

## Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest

Oon Wira Yuda<sup>1</sup>, Darmawan Tuti<sup>2</sup>, Lim Sheih Yee<sup>3</sup>, Susanti<sup>4</sup>

<sup>1</sup>STMik Amik Riau, [oonwirayuda@stmik-amik-riau.ac.id](mailto:oonwirayuda@stmik-amik-riau.ac.id), Pekanbaru, Indonesia

<sup>2</sup>STMik Amik Riau, [darmawantuti@stmik-amik-riau.ac.id](mailto:darmawantuti@stmik-amik-riau.ac.id), Pekanbaru, Indonesia

<sup>3</sup>STMik Amik Riau, [limsheihyee@stmik-amik-riau.ac.id](mailto:limsheihyee@stmik-amik-riau.ac.id), Pekanbaru, Indonesia

<sup>4</sup>STMik Amik Riau, [susanti@stmik-amik-riau.ac.id](mailto:susanti@stmik-amik-riau.ac.id), Pekanbaru, Indonesia

### Informasi Makalah

Submit : Oktober 19, 2022  
Revisi : November 28, 2022  
Diterima : Desember 6, 2022

### Kata Kunci :

Data Mining  
Klasifikasi  
Kelulusan mahasiswa  
Random Forest

### Abstrak

Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentasi kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Beberapa faktor dapat dilihat dari Indeks Prestasi Kumulatif (IPK) yang mencerminkan seluruh nilai yang diperoleh mahasiswa setiap semester yang sedang berjalan, dan jumlah SKS yang telah ditempuh mahasiswa. Berdasarkan data kelulusan yang di peroleh dari PDPT STMik Amik Riau dari tahun 2019 sampai 2022 berjumlah 511 data. Masalah tingkat kelulusan mahasiswa banyak dialami oleh perguruan tinggi. Faktor utama yang menentukan reputasi suatu perguruan tinggi salah satunya adalah kelulusan mahasiswa yang tepat waktu. Sehubungan dengan hal itu dilakukanlah penelitian yang bertujuan menghasilkan sebuah analisa yang dapat dijadikan acuan bagi STMik Amik Riau untuk mengatasi masalah tersebut. Penelitian ini menggunakan metode Random Forest. Variabel klasifikasi yang dihasilkan berdasarkan kategori Tepat waktu dengan ipk tinggi, Tepat waktu dengan ipk sedang, Tepat waktu dengan ipk rendah, Tidak tepat waktu dengan ipk tinggi, Tidak tepat waktu dengan sedang, Tidak tepat waktu dengan ipk rendah. Dengan menggunakan algoritma *random forest* dan variable yang digunakan, diperoleh tingkat akurasi sebesar 98%. Nilai akurasi yang diperoleh ini menunjukkan bahwa proses klasifikasi yang dilakukan akurat.

### Abstract

One of the factors that determine the quality of tertiary institutions is the percentage of students' ability to complete their studies on time. Several factors can be seen from the Grade Point Average (GPA), which reflects all grades earned by students each ongoing semester, and the number of credits students have taken. Based on the graduation data obtained from PDPT STMik Amik Riau from 2019 to 2022 there are 511 data. The problem of student graduation rates is experienced by many tertiary institutions. One of the main factors that determine the reputation of a tertiary institution is the timely graduation of students. In connection with this, research was carried out which aims to produce an analysis that can be used as a reference for STMik Amik Riau to overcome this problem. This study uses the Random Forest method. The resulting classification variables are based on the categories On time with high GPA, On time with moderate GPA, On time with low GPA, Not on time with high GPA, Not on time with moderate, Not on time with low GPA. By using the random forest algorithm and the variables used, an accuracy rate of 98% is obtained. The accuracy value obtained indicates that the classification process is accurate.

Oon Wira Yuda,

Email: [oonwirayuda@stmik-amik-riau.ac.id](mailto:oonwirayuda@stmik-amik-riau.ac.id).

## 1. Pendahuluan

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa. Lima lembaga perguruan tinggi diantaranya adalah universitas, institut, sekolah tinggi, akademi dan politeknik. Perguruan tinggi diharapkan menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang berilmu, cakap dan kreatif (Nugroho & Wibowo, 2017)

Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Berdasarkan matriks penilaian instrument akreditasi program studi Badan Akreditasi Nasional Perguruan Tinggi bahwa persentase mahasiswa yang lulus tepat waktu merupakan salah satu elemen penilaian akreditasi universitas (BAN-PT, 2008).

