

## Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)

Nurhidayati<sup>1\*</sup>, Yahya<sup>2</sup>, Fathurrahman<sup>3</sup>, L.M Samsu<sup>4</sup>, Wajizatul Amnia<sup>5</sup>

<sup>1,2,3,4</sup>Program Studi Sistem Informasi, Universitas Hamzanwadi

<sup>5</sup>Program Studi Teknik Lingkungan, Universitas Hamzanwadi

\*[hidayati2188@gmail.com](mailto:hidayati2188@gmail.com)

### Abstrak

Setiap tahun Universitas selalu menawarkan berbagai jenis beasiswa kepada mahasiswa tidak terkecuali Universitas Hamzanwadi Selong Lombok Timur. Salah satu jenis beasiswa yang ditawarkan adalah beasiswa Bidikmisi (KIP/K) yang diperuntukkan bagi mahasiswa yang memiliki tingkat ekonomi menengah ke bawah, tapi memiliki prestasi atau potensi akademik yang baik. Setiap tahun jumlah pendaftar beasiswa ini terus mengalami peningkatan namun jumlah yang diterima tiap tahunnya terbatas. Banyaknya tumpukan berkas yang mengajukan beasiswa serta proses seleksi yang manual, cenderung tidak efektif dan efisien dan hasil seleksi kurang akurat, maka diperlukan sebuah sistem yang mampu membantu dalam proses seleksi dengan cepat, mudah dan tepat sasaran. Metode yang digunakan adalah CRIS-DM dengan pemodelan algoritma Naive Bayes Classifier, metode ini merupakan pendekatan yang mengacu pada teorema Bayes yang mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Adapun variabelnya terdiri dari 7 atribut, yakni : Nama, Status DTKS, prestasi, pekerjaan orang tua, jumlah penghasilan orang tua, kepemilikan rumah, dan jumlah tanggungan keluarga[1]. Pengujian dilakukan dengan menggunakan k-fold cross validation, dan didapatkan hasil akurasi tertinggi dari k-fold 4 sebesar 91,43%, sementara AUC sebesar 0,996% dengan diagnosa excellent classification. Dengan demikian dapat disimpulkan, bahwa algoritma naive bayes ini sangat baik digunakan dalam seleksi kelulusan pemberkasan beasiswa bidikmisi pada Universitas Hamzanwadi.

**Kata kunci :** AUC, Beasiswa KIP/K, CRIS-DM, Cross Validation, Excellent classification, Naive Bayes classifier

### Abstract

Every year the University offers various types of scholarships to its students, including Hamzanwadi Selong University, East Lombok. One type of scholarship offered is the Bidikmisi scholarship (KIP/K) which is intended for students who have middle to lower economic levels but have good academic achievement or potential. Every year the number of applicants for this scholarship continues to increase, but the number received each year is limited. The large number of piles of files applying for scholarships and the manual selection process tends to be ineffective and efficient and the results of the selection are inaccurate, so a system is needed that is able to assist in the selection process quickly, easily and on target. The method used is CRIS-DM with the Naive Bayes Classifier algorithm modeling, this method is an approach that refers to the Bayes theorem which combines previous knowledge with new knowledge. The variables consist of 7 attributes, namely: Name, DTKS status, achievement, parents' occupation, total income of parents, home ownership, and number of family dependents[1]. Testing was carried out using k-fold cross validation, and the highest accuracy results were obtained from k-fold 4 of 91.43%, while the AUC was 0.996% with a very good diagnostic classification. Thus it can be interpreted that the Naive Bayes algorithm is very well used in the selection of scholarships for bidikmisi scholarships at Hamzanwadi University.

**Keywords :** AUC, Beasiswa KIP/K, CRIS-DM, Cross Validation, Excellent classification, Naive Bayes classifier.

## 1. Pendahuluan

Setiap warga negara Indonesia berhak mendapatkan pendidikan sebagaimana tercantum dalam UUD 1945 pasal 31 ayat 1-2, namun pada kenyataannya tidak semua warga negara Indonesia mendapatkan kesempatan untuk memperoleh pendidikan yang layak dan sesuai serta berkelanjutan. Oleh karena itu peserta didik yang kurang mampu berhak untuk memperoleh pendidikan dan bagi peserta didik yang berprestasi berhak mendapatkan beasiswa. Sebagaimana yang tercantum dalam

Undang-undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, menyebutkan bahwa setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan Beasiswa bagi yang berprestasi yang orang tuanya kurang mampu membiayai pendidikannya. Beasiswa bidik misi atau yang sekarang disebut dengan beasiswa kartu Indonesia pintar kuliah (KIP-K) merupakan program beasiswa yang diperuntukkan bagi mahasiswa tidak mampu dan berprestasi yang diselenggarakan oleh dikti[2].

Berdasarkan identifikasi dan observasi peneliti dengan salah satu staf biro kemahasiswaan di Universitas Hamzanwadi mengenai sistem seleksi penerima beasiswa menjelaskan bahwa prosesnya masih belum memiliki sistem informasi seleksi beasiswa yang terpadu. Perangkat yang digunakan masih konvensional berupa buku arsip dan Microsoft excel. Seleksi penerimaan

beasiswa yang dilaksanakan secara konvensional cenderung tidak efektif dan efisien, dan hasil seleksi kurang akurat karena hanya mengandalkan pengamatan indera penglihatan, dimana dalam kondisi tertentu bisa saja terjadi kesalahan (human error). Selain itu permasalahan yang lain adalah, tiap tahun berkas pengajuan beasiswa bidikmisi ini selalu bertambah, sementara kuota yang disediakan perguruan tinggi terbatas dan membutuhkan waktu yang lama dalam proses seleksi. Sedangkan penerima beasiswa harus tepat sasaran sesuai dengan kriteria yang berlaku. Banyaknya peserta calon yang mengajukan beasiswa mengakibatkan penentuan calon penerima beasiswa yang kurang tepat. Untuk itu perlu klasifikasi untuk membantu menentukan keputusan penerima beasiswa bidikmisi tersebut. Adapun algoritma klasifikasi yang digunakan dalam penelitian ini adalah Naive Bayes classifier, merupakan suatu metode klasifikasi dalam data mining dengan menggunakan metode probabilitas dan statistik. Algoritma ini merupakan algoritma klasifikasi sederhana namun memiliki akurasi yang baik dan handal sehingga bisa digunakan untuk proses seleksi penerima beasiswa agar tepat sasaran. Metode naive bayes memiliki kelebihan, yaitu cepat dalam perhitungan, algoritma yang sederhana dan berakurasi tinggi. Metode Naive Bayes hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk

menentukan parameter yang diperlukan dalam proses pengklasifikasian. Metode Algoritma Naïve Bayes Classifier lebih mudah digunakan karena memiliki alur perhitungan yang tidak panjang[3].

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

Untuk mendukung penelitian ini, peneliti menggunakan beberapa penelitian sebelumnya, antara lain sebagai berikut :

- Antony Anwari Rahman dan Agus Suryanto yang berjudul "Implementasi Sistem Informasi Seleksi Penerima Beasiswa Dengan Metode Naive Bayes Classifier" Metode penelitian yang digunakan Research and Development (R&D). Perangkat dibuat dengan framework Microsoft Visual Studio 2013. Uji kelayakan dilakukan dengan pengujian 1 dan 2. Hasil penelitian dari pengujian 1 sebesar 96,56%, dan pengujian 2 sebesar 90,33%, termasuk dalam kategori sangat layak. Pada penelitian ditemukan bahwa sistem menggunakan algoritma Naive Bayes Classifier menghasilkan akurasi yang tinggi. Sistem dengan akurasi tinggi dan efisien bermanfaat dalam menentukan calon penerima beasiswa. Manfaat lainnya berupa pengembangan ilmu pengetahuan tentang implementasi data mining yaitu algoritma Naive Bayes Classifier[4].
- Penerapan Naive Bayesian Classifier Dalam Penyeleksian Beasiswa Ppa, Pada penelitian ini,

penulis menggunakan metode Naïve Bayes Classifier untuk melakukan klasifikasi penerima beasiswa di Universitas Budi Darma. Penggunaan metode Naïve Bayes Classifier dipilih untuk melakukan pengambilan keputusan agar dapat menentukan diterima dan ditolak mahasiswa yang mendaftar program beasiswa. Metode classifier yaitu untuk mengklasifikasikan teks berdasarkan probabilitas tertinggi yang ditetapkan sebagai kategori dokumen baru. Dari hasil penelitian ini, maka mendapatkan hasil bahwa nilai "Menerima" lebih tinggi daripada "Tidak" yaitu 0.0351 dibanding 0 maka dapat disimpulkan mahasiswa atas nama Riska Ramadhani menerima beasiswa PPA 2019[5].

- Gagan Suganda dkk, dengan judul "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier" Penelitian ini bertujuan untuk membantu mengklasifikasi data-data calon penerima bantuan KIP Kuliah menggunakan algoritma Naïve Bayes Classifier. Dataset yang digunakan dalam penelitian ini sebanyak 100 data calon penerima bantuan KIP Kuliah diperiode 2019/2020. Dataset tersebut dibagi menjadi 90 data training serta 10 data uji. Kriteria penentuan klasifikasi diperoleh dari hasil verifikasi data calon penerima bantuan KIP Kuliah, seperti penghasilan ayah dan ibu, pendapatan ayah dan ibu, status ayah dan ibu, jumlah tanggungan, kepemilikan rumah dan MCK, sumber air dan

listrik, luas tanah dan bangunan, prestasi, serta hasil tes ujian yang telah diikuti. Hasil pengujian klasifikasi terhadap 10 data uji diperoleh 6 data terklasifikasi 'Layak' sedangkan 4 data uji terklasifikasikan 'tidak layak'. Sedangkan hasil pengujian terhadap implementasi sistem yang dirancang diperoleh persentase tingkat kesesuaian sebesar 88.21% yang artinya aplikasi yang telah dibangun telah sesuai dengan kebutuhan[6].

- Said Iskandar dkk, "Metode Naive Bayes Classifier Dalam Penentuan Penerima Beasiswa Bidikmisi Di Universitas Negeri Medan", Penelitian ini adalah Studi Kasus yang dilakukan di Jurusan Matematika Angkatan 2019 Universitas Negeri Medan. Data yang diambil adalah pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan, daya listrik (watt), dan nilai ujian nasional. Adapun beberapa tahapan dari proses prosedur dari Klasifikasi Naive Bayes Classifier (NBC) yaitu : Pengumpulan Data, Data Cleaning, Data Transformation, dan Proses Perhitungan Naive Bayes Classifier. Dari hasil penelitian didapat Pengujian pada perbandingan data training dan data testing sebesar 80:20 menghasilkan akurasi tertinggi dengan 79% dan dilihat Metode Naive Bayes Classification digunakan untuk mengklasifikasikan Penerima beasiswa menghasilkan akurasi yang baik[7].

- Klasifikasi Mahasiswa Penerima Program Beasiswa Bidik Misi Menggunakan Naive Bayes,

Dalam penelitian ini, digunakan metode Naive Bayes classifier dengan menggunakan variabel penghasilan orang tua, prestasi akademik, prestasi non akademik, uang kuliah tunggal, daya listrik serta jumlah tanggungan keluarga. Keenam variabel tersebut digunakan dalam proses klasifikasi sebagai indikator penentuan kelayakan mahasiswa mendapatkan beasiswa. Sistem yang dibuat dengan metode Naive Bayes classifier sudah baik, dibuktikan dengan hasil akurasi hingga 83,33 %[2].

## 2.2. Landasan Teori

### 1. Naive Bayes

Naive Bayes Classifier adalah algoritma yang efisien dan efektif untuk pembelajaran mesin dan penambangan data. Berbasiskan Naive Bayes Classifier pada kombinasi Bayes Theorem dan atribut independensi anggapan. Naive Bayes Classifier didasarkan pada asumsi yang disederhanakan nilai-nilai atribut independen secara kondisional, dengan asumsi dari nilai target yang diberikan[8]. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode Naive Bayes juga dinilai berpotensi baik dalam mengklasifikasikan dokumen dibandingkan dengan metode

pengklasifikasian lain dalam hal akurasi dan efisiensi komputasi (Bustami,2014)[9].

$$P(H | X) = \frac{P(X | H) \times P(H)}{P(X)} \dots \dots \dots (1)$$

Penjelasan dari persamaan (1) sebagai berikut:

X : Data dengan class yang belum diketahui.

H : Hipotesis pada data X yang merupakan suatu class khusus.

P(H|X) : Nilai probabilitas pada hipotesis H berdasarkan kondisi X.

P(H) : Nilai probabilitas pada hipotesis H.

P(X|H) : Nilai probabilitas X yang berdasarkan dengan kondisi H.

P(X) : Nilai probabilitas pada X.

## 2. Teknik klasifikasi

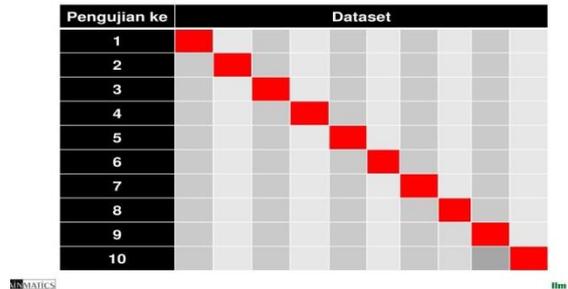
Dalam klasifikasi, terdapat dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai prototype untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya[2].

### a). K-Fold Cross Validation

Salah satu metode cross-validation yang populer adalah K-Fold Cross Validation. Dalam teknik ini dataset dibagi menjadi sejumlah K-buah partisi secara acak. Kemudian dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai data testing dan memanfaatkan sisa partisi lainnya sebagai data training. Metode cross-validation

digunakan untuk menghindari overlapping pada data testing[10].

■ Merah: k-subset (data testing)



Gambar 1. Pengujian Cross Validation

Dari gambar diatas bahwa bisa di jelaskan warna merah yang berada pada data set merupakan data testing dan selebihnya merupakan data training. cross validation merupakan metode evaluasi standar yaitu menggunakan 10 kali pengujian. Hasil dari penelitian beserta teori menunjukkan bahwa untuk mendapatkan hasil validasi yang akurat menggunakan 10-fold crossvalidation (pengujian 10 kali menggunakan cross validation). 10-fold cross-validation akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian[11].

### b). Confusion Matrix

Confusion Matrix ini meliputi informasi tentang kelas yang sebenarnya dan kelas prediksi. Confusion Matrix adalah salah satu alat ukur berbentuk matrik 2x2 yang digunakan untuk mendapatkan jumlah ketepatan algoritma yang dipakai[3].

Rumus untuk menghitung nilai akurasi adalah sebagai berikut [12]:

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

Dimana :

TP = Proporsi kasus positif yang diidentifikasi dengan benar

TN = Proporsi kasus negatif yang diklasifikasikan dengan benar

FP = Proporsi kasus negatif yang salah diklasifikasikan sebagai positif

FN = Proporsi kasus positif yang salah diklasifikasikan sebagai negatif

#### c). AUC (Area Under The Curve)

Area under the curve (AUC) adalah suatu daerah di bawah receiver operating characteristic (ROC). Receiver operating characteristic (ROC) merupakan kurva yang dihasilkan dari tarik ulur antara sensitivitas dan spesifisitas pada berbagai titik potong. Nilai AUC secara teoritis berada di antara 0 dan 1. Nilai AUC memberikan gambaran tentang keseluruhan pengukuran atas kesesuaian dari model yang digunakan. Semakin besar area under curve maka semakin baik variabel yang diteliti dalam memprediksi kejadian[13]

#### d). Rapid Miner

Rapidminer adalah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi yang menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik[14]

### 2.3. Tahapan Penelitian

Pada tahap penelitian ini, peneliti menggunakan model metode CRIS-DM (Cross Industry Standard Process for Data Mining) yang terdiri dari enam tahapan yaitu[11] :



Gambar 2. Model CRIS-DM

1. Pemahaman bisnis (*business understanding*)  
Merupakan tahap awal yaitu pemahaman penelitian, penentuan tujuan, dan rumusan masalah data mining. Pada tahap ini peneliti mengidentifikasi permasalahan yang ditemukan pada proses seleksi pemberkasan dan penetapan mahasiswa yang diterima dan ditolak untuk mendapatkan beasiswa bidikmisi yang ada di Universitas Hamzanwadi adalah masih menggunakan proses manual yakni dengan seleksi kelