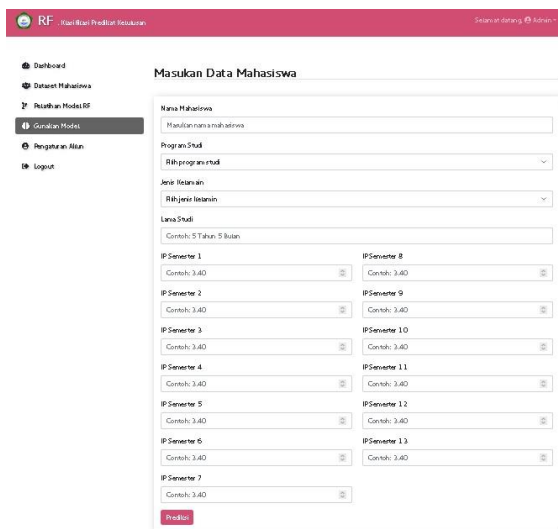
dapat membuat prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya dan hasil palatihan berupa model *Random forest* yang sudah siap pakai akan disimpan secara otomatis pada sistem kita dalam format pkl, Dapat dilihat pada gambar 15.



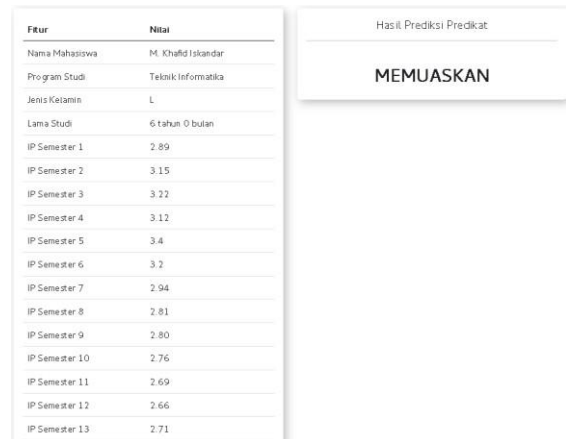
Gambar 15. Tampilan Menu Pelatihan Model *Random Forest*

#### 4.3.4. Tampilan Menu Gunakan Model

Pada tampilan menu gunakan model memungkinkan kita untuk menggunakan *Inference/Prediction* model *Random Forest* yang telah kita latih sebelumnya dengan cara menginput data baru memiliki label Model *Random Forest* yang sudah terlatih (yang telah disimpan dalam format.pkl (pickle)). dapat dilihat pada gambar 16.



Gambar 16. Tampilan Menu Gunakan Model



Gambar 17. Tampilan Menu Gunakan Model

## 5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa Penelitian ini berhasil dilakukan dengan jumlah *dataset* yang digunakan sebanyak 1.950 data yang dikumpulkan dari BAKP dan data Mahasiswa dengan rincihan 650 Kelas Predikat Terpuji, 650 Kelas Predikat Sangat Memuaskan dan 650 Kelas Predikat Memuaskan. Data tersebut kemudian dilakukan tahapan *preprocessing* yang meliputi *data cleaning* dan *data transformation* mampu mempersiapkan data dengan baik sehingga model dapat mengenali pola pada data secara optimal. Kemudian data dibagi kedalam data latih sebanyak 60% atau 1.170 data dan data Test sebanyak 40% atau 780 data, evaluasi performa model menggunakan *Confusion Matrix* menghasilkan nilai akurasi 100%, *Precision* 100%, *Recall* 100% dan *F1-Score* 100%. Pengujian *Black Box* terhadap sistem aplikasi menunjukkan semua fitur utama seperti *login*, pengelolaan *dataset*, pelatihan model, dan penggunaan model berjalan dengan baik sesuai keinginan dan tidak ditemukan kesalahan selama proses pengujian. Hasil penelitian ini menunjukkan bahwa algoritma *Random Forest* sangat efektif untuk mengklasifikasi Predikat kelulusan mahasiswa.

## DAFTAR PUSTAKA

- [1] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, & Susanti. 2022. Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Random Forest*. SATIN - Sains Dan Teknologi Informasi, 8(2), 122–131.
- [2] khoerul ummah. (2022). SISTEM PREDIKSI PREDIKAT KELULUSAN MAHASISWA MENGGUNAKAN METODE

- REGRESI LOGISTIK  
MULTINOMIAL., 8.5.2017,  
2003–2005.
- [3] Alturki, S., Alturki, N., & Stuckenschmidt, H. (2021). Using Educational Data Mining To Predict Students' Academic Performance For Applying Early Interventions. *Journal of Information Technology Education: Innovations in Practice*, 20, 121–137.  
<https://doi.org/10.28945/4835>
- [4] Luvia, Y. S., Windarto, A. P., Solikhun, S., & Hartama, D. 2017. Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 1(1), 75.
- [5] Ramadhani, L. K., & Widyaningrum, B. N. 2022. Perbandingan Metode Klasifikasi *Naïve Bayes* dan *Support Vector Machine* Pada Predikat Kelulusan Mahasiswa. 8798, 150–154.
- [6] Breiman, L. 2001. *Using iterated bagging to debias regressions. Machine Learning*, 45(3), 261–277.
- [7] Breiman, L., & Ihaka, R. 1984. *Nonlinear discriminant analysis via scaling and ACE. Technical Report No. 40*, 40.
- [8] Srirahayu, A., & Pribadie, L. S. 2023. *Review Paper Data Mining Klasifikasi Data Mining. Jurnal Ilmiah Informatika Global*, 14
- [9] Sains, S., Kelulusan, P., Di, M., Kampar, P., Saputra, A., & Fitri, T. A. 2023. Penerapan Data Mining Algoritma C4.5 Dalam Memprediksi.
- [10] Wibowo, A., & Rohman, A. 2022. Prediksi Predikat Kelulusan Mahasiswa Menggunakan *Naive Bayes* dan *Decision Tree* pada Universitas XYZ. *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, 12(2), 104.
- [11] Zailani, A. U., & Hanun, N. L. 2020. Penerapan Algoritma Klasifikasi *Random Forest* Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera. *Infotech: Journal of Technology Information*, 6(1), 7–14.
- [12] Rahmah Muthia, 2018. Klasifikasi Ketepatan Lama Studi Mahasiswa Menggunakan Metode *Support Vector Machine* dan *Random Forest*. 1–26.
- [13] Nurvania, J., Jondri, & Lhaksamana, K. M. 2021. Analisis Sentimen Pada Ulasan di *TripAdvisor* Menggunakan Metode *Long Short-Term Memory (LSTM)*. *E-Proceeding of Engineering*, 8(4), 4124–4135.
- [14] Renaldo Yosia Rafael, F. A. 2023. Pengimplmentasian Algoritma *Long ShortTerm Memory* Untuk Mendeteksi Ujaran Kebencian Pada Aplikasi *Twitter*. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(2), 551–560.
- [15] Schuerer, K., Maufrais, C., Letondal, C., & Deveaud, E. 2004. *Introduction to Programming using Python Programming Course for Biologists at the Pasteur Institute*.
- [16] Wahyono, T. 2021. *Fundamental of Python for Machine Learning Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan*.