Masalah tingkat kelulusan mahasiswa banyak dialami oleh perguruan tinggi. Faktor utama yang menentukan reputasi suatu perguruan tinggi salah satunya adalah kelulusan mahasiswa yang tepat waktu. Semakin tinggi tingkat mahasiswa pada perguruan tinggi maka dengan rasio yang sama juga wajib ada mahasiswa yang lulus tepat waktu. Peningkatan jumlah data mahasiswa terjadi jika banyak mahasiswa yang tidak lulus tepat waktu dari semua mahasiswa yang terdaftar. Sehingga akan mempengaruhi reputasi dari perguruan tinggi yang dapat mengancam akreditasi perguruan tinggi.

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) AMIK Riau adalah sebuah perguruan tinggi swasta (PTS) di Provinsi Riau. STMIK Amik Riau memiliki tiga program studi yaitu Teknik Informatika, Sistem Informasi dan Teknologi Informasi. Kualitas lulusan dari sebuah perguruan tinggi salah satunya dapat dilihat dari lama studi mahasiswa. Selain itu lama studi mahasiswa menggambarkan tingkat capaian mahasiswa

dalam pendidikannya. Lama studi juga sangat berpengaruh pada kualitas program studi karena lama studi mahasiswa merupakan salah satu kriteria penilaian akreditasi. Lama studi merupakan jangka waktu yang diperlukan mahasiswa dalam menyelesaikan pendidikannya (Endang Etriyanti, 2021)

Salah satu kriteria penilaian pada akreditasi perguruan tinggi adalah penilaian terhadap lama studi mahasiswa yang lulus. Lama studi mahasiswa dapat dipengaruhi oleh beberapa faktor dapat dilihat dari Indeks Prestasi Kumulatif (IPK) yang mencerminkan seluruh nilai yang diperoleh mahasiswa setiap semester yang sedang berjalan. IPK diperoleh dengan menjumlahkan nilai dari semua mata kuliah yang telah diambil dan membaginya dengan total SKS (Satuan Kredit Semester) (Kurnawan et al., 2018).

Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan Teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Mardi, 2017). Data mining adalah metodologi analisis data yang digunakan untuk mengidentifikasi pola-pola tersembunyi diset data yang besar (Pratiwi & Adrianto, 2017).

Klasifikasi Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Annur, 2018). Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/ Naïve Bayes classifiers*, *Neural networks*, *Analisa Statistik*, *Algoritma Genetika*, *Rough sets*, *k-nearest neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)*.

Penelitian yang sudah dilakukan oleh beberapa peneliti terkait Klasifikasi kelulusan

mahasiswa tepat waktu sangat beragam. Seperti penelitian yang dilakukan terkait penerapan metode klasifikasi *decision tree* untuk memprediksi kelulusan tepat waktu dengan tingkat akurasi 68,49% (Ma'sum et al., 2021).

Selanjutnya implementasi sistem klasifikasi ketepatan waktu kelulusan dengan menggunakan metode backpropagation neural network pada kasus studi sistem informasi universitas Jember dengan memperoleh akurasi 98,82% (Hizham et al., 2018).

Selain itu, penelitian selanjutnya klasifikasi waktu kelulusan mahasiswa STIKOM Bali menggunakan metode chaid regression trees dan regresi logistik biner. Hasil perbandingan ketepatan klasifikasi diperoleh bahwa klasifikasi dengan menggunakan metode CHAID regression trees lebih baik yaitu 91,2% sedangkan dengan regresi logistik ketepatan klasifikasi mencapai 90,2%. Dengan menghasilkan perbandingan metode *chaid regression trees* lebih baik dibandingkan dengan menggunakan metode regresi logistik biner (Ketut et al., 2017).

Penelitian selanjutnya tentang klasifikasi predikat tingkat kelulusan mahasiswa program studi Teknik Informatika dengan menggunakan algoritma C45 dengan diperoleh nilai true positif 99, true negatif 10, false negatif 8, false positif 10 serta nilai accuracy sebesar 85,83% (Suhada et al., 2021).

Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes dengan hasil akurasi tes classifier untuk klasifikasi lama studi sebesar 0,74 dan untuk akurasi tes klasifikasi predikat kelulusan sebesar 0,61 pada kelompok data Program (Suhada et al., 2021) Studi Sistem Komputer. Kemudian untuk kelompok data Program Studi Sistem Informasi akurasi tes klasifikasi lama studi sebesar 0,73 dan untuk

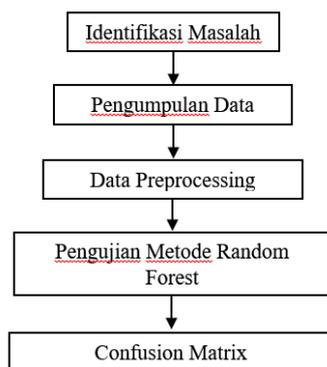
akurasi klasifikasi predikat kelulusan sebesar 0,67 (Rudy Hendrawan et al., 2022).

Penelitian terakhir yaitu perbandingan metode fuzzy C-Means dan K-Means dalam klasifikasi kelulusan mahasiswa (studi kasus : jurusan manajemen, universitas Maritim Raja Ali Haji). Perbandingan metode *fuzzy c-means* dan *k-means* digunakan untuk mendapatkan hasil klasifikasi yang tepat dan akurat. Berdasarkan hasil pengujian yang telah dilakukan pada *fuzzy c-means* nilai pangkat (w) 3, maksimum iterasi 50 dan nilai eror 0.01 didapatkan nilai akurasi sebesar 81.91%. Sedangkan pada pengujian tingkat akurasi metode *k-means* didapatkan nilai tingkat akurasi sebesar 63%. Hal ini membuktikan bahwa metode *fuzzy c-means* lebih akurat dibandingkan dengan metode *k-means* dalam klasifikasi kelulusan. (Uperiati et al., 2020).

Berdasarkan dari penelitian sebelumnya, algoritma yang digunakan dalam klasifikasi kelulusan mahasiswa masih menggunakan algoritma *deep learning*. Sedangkan pada penelitian ini menggunakan algoritma *random forest* yang termasuk dalam kategori algoritma *machine learning* untuk mengklasifikasi kelulusan mahasiswa yang pada penelitian sebelumnya belum ada yang menggunakan algoritma *random forest*. Hasil penelitian ini digunakan untuk mengetahui tingkat akurasi dari proses klasifikasi yang telah diperoleh.

## 2. Metode Penelitian

Metode penelitian merupakan teknik yang disusun oleh peneliti untuk mengumpulkan data dan informasi dalam melakukan penelitian yang sesuai subjek dan objek yang diteliti, dengan adanya data tersebut diharapkan mampu memberikan hasil yang berkualitas dan optimal. Metode penelitian disajikan pada gambar 1.



Gambar 1. Metode penelitian

### 3.1. Identifikasi Masalah

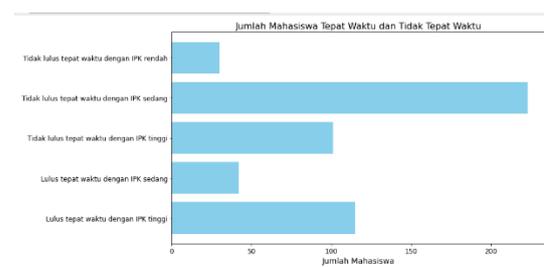
Identifikasi masalah merupakan langkah awal dalam proses penelitian, identifikasi masalah dalam penelitian ini adalah menemukan permasalahan yang terjadi pada mahasiswa STMIK Amik Riau. Permasalahan yang terjadi yaitu untuk evaluasi terkait tingkat kelulusan mahasiswa menggunakan data masa lalu yang diambil dari data kelulusan mahasiswa tahun 2016 hingga 2022 untuk dijadikan sebuah informasi untuk kelulusan setiap tahunnya. Pengetahuan ini dapat digunakan sebagai bahan pertimbangan Program Studi untuk menentukan kebijakan atau pembinaan terkait tingkat kelulusan mahasiswa disuatu perguruan tinggi. Untuk mengetahui tingkat kelulusan mahasiswa disuatu perguruan tinggi, dilakukan klasifikasi lama masa studi mahasiswa disuatu berdasarkan IPK.

### 3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini yaitu data mahasiswa aktif STMIK Amik Riau tahun angkatan 2016 hingga 2022. Data didapatkan dari PDPT STMIK Amik Riau yang berjumlah 511 data. Data dikategorikan berdasarkan predikat, yang disajikan pada Tabel 1.

Table 1. Kategori Data

| Kategori | Keterangan                                | Jumlah |
|----------|---|--------|
| 1        | Lulus tepat waktu dengan IPK tinggi       | 115    |
| 2        | Lulus tepat waktu dengan IPK sedang       | 42     |
| 3        | Lulus tepat waktu dengan IPK rendah       | 0      |
| 4        | Tidak lulus tepat waktu dengan IPK tinggi | 101    |
| 5        | Tidak lulus tepat waktu dengan IPK sedang | 223    |
| 6        | Tidak lulus tepat waktu dengan IPK rendah | 30     |



Gambar 2 . Grafik Distribusi Kelas

### 3.3. Preprocessing

*Data preprocessing* merupakan sekumpulan teknik yang diterapkan pada ngkap dan tidak konsisten sehingga mengakibatkan database untuk menghapus *noise*, *missing value*, dan data yang tidak konsisten. *Data preprocessing* ini digunakan karena dalam data *realtime database* sering kali tidak le hasil data mining tidak tepat dan kurang akurat. Oleh karena itu, untuk meningkatkan kualitas data yang akan dianalisis, perlu dilakukan *preprocessing data*. Proses *preprocessing data* yang dilakukan yaitu *cleaning data*. *Cleaning data* adalah proses mempersiapkan data untuk analisis dengan menghapus atau memodifikasi data yang tidak benar, tidak lengkap, tidak

relevan. Data yang digunakan pada penelitian ini sudah bersih. Tahap selanjutnya dilakukan proses cleaning data hanya memperlihatkan data apakah ada data yang kosong.

### 3.4. Pengujian Metode

#### 1. Data Mining

*Data mining* adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari *database* besar. *Data mining* dapat diartikan sebagai proses analisis data yang dapat dilakukan dengan banyak metode (Sukarna Royan, Ansori Yulian, 2021). *Data Mining* tidak hanya mengumpulkan data, tetapi mencakup analisis dan analisis prediksi informasi yang ingin ditampilkan. Data yang terkumpul disimpan dalam *database* kemudian diolah sehingga dapat digunakan untuk pengambilan keputusan dalam melihat informasi yang akan digunakan (Lumban Gaol et al., 2021). Menurut (Asriningtias & Mardhiyah, 2014) tahap-tahap dari *data mining* meliputi:

##### 1) *Data cleaning*

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak relevan. Berikut hasil dari proses *cleaning data*

```
In [5]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 511 entries, 0 to 510
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   IPK          511 non-null    float64
1   Total SKS   511 non-null    int64
2   Total Studi 511 non-null    int64
3   Target      511 non-null    int64
dtypes: float64(1), int64(3)
memory usage: 16.1 KB
```

Gambar 3. Data Cleaning

##### 2) *Data integration*

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

##### 3) *Data selection*

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

##### 4) *Data transformation*

Data diubah atau digabung ke dalam format yang sesuai untuk diproses.

##### 5) Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

##### 6) Evaluasi pola

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasilnya berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

##### 7) *Knowledge presentation*

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

#### 2. Klasifikasi

Klasifikasi data mining adalah sebuah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas. Klasifikasi adalah salah satu metode yang paling umum untuk digunakan dalam data mining (Mardi, 2017).

#### 3. Machine Learning

*Machine learning* atau pembelajaran mesin adalah cabang dari Ilmu Komputer terkait dengan *Artificial Intelligence* atau ilmu kecerdasan buatan yang terutama berfokus pada pengembangan dan studi suatu sistem dengan tujuan untuk dapat

belajar dari data yang diperolehnya (Rahmansyah et al., 2018). Hal dasar dan tidak terpisahkan untuk menerapkan algoritma *machine learning* yaitu data. *Machine learning* tidak akan pernah berfungsi tanpa data yang tersedia. Dari data tersebut biasanya dibagi menjadi dua kelompok yaitu *data training* dan *data testing*. *Data training* digunakan untuk pengembangan model dan algoritma latih, sedangkan *data testing* digunakan sebagai acuan dan untuk mengetahui bagaimana performansi algoritma yang dilatih sebelumnya saat menemukan data baru yang belum pernah diketahui. Berikut hasil dari pembagian dari data training dan data testing

```
In [58]: M from sklearn.model_selection import train_test_split
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 42)

In [59]: M X_train.shape, X_test.shape
Out[59]: ((408, 3), (103, 3))
```

Gambar 4. Splitting Data

Data training berjumlah 408 data dan data training berjumlah 103 data

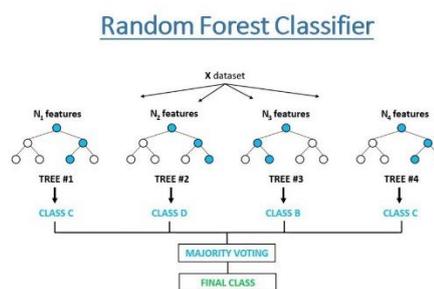
#### 4. Supervised Learning

*Machine learning* atau pembelajaran mesin adalah cabang dari Ilmu Komputer terkait dengan *Artificial Intelligence* atau ilmu kecerdasan buatan yang terutama berfokus pada pengembangan dan studi suatu sistem dengan tujuan untuk dapat belajar dari data yang diperolehnya (Utami et al., 2020). Metode *supervised learning* digunakan untuk menyelesaikan berbagai persoalan yang terkait klasifikasi, dan regresi. Algoritma yang tergolong dari *supervised learning* yaitu *linear regression*, *decision tree*, *random forest*, *naïve bayes*, *support vector machine*, *Artificial Neural Network*, *K-Nearest Neighbor* dan lain sebagainya.

#### 5. Random Forest

*Random forest classifier* merupakan metode klasifikasi yang terdiri dari kumpulan pohon keputusan yang nantinya akan dijadikan *vote* untuk mendapatkan hasil

terakhir dari pendeteksian sarkasme dengan pendukung berupa data latih dan fitur acak yang independen dengan fitur yang berbeda-beda (Alita & Isnain, 2020). Pohon keputusan dibuat dengan menentukan node akar dan berakhir dengan beberapa node daun untuk mendapatkan hasil akhir. Berikut merupakan ilustrasi dari pohon keputusan yang dapat dilihat pada Gambar 4.



Gambar 5. Random Forest Classifier

Persamaan yang digunakan dalam metode random forest dapat ditunjukkan pada persamaan (1).

$$f_i = \frac{\sum j: \text{node } j \text{ splits on feature } i \text{ } n_{ij}}{\sum k \in \text{all nodes } n_{ik}} \quad (1)$$

Keterangan:

$f_i$  sub(i)= pentingnya fitur  $i$

$n_{ij}$  sub(j)= pentingnya simpul  $j$

#### 3.5 Confusion Matrix

*Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Confusion matrix* digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan (Rahman et al., 2017).

Table 2. Metode *Confusion Matrix*

|                 |       | Observed Class      |                     |
|-----------------|-------|---------------------|---------------------|
|                 |       | True                | False               |
| Predicted Class | True  | True Positive (TP)  | False Positive (FP) |
|                 | False | False Negative (FN) | True Negative (TN)  |

Dimana :

1. TP adalah *True Positif*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
2. TN adalah *True Negatif*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
3. FN adalah *False Negatif*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
4. FP adalah *False Positif*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi diperoleh dengan persamaan berikut.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasi secara benar dibagi dengan total data yang diklasifikasi positif, Presisi dapat diperoleh dengan persamaan.

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

Sementara itu, recall menunjukkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

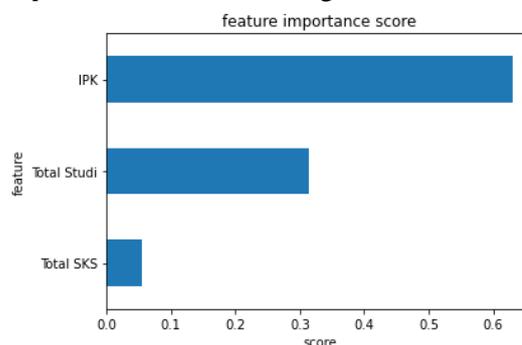
Error adalah kasus yang diidentifikasi salah dalam sejumlah data, sehingga dapat dilihat seberapa besar tingkat kesalahan pada sistem yang digunakan. Persentase error dapat dilakukan perhitungan menggunakan persamaan 5 di bawah ini.

$$Error = \frac{FP}{TP} \times 100\% \quad (5)$$

#### 4. Hasil dan Pembahasan

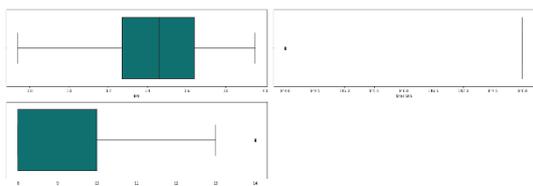
Pada implementasi metode *random forest* dengan menggunakan Bahasa pemrograman *Python*, data yang telah diperoleh dilakukan *split data* 80:20.

Berdasarkan pada gambar 1, menampilkan hasil *feature importance score* yang diperoleh yaitu nilai *feature importance* IPK dengan *score* 0.6305, *feature importance* Total Studi dengan *score* 0.3138 dan *feature importance* Total SKS dengan *score* 0.0555.



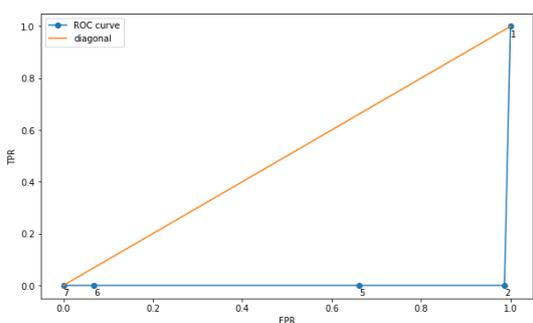
Gambar 6. Feature Importance Score

Pada gambar 7 merupakan visualisasi *boxplot* horizontal yang menjelaskan bahwa *boxplot* IPK nilai minimum yaitu 2.74, nilai maksimum 3.95, dan nilai median 3.46. Untuk *boxplot* Total Studi nilai minimum yaitu 8, nilai maksimum 13, dan *outlier* yaitu 14. Sedangkan pada *boxplot* Total SKS hanya menjelaskan bahwa *outlier* 144.



Gambar 7. Boxplot

Pada gambar 8 menampilkan ROC kurva, dimana ROC kurva menunjukkan visualisasi antara *true positive rate* (TPR) dan *false positive rate* (FPR). *Classifier* yang memberikan kurva semakin mendekati sudut kiri atas (*perfect classifier*) menunjukkan kinerja yang semakin baik. Pengujian dengan menggunakan metode random forest ini, diperoleh akurasi 0.9806 atau 98% yang dapat dilihat pada ROC kurva.



Gambar 8. ROC Kurva

### 1. Confusion Matrix

Berdasarkan pada gambar 9 dapat dijelaskan bahwa nilai akurasi yang diperoleh yaitu 0.98 atau 98%. Nilai akurasi yang didapat tergolong sangat baik.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 1            | 0.97      | 1.00   | 0.98     | 29      |
| 2            | 0.92      | 1.00   | 0.96     | 11      |
| 4            | 1.00      | 1.00   | 1.00     | 12      |
| 5            | 1.00      | 0.96   | 0.98     | 46      |
| 6            | 1.00      | 1.00   | 1.00     | 5       |
| accuracy     |           |        | 0.98     | 103     |
| macro avg    | 0.98      | 0.99   | 0.98     | 103     |
| weighted avg | 0.98      | 0.98   | 0.98     | 103     |

Gambar 9. Confusion Matrix

## 5. Simpulan

Dari hasil penelitian yang telah dilakukan di Perguruan Tinggi STMIK Amik Riau maka

dapat disimpulkan bahwa penelitian ini dilakukan untuk mengetahui proses klasifikasi kelulusan mahasiswa di STMIK Amik Riau menggunakan metode random forest. Berdasarkan hasil pengujian metode random forest dengan menggunakan Bahasa pemrograman Python, diperoleh tingkat akurasi sebesar 0.98 atau 98%. Dengan diperolehnya tingkat akurasi 98%, dapat disimpulkan bahwa pengujian metode random forest untuk mengklasifikasi kelulusan mahasiswa tergolong sangat akurat.

## 6. Referensi

- Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 677–690. <https://doi.org/10.30812/matrik.v21i3.1726>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58. <https://doi.org/10.23960/komputasi.v8i2.2615>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Asriningtias, Y., & Mardhiyah, R. (2014). Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa. *Informatika*, 8(1), 837–848.
- Endang Etriyanti. (2021). Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, 3(1), 6–14. <https://doi.org/10.52303/jb.v3i1.40>
- Hizham, F. A., Nurdiansyah, Y., & Firmansyah, D. M. (2018). Implementasi metode Backpropagation Neural Network (BNN) dalam sistem klasifikasi ketepatan waktu kelulusan mahasiswa. *Berkala Sainstek*, 6(2), 97–105. [https://www.researchgate.net/publication/330446472\\_Implementasi\\_Metode\\_Backprop](https://www.researchgate.net/publication/330446472_Implementasi_Metode_Backprop)

- agation\_Neural\_Network\_BNN\_dalam\_Sistem\_Klasifikasi\_Ketepatan\_Waktu\_Kelulusan\_Mahasiswa\_Studi\_Kasus\_Program\_Studi\_Sistem\_Informasi\_Universitas\_Jember
- Ketut, I., Suniantara, P., & Rusli, M. (2017). Klasifikasi Waktu Kelulusan Mahasiswa Stikom Bali Menggunakan Chaid Regression – Trees dan Regresi Logistik Biner. *Statistika*, 5(1), 27–32.
- Kurnawan, I., Marisa, F., & Purnomo, P. (2018). Implementasi Data Mining Dengan Algoritma Apriori Untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Teknologi Dan Manajemen Informatika*, 4(1).  
<https://doi.org/10.26905/jtmi.v4i1.1894>
- Lumban Gaol, L. Y., Safii, M., & Suhendro, D. (2021). Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 2(2), 97–106.  
<https://doi.org/10.30645/brahmana.v2i2.71>
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219.  
<https://doi.org/10.22202/ei.2016.v2i2.1465>
- Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Upgris*, 3(1), 63–70.  
<https://doi.org/10.26877/jiu.v3i1.1669>
- Pratiwi, F., & Adrianto, S. (2017). Peningkatan Jumlah Mahasiswa melalui Promosi dengan Penerapan Analisa Data Mining. *SATIN - Sains Dan Teknologi Informasi*, 3(2), 29–37.
- Rahman, M. F., Alamsah, D., & Darmawidjadja, M. I. (2017). Klasifikasi Untuk Diagnosa
- Rahmansyah, A., Dewi, O., Andini, P., Hastuti, T., Ningrum, P., & Suryana, M. E. (2018). Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 1907–5022.
- Rudy Hendrawan, I. N., Budhi Saputra, I. M. A., Cahya Dewi, G. A. P., Adi Pranata, I. G. S., & Wedasari, N. L. N. (2022). Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes. *Jurnal Eksplora Informatika*, 11(1), 50–56.  
<https://doi.org/10.30864/eksplora.v11i1.606>
- Suhada, K., Elanda, A., & Aziz, A. (2021). Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang). *Dirgamaya: Jurnal Manajemen Dan Sistem Informasi*, 1(2), 14–27.  
<https://doi.org/10.35969/dirgamaya.v1i2.182>
- Sukarna Royan, Ansori Yulian, S. (2021). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. In *SAINTEK / Jurnal Sains & Teknologi* (Vol. 5, Issue 2, pp. 9–22).
- Uperiati, A., Bettiza, M., & Puspasari, A. (2020). Perbandingan Metode Fuzzy C-Means Dan K-Means Dalam Klasifikasi Kelulusan Mahasiswa (Studi Kasus: Jurusan Manajemen, Universitas Maritim Raja Ali Haji). *Jurnal Sustainable: Jurnal Hasil Penelitian Dan Industri Terapan*, 9(2), 75–81.  
<https://doi.org/10.31629/sustainable.v9i2.1409>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58.  
<https://doi.org/10.23960/komputasi.v8i2.2615>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165.  
<https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Asriningtias, Y., & Mardhiyah, R. (2014). Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa. *Informatika*, 8(1), 837–848.
- Endang Etriyanti. (2021). Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa. *Jurnal Ilmiah Binary STMIK*

- Bina Nusantara Jaya Lubuklinggau*, 3(1), 6–14. <https://doi.org/10.52303/jb.v3i1.40>
- Hizham, F. A., Nurdiansyah, Y., & Firmansyah, D. M. (2018). Implementasi metode Backpropagation Neural Network (BNN) dalam sistem klasifikasi ketepatan waktu kelulusan mahasiswa. *Berkala Sainstek*, 6(2), 97–105. [https://www.researchgate.net/publication/330446472\\_Implementasi\\_Metode\\_Backpropagation\\_Neural\\_Network\\_BNN\\_dalam\\_Sistem\\_Klasifikasi\\_Ketepatan\\_Waktu\\_Kelulusan\\_Mahasiswa\\_Studi\\_Kasus\\_Program\\_Studi\\_Sistem\\_Informasi\\_Universitas\\_Jember](https://www.researchgate.net/publication/330446472_Implementasi_Metode_Backpropagation_Neural_Network_BNN_dalam_Sistem_Klasifikasi_Ketepatan_Waktu_Kelulusan_Mahasiswa_Studi_Kasus_Program_Studi_Sistem_Informasi_Universitas_Jember)
- Ketut, I., Suniantara, P., & Rusli, M. (2017). Klasifikasi Waktu Kelulusan Mahasiswa Stikom Bali Menggunakan Chaid Regression – Trees dan Regresi Logistik Biner. *Statistika*, 5(1), 27–32.
- Kurnawan, I., Marisa, F., & Purnomo, P. (2018). Implementasi Data Mining Dengan Algoritma Apriori Untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Teknologi Dan Manajemen Informatika*, 4(1). <https://doi.org/10.26905/jtmi.v4i1.1894>
- Lumban Gaol, L. Y., Safii, M., & Suhendro, D. (2021). Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 2(2), 97–106. <https://doi.org/10.30645/brahmana.v2i2.71>
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Upgris*, 3(1), 63–70. <https://doi.org/10.26877/jiu.v3i1.1669>
- Pratiwi, F., & Adrianto, S. (2017). Peningkatan Jumlah Mahasiswa melalui Promosi dengan Penerapan Analisa Data Mining. *SATIN - Sains Dan Teknologi Informasi*, 3(2), 29–37.
- Rahmansyah, A., Dewi, O., Andini, P., Hastuti, T., Ningrum, P., & Suryana, M. E. (2018). Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 1907–5022.
- Rudy Hendrawan, I. N., Budhi Saputra, I. M. A., Cahya Dewi, G. A. P., Adi Pranata, I. G. S., & Wedasari, N. L. N. (2022). Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes. *Jurnal Eksplora Informatika*, 11(1), 50–56. <https://doi.org/10.30864/eksplora.v11i1.606>
- Suhada, K., Elanda, A., & Aziz, A. (2021). Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang). *Dirgamaya: Jurnal Manajemen Dan Sistem Informasi*, 1(2), 14–27. <https://doi.org/10.35969/dirgamaya.v1i2.182>
- Sukarna Royan, Ansori Yulian, S. (2021). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. In *SAINTEK / Jurnal Sains & Teknologi* (Vol. 5, Issue 2, pp. 9–22).
- Uperiati, A., Bettiza, M., & Puspasari, A. (2020). Perbandingan Metode Fuzzy C-Means Dan K-Means Dalam Klasifikasi Kelulusan Mahasiswa (Studi Kasus: Jurusan Manajemen, Universitas Maritim Raja Ali Haji). *Jurnal Sustainable: Jurnal Hasil Penelitian Dan Industri Terapan*, 9(2), 75–81. <https://doi.org/10.31629/sustainable.v9i2.14097>