

# Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu

Baskoro<sup>1\*</sup>, Sriyanto<sup>2</sup>, Lilis Setya Rini<sup>3</sup>

<sup>1,2,3</sup>Magister Teknik Informatika, Institut Informatika dan Bisnis Darmajaya Lampung  
Email: <sup>1</sup>baskoro.jogja@gmail.com, <sup>2</sup>sriyanto@ darmajaya.ac.id, <sup>3</sup>orin.iphone6@gmail.com

## Abstract

Scholarships are donations from the government, institutions, and foundations given to students or students who are still studying in order to seek knowledge. This donation is usually in the form of funds to pay tuition fees incurred by students or students in pursuing education at school or in college. The number of students at Pringsewu Muhammadiyah University is currently approximately 3500 active students consisting of 4 faculties. With the amount of data used, researchers use data mining techniques for student data management in predicting scholarship recipients. Data Mining is a technology used to find new information from large amounts of data. From several algorithms, we can use it to predict scholarship recipients by adding classification techniques. From the results of the classification process, it can be formulated an analysis for the Random forests algorithm with an accuracy rate of 98.08% on the scholarship dataset with AUC graph of 0.976. The Naive Bayes algorithm gets an accuracy value of 94.23% with an AUC graph of 0.973. Then for the k-NN algorithm to get an accuracy value of 94.23% with an AUC graph of 0.845.

**Keywords :** *Scholarship, Data mining, Prediction, Classification, Algorithm*

## Abstrak

Beasiswa merupakan donasi dari pemerintah, lembaga, yayasan yang diberikan kepada siswa atau mahasiswa yang masih menempuh pendidikan dalam rangkamencari ilmu pengetahuan. Donasi ini biasanya berbentuk dana untuk membayar biaya pendidikan yang dikeluarkan siswa atau mahasiswa dalam menempuh pendidikan disekolah maupun di perguruan tinggi. Jumlah mahasiswa Universitas Muhammadiyah Pringsewu saat ini kurang lebih 3500an mahasiswa aktif yang terdiri dari 4 Fakultas. Dengan banyaknya data yang digunakan peneliti menggunakan teknik data mining untuk pengelolaan data mahasiswa dalam memprediksi penerima beasiswa. Data Mining merupakan teknologi yang digunakan untuk menemukan informasi baru dari data yang jumlahnya sangat banyak. Dari beberapa algoritma kita bisa menggunakannya untuk memprediksi penerima beasiswa dengan menambahkan teknik klasifikasi. Dari hasil proses klasifikasi maka dapat dirumuskan analisis untuk algoritma Random forests tingkat akurasi pada dataset beasiswa sebesar 98.08% dengan grafik AUC 0.976. Pada algoritma Naive Bayes mendapatkan nilai akurasi sebesar 94.23% dengan grafik AUC 0.973. Kemudian untuk algoritma k-NN mendapatkan nilai akurasi 94.23% dengan grafik AUC 0.845.

**Keywords :** *Beasiswa, Data mining, Prediksi, Klasifikasi, Algoritma*

## 1. PENDAHULUAN

Universitas Muhammadiyah Pringsewu (UMPRI) merupakan salah satu perguruan tinggi muhammadiyah yang ada di kabupaten pringsewu yang berdiri pada tahun 2019 merupakan penggabungan antara tiga sekolah tinggi muhammadiyah yaitu Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Muhammadiyah Pringsewu, Sekolah Tinggi Ilmu Kesehatan (STIKes) Muhammadiyah Pringsewu dan Sekolah Tinggi Ilmu Ekonomi (STIE) Muhammadiyah Pringsewu. Dengan banyaknya jumlah mahasiswa dan untuk menjadikan mahasiswa berprestasi dalam bidang akademik maka universitas mempunyai program beasiswa.

Beasiswa merupakan donasi dari pemerintah, lembaga, yayasan yang diberikan kepada siswa atau mahasiswa yang masih menempuh pendidikan dalam rangkamencari ilmu pengetahuan. Donasi ini biasanya berbentuk dana untuk membayar biaya pendidikan yang dikeluarkan siswa atau mahasiswa dalam menempuh pendidikan disekolah maupun di perguruan tinggi.

Jumlah mahasiswa Universitas Muhammadiyah Pringsewu saat ini kurang lebih 3500an mahasiswa aktif yang terdiri dari 4 Fakultas. Program beasiswa dirancang untuk digunakan bagi mahasiswa yang berprestasi sebagai bentuk apresiasi atau penghargaan terhadap mahasiswa tersebut, program beasiswa juga dibentuk sebagai motivasi mahasiswa untuk mendapatkan IPK tinggi. Salah satu permasalahan yang dihadapi dalam penyaluran dana beasiswa adalah masih adanya pemberian beasiswa yang kurang tepat, sehingga menyebabkan kurang maksimalnya penggunaan dana beasiswa tersebut, prosesnya pun masih dilakukan secara manual terkadang membutuhkan waktu

cukup lama sehingga kurang efisien. Selain itu banyaknya jumlah peserta didik dan terbatasnya jumlah beasiswa yang diberikan karena universitas ini tergolong baru dan merupakan perguruan tinggi swasta maka dibutuhkan suatu metode untuk menentukan mahasiswa penerima beasiswa sehingga tidak melebihi kapasitas yang telah ditentukan oleh pihak penyedia beasiswa. Dengan banyaknya data yang digunakan peneliti menggunakan teknik data mining untuk pengelolaan data mahasiswa.

Data Mining merupakan teknologi yang digunakan untuk menemukan informasi baru dari data yang jumlahnya sangat banyak. Perkembangan yang pesat di bidang pengumpulan data serta teknologi penyimpanan pada banyak sekali bidang, menghasilkan data yang terlalu besar. data yang dikumpulkan tersebut akan diteliti kembali karena terlalu panjang, membosankan, serta tidak menarik. keputusan yang dibuat berdasarkan data seringkali dibuat tidak lagi sesuai data, melainkan hanya ide dari para pembuat keputusan, oleh sebab itu maka lahirlah ilmu data mining ini. Pada dasarnya data mining dibedakan menjadi dua fungsionalitas, yaitu deskripsi dan prediksi, kemudian terdapat beberapa macam algoritma seperti *C4.5*, *k-Means*, *SVM* atau *Support Vector Machine*, *Apriori*, *EM*, *PageRank*, *AdaBoost*, *kNN*, *Naive Bayes*. Dari beberapa algoritma yang disebutkan kita bisa menggunakannya untuk memprediksi penerima beasiswa dengan menambahkan teknik klasifikasi.

Beberapa penelitian telah dilakukan untuk memprediksi penerima beasiswa agar supaya beasiswa yang diberikan memang pantas diterima oleh mahasiswa tersebut. beberapa metode yang sering digunakan ialah Decision Tree, k-Nearest Neighbor dan Naive Bayes seperti penelitian yang dilakukan oleh Heni Sulistiani dan Yohana Tri Utami dengan hasil bahwa metode Naive Bayes dengan penggunaan data 10% untuk testing mampu memberikan nilai akurasi sebesar 80%. Selain itu, Choirul Anam dan Harry Budi Santoso melakukan perbandingan antara Algoritma C4.5 dan Naive Bayes untuk klasifikasi penerima beasiswa. Hasilnya algoritma C4.5 menghasilkan akurasi 96.40% lebih baik dari algoritma Naive Bayes yaitu 95.11%.

Peneliti ingin melakukan pengujian dari penerapan teknik data mining untuk melihat nilai akurasi mana yang lebih baik agar supaya bisa membantu dalam prediksi penerima beasiswa di Universitas Muhammadiyah Pringsewu. Diharapkan dengan penerapan teknik data mining tersebut mampu meminimalisir kesalahan dari penyedia beasiswa, sehingga penerima beasiswa dapat dilakukan dengan cepat dan efisien serta pembagiannya merata tidak melebihi kapasitas.

## 2. KERANGKA TEORI

### 2.1. Prediksi

Dalam arti non-statistik, istilah "prediksi" sering digunakan untuk merujuk pada tebakan atau opini yang diinformasikan. Prediksi semacam ini mungkin diinformasikan oleh penalaran abduktif, penalaran induktif, penalaran deduktif, dan pengalaman seseorang yang memprediksi; dan mungkin berguna jika orang yang memprediksi adalah orang yang berpengetahuan luas di bidangnya. *The Metode Delphi* adalah teknik untuk memunculkan prediksi berbasis ahli-penilaian seperti dengan cara yang terkontrol. Jenis prediksi ini mungkin dianggap konsisten dengan teknik statistik dalam arti bahwa, minimal, "data" yang digunakan adalah pengalaman kognitif pakar prediksi yang membentuk "kurva probabilitas" intuitif.

Dalam statistika, prediksi merupakan bagian dari inferensi statistik. Salah satu pendekatan khusus untuk inferensi semacam itu dikenal sebagai inferensi prediktif, tetapi prediksi dapat dilakukan dalam salah satu dari beberapa pendekatan untuk inferensi statistik. Memang, satu gambaran statistik yang mungkin adalah bahwa statistik menyediakan sarana untuk mentransfer pengetahuan tentang sampel populasi ke seluruh populasi, dan populasi terkait lainnya, yang tidak selalu sama dengan prediksi dari tahun ke tahun. Ketika informasi ditransfer lintas waktu, seringkali ke titik waktu tertentu, prosesnya dikenal sebagai peramalan. Peramalan biasanya membutuhkan metode time series, sedangkan prediksi sering dilakukan pada data cross-sectional.

Teknik statistik yang digunakan untuk prediksi termasuk analisis regresi dan berbagai sub-kategorinya seperti regresi linier, model linier umum (regresi logistik, regresi Poisson, regresi Probit), dll. Dalam hal peramalan, model rata-rata bergerak autoregresif dan model regresi vektor dapat dimanfaatkan. Saat ini dan/atau terkait, kumpulan regresi atau metode pembelajaran mesin yang diterapkan dalam penggunaan komersial, bidang ini dikenal sebagai analitik prediktif.

### 2.2. Data Mining

Data mining merupakan sebuah proses ekstraksi untuk menemukan sebuah pola yang berada didalam sejumlah data yang sangat luas dan besar yang melibatkan metode pembelajaran mesin, database dan ilmu statistik. Data mining adalah subbidang interdisipliner ilmu komputer dan statistik dengan tujuan keseluruhan untuk mengekstrak informasi dengan metode cerdas dari kumpulan data dan mengubah informasi menjadi struktur yang dapat dipahami untuk digunakan lebih lanjut. Proses "penemuan pengetahuan dalam database", atau KDD merupakan langkah analisis dari data mining. Selain langkah analisis mentah, juga melibatkan aspek database dan manajemen data, pra-

pemrosesan data, pertimbangan model dan inferensi, metrik ketertarikan, pertimbangan kompleksitas, pasca-pemrosesan struktur yang ditemukan, visualisasi, dan pembaruan online. Istilah data mining dari sejumlah data yang diekstraksi pola dengan pengetahuan itu adalah sebuah kesalahan karena bukan ekstraksi dari data mining itu sendiri. Ini juga merupakan kata kunci dan sering diterapkan pada segala bentuk pemrosesan data atau informasi skala besar pengumpulan ekstraksi pergudangan analisis dan statistik serta aplikasi sistem pendukung keputusan komputer termasuk Kecerdasan buatan.

Pengumpulan data, persiapan data, maupun interpretasi dan pelaporan hasil bukanlah bagian dari langkah penambangan data, tetapi termasuk dalam proses KDD secara keseluruhan sebagai langkah tambahan. Perbedaan antara analisis data dan penggalian data adalah bahwa analisis data digunakan untuk menguji model dan hipotesis pada kumpulan data, misalnya, menganalisis efektivitas kampanye pemasaran, berapa pun jumlah datanya; sebaliknya, penambangan data menggunakan pembelajaran mesin dan model statistik untuk mengungkap pola rahasia atau tersembunyi dalam volume data yang besar.

Istilah terkait pengumpulan data, penangkapan data, dan pengintaian data mengacu pada penggunaan metode penambangan data untuk mengambil sampel bagian dari kumpulan data populasi yang lebih besar yang (atau mungkin) terlalu kecil untuk membuat kesimpulan statistik yang andal tentang validitas dari setiap pola ditemukan. Metode ini dapat, bagaimanapun, digunakan dalam membuat hipotesis baru untuk menguji populasi data yang lebih besar.

### 2.3. Klasifikasi

Klasifikasi merupakan tugas mensurvei objek informasi untuk memasukan kedalam kelas tertentu dari jumlah kelas yang dapat diakses. klasifikasi dapat dicirikan sebagai tugas yang melakukan persiapan/pembelajaran pada kapasitas objektif yang memetakan setiap set properti (sorotan) ke satu nomor tanda kelas yang dapat diakses. Sebuah framework yang melakukan klasifikasi diharuskan memiliki pilihan untuk mengkarakterisasi semua indeks informasi secara efektif, namun tidak dapat dipungkiri bahwa eksekusi framework tidak dapat 100% benar sehingga situasi klasifikasi juga harus mengukur presentasinya. Sebagian besar, estimasi eksekusi klasifikasi diselesaikan dengan kisi-kisi yang kacau.

Dalam statistik, klasifikasi adalah masalah mengidentifikasi yang mana dari satu set kategori (sub-populasi) pengamatan, (atau pengamatan) milik. Seringkali, pengamatan individu dianalisis ke dalam satu set sifat terukur, yang dikenal sebagai variabel penjelas atau fitur. Sifat-sifat ini bisa bermacam-macam kategori (misalnya "A", "B", "AB" atau "O", untuk golongan darah), ordinal (misalnya "besar", "sedang" atau "kecil"), bernilai integer (misalnya jumlah kemunculan kata tertentu dalam email) atau bernilai nyata (misalnya pengukuran tekanan darah).

Sebuah algoritma yang mengimplementasikan klasifikasi, terutama dalam implementasi konkret, dikenal sebagai classifier. Istilah "pengklasifikasi" terkadang juga mengacu pada fungsi matematika, yang diimplementasikan oleh algoritme klasifikasi, yang memetakan data masukan ke suatu kategori.

Terminologi lintas bidang cukup bervariasi. Dalam statistik, di mana klasifikasi sering dilakukan dengan regresi logistik atau prosedur serupa, sifat-sifat pengamatan disebut variabel penjelas (atau variabel bebas, regresi, dll.), dan kategori yang diprediksi dikenal sebagai hasil, yang dianggap sebagai menjadi nilai-nilai yang mungkin dari variabel dependen. Dalam pembelajaran mesin, pengamatan sering dikenal sebagai instance, variabel penjelas disebut fitur (dikelompokkan ke dalam vektor fitur), dan kategori yang mungkin untuk diprediksi adalah kelas. Bidang lain mungkin menggunakan terminologi yang berbeda: misalnya dalam ekologi komunitas, istilah "klasifikasi" biasanya mengacu pada analisis kluster.

### 2.4. Confusion matrix

*Confusion Matrix* juga dikenal sebagai *error matrix* merupakan tata letak tabel khusus yang memungkinkan visualisasi kinerja suatu algoritma, biasanya pembelajaran yang diawasi (dalam pembelajaran tanpa pengawasan biasanya disebut matriks yang cocok). Setiap baris matriks mewakili instance di kelas aktual sementara setiap kolom mewakili instance di kelas yang diprediksi, atau sebaliknya – kedua varian ditemukan dalam literatur. Nama berasal dari fakta bahwa memudahkan untuk melihat apakah sistem ini membingungkan dua kelas (yaitu biasanya salah memberi label satu sama lain). Dalam penelitian ini menggunakan perhitungan akurasi dimana jumlah data yang telah diklasifikasikan secara benar dan salah dibagi dengan total sampel data testing.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

### 3. METODOLOGI

#### 3.1. Alur Penelitian

Alur penelitian dimulai pengumpulan data, data yang diperoleh tersebut diolah menjadi data set, kemudian melakukan pengujian cross validation setiap model metode, setelah itu peneliti melakukan evaluasi dan yang terakhir peneliti melakukan perbandingan akurasi kinerja.



Gambar 1. Alur Penelitian

#### 3.2. Pengumpulan Data

##### a) Observasi

Melakukan observasi ke beberapa fakultas dengan tujuan untuk mengumpulkan data-data yang diperlukan dalam menyusun tesis ini. Data-data yang diperlukan adalah data mahasiswa dan mahasiswi dari 4 fakultas yang nantinya akan berkaitan dengan penentuan penerimaan beasiswa.

##### b) Wawancara

Upaya mendapatkan data yang lebih jelas, maka dilakukan wawancara secara langsung dengan bagian kemahasiswaan yang ada di Universitas Muhammadiyah Pringsewu tentang bagaimana ketentuan penerimaan beasiswa di kampus tersebut.

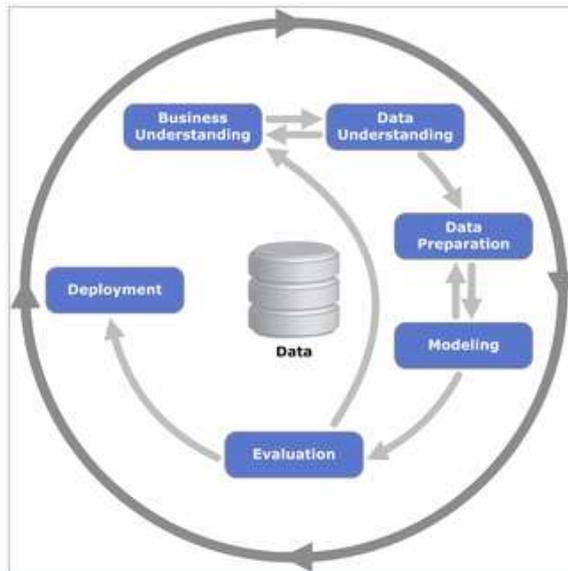
#### 3.3 Pengolahan dan Analisa data

Pengolahan data awal menjadi dataset mahasiswa yaitu dengan menentukan atribut yang digunakan untuk menentukan prediksi penerima beasiswa. Gambar tabel dibawah ini akan menjelaskan tentang atribut yang digunakan.

Tabel 1. Tabel Atribut Dataset

Atribut	Nilai
Jarak Tempat Tinggal kekampus	Jauh/Dekat
JenisKelamin	L/P
Ikut Organisasi	Ikut/Tidak
Ikut UKM	Ikut/Tidak
IPK	Sangat Baik/Baik/Cukup
Pekerjaan Orang Tua	PNS/Karyawan Swasta/Wiraswasta/Pedagang/Petani
Penghasilan	Rendah/Sedang/Tinggi
Tanggungan	Banyak/Sedang/Sedikit

Analisis data menunjukkan bahwa metode CRISP-DM adalah metodologi utama yang digunakan oleh para penambang data. Proses standar lintas industri untuk data mining (CRISP-DM) yang mendefinisikan enam fase yaitu Pemahaman bisnis, Pemahaman data, Persiapan data, Pemodelan, Evaluasi, dan Penyebaran atau proses yang disederhanakan seperti Pra-pemrosesan, Data Mining, dan Validasi Hasil.



Gambar 2. Diagram proses CRISP-DM

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Hasil

Pengolahan dataset yang dibagi menjadi dataset training dan testing 90%:10% dengan jumlah 596 record yang terdiri dari beberapa atribut. Data tersebut bisa dilihat pada table dibawah ini atau untuk melihat lebih jelasnya di [https://bit.ly/Dataset\\_Beasiswa](https://bit.ly/Dataset_Beasiswa).

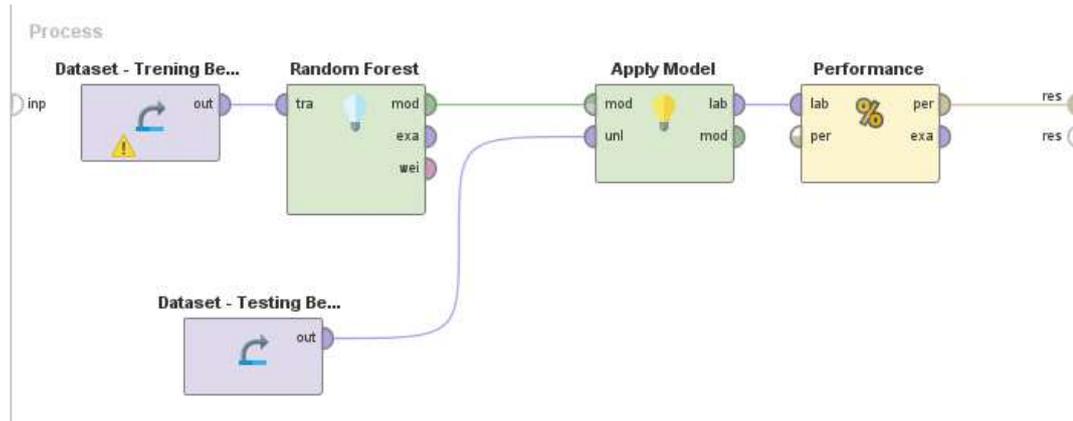
Tabel 2. Potongan Dataset

Ikut UKM <i>polynomial</i>	IPK <i>real</i>	Pekerjaan ... <i>polynomial</i>	Penghasilan <i>polynomial</i>	Tanggung <i>integer</i>	Status Bea... <i>binomial label</i>
Ikut	3.570	Wiraswasta	Sedang	4	Terima
Ikut	2.950	Buruh	Sedang	2	Tidak
Ikut	3.670	Petani	Sedang	4	Terima
Ikut	3.190	Wiraswasta	Tinggi	2	Tidak
Ikut	3.190	Wiraswasta	Sedang	2	Tidak
Ikut	3.780	Wiraswasta	Rendah	3	Terima
Ikut	3.330	Petani	Sedang	3	Tidak
Ikut	3.290	Petani	Rendah	1	Tidak
Ikut	3.480	Wiraswasta	Sedang	3	Tidak
Ikut	3.380	Pedagang	Sedang	1	Tidak
Ikut	3.520	Petani	Rendah	2	Tidak
Ikut	2.810	Pedagang	Rendah	2	Tidak

##### 4.2. Pembahasan

###### 1. Eksperimen dan Pengujian

Hasil dua file dataset kemudian dilakukan pengujian model yang dilakukan satu persatu dengan beberapa algoritma untuk membandingkan algoritma mana yang lebih akurat. Pengujian dengan metode *Random forests* yang merupakan metode klasifikasi dengan data yang jumlahnya besar, klasifikasi tersebut menggunakan gabungan pohon keputusan pada sampel data yang dimiliki dilakukan training. Langkah pengolahan dataset beasiswa menggunakan *tools* rapid miner dengan menggunakan metode *random forests*. pertama menggunakan dua file dataset yang telah di olah data training dihubungkan dengan metode *random forests*, data testing dihubungkan langsung ke modul *apply model*, kemudian keduanya dihubungkan ke modul *performance*.



Gambar 3. Skema pengujian dengan rapid miner

2. Evaluasi dan Validasi Hasil

Setelah data di olah menggunakan *tools* rapid miner maka dapat di uji tingkat akurasi nya dilakukan dengan menggunakan *confusion matrix*. dan kurva ROC/AUC. (*Area Under Cover*).

Tabel 3. Confusion Matrix Random forests

accuracy: 98.08%

	true Terima	true Tidak	class precision
pred. Terima	4	0	100.00%
pred. Tidak	1	47	97.92%
class recall	80.00%	100.00%	

Tabel 4.2 merupakan hasil perhitungan akurasi data testing menggunakan Algoritma *Random Forests*, Tingkat akurasi nya 98.08%. Dari 52 data testing dikelompokkan menjadi dua kelas yaitu “Terima” dan ”Tidak”. Dimana terdapat lima data yang masuk dalam kelompok “Terima” dan empat puluh tujuh masuk kedalam kelompok “Tidak”. Dari kelas “Terima” terdapat empat data yang diklasifikasikan dengan benar dan satu data diklasifikasikan dengan salah. Sedangkan kelas “Tidak” terdapat nol data yang diklasifikasikan dengan benar dan empat puluh tujuh data diklasifikasikan dengan salah.



Gambar 4. Grafik AUC Random forests

Gambar 4.2 adalah grafik AUC (Area Under Cover) dari model Algoritma Random Forests yaitu 0.976. Garis horizontal adalah false positif dan garis vertikal false negatif. *Performance* keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu:

1. 0.90 – 1.00 = *Excellent Classification*
2. 0.80 – 0.90 = *Good Classification*
3. 0.70 – 0.80 = *Fair Classification*
4. 0.60 – 0.70 = *Poor Classification*
5. 0.50 – 0.60 = *Failure Classification*

Berdasarkan klasifikasi tersebut maka dapat disimpulkan bahwa Algoritma *Random Forests* termasuk algoritma yang akurat untuk memprediksi penerima beasiswa karena nilai AUC termasuk dalam predikat *Excellent Classification* (0.90 – 1.00).

## 5. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan untuk prediksi penerima beasiswa menggunakan tiga algoritma yaitu Random Forests, Naive Bayes dan k-NN, karena kinerja algoritma Random forests lebih baik dari pada algoritma Naive Bayes dan k-NN, Dalam Hal ini peneliti bisa dilihat dari tingkat akurasi maupun grafik AUC yang didapatkan. Algoritma Random forests lebih baik dari dari pada algoritma Naive Bayes dan k-NN karena kemampuan klasifikasi dengan data yang jumlahnya besar, klasifikasi tersebut menggunakan gabungan pohon keputusan pada sampel data yang dimiliki untuk dilakukan training. Untuk penelitian berikutnya diperlukan penelitian lebih lanjut mengenai atribut yang mempengaruhi pengolahan data yang digunakan, serta menggunakan lebih banyak data guna meningkatkan kualitas hasil penelitian.

## UCAPAN TERIMA KASIH

Pada kesempatan ini penulis mengucapkan rasa syukur dan terimakasih kepada :

1. Keluarga tercinta yang selalu memberikan dukungan dalam penulisan artikel ini
2. Bapak Sriyanto yang selalu mengingatkan dan memberikan masukan dalam penulisan artikel ini
3. Pengelola kampus Universitas Muhammadiyah Pringsewu
4. Teman-teman seperjuangan MTI IIB Darmajaya.

## DAFTAR PUSTAKA

- Fitriani, E. (2020) 'Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan', *Sistemasi*, P. 103. Doi: 10.32520/Stmsi.V9i1.596.
- Hasan, F. N., Hikmah, N. And Utami, D. Y. (2018) 'Perbandingan Algoritma C4.5, Knn, Dan Naive Bayes Untuk Penentuan Model Klasifikasi Penanggung Jawab Bsi Entrepreneur Center', *Jurnal Pilar Nusa Mandiri*, P. 169. Doi: 10.33480/Pilar.V14i2.908.
- Sabita, H., Fitria, F., & Herwanto, R. (2021). Analisa Dan Prediksi Iklan Lowongan Kerja Palsu Dengan Metode Natural Language Programing Dan Machine Learning. *Jurnal Informatika*, 21(1), 14-22.
- Linawati, S. Et Al. (2020) 'Perbandingan Algoritma Klasifikasi Naive Bayes Dan Svm Pada Studi Kasus Pemberian Penerima Beasiswa Ppa', *Swabumi*, Pp. 71–75. Doi: 10.31294/Swabumi.V8i1.7708.
- Nalattisifa, H. Et Al. (2021) 'Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (Svm), Dan Random Forest Untuk Prediksi Ketidakhadiran Di Tempat Kerja', *Jurnal Informatika Universitas Pamulang*, P. 578. Doi: 10.32493/Informatika.V5i4.7575.
- Nanni, N. And Sudransyah, A. (2020) 'Perbandingan Kinerja Algoritma Naive Bayes Dan C4.5 Untuk Klasifikasi Harga Pangan', *Protek : Jurnal Ilmiah Teknik Elektro*, Pp. 20–24. Doi: 10.33387/Protk.V7i1.1710.
- Putra, Y. K. Et Al. (2019) 'Perbandingan Algoritma Naive Bayes Dan Naive Bayes Berbasis Pso Untuk Analisis Kredit Pada Pt. Bpr Syariah Paokmotong', *Infotek : Jurnal Informatika Dan Teknologi*, Pp. 61–69. Doi: 10.29408/Jit.V2i2.1460.
- Riyanah, N. And Fatmawati, F. (2021) 'Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu', *Jtim : Jurnal Teknologi Informasi Dan Multimedia*, Pp. 206–213. Doi: 10.35746/Jtim.V2i4.117.
- Saputra, P. S. (2021) 'Perbandingan Algoritma Fuzzy C-Means Dan Algoritma Naive Bayes Dalam Menentukan Keluarga Penerima Manfaat (Kpm) Berdasarkan Status Sosial Ekonomi (Sse) Terendah', *Jst (Jurnal Sains Dan Teknologi)*, P. 1. Doi: 10.23887/Jst-Undiksha.V10i1.23340.

- Wibowo, H., & Indriyani, F. (2018, October). K-Nearest Neighbor Method For Monitoring Of Production And Preservation Information (*Treatment*) Of Rubber Tree Plant. In *International Conference on Information Technology and Business (ICITB)* (pp. 29-44).
- Setifani, N. A., Fitriana, D. N. And Yusuf, A. (2020) 'Perbandingan Algoritma Naïve Bayes, Svm, Dan Decision Tree Untuk Klasifikasi Sms Spam', *Jusim (Jurnal Sistem Informasi Musirawas)*, Pp. 153–160. Doi: 10.32767/Jusim.V5i02.956.
- Siswa, T. A. Y. (2018) 'Analisis Penerapan Optimasi Perbandingan Kinerja Algoritma C4.5 Dan Naïve Bayes Berbasis Particle Swarm Optimization (Pso) Untuk', *Jurnal Bangkit Indonesia*, P. 1. Doi: 10.52771/Bangkitindonesia.V7i2.48.
- Pratama, T. (2018). Implementasi Algoritma Naive Bayes Dalam Menentukan Konsentrasi Skripsi Dan Rekomendasi Bahasa Pemrograman. *Jurnal Informatika*, 18(1), 1-13.  
Website: <https://id.wikipedia.org/> diakses tanggal 24 Mei.
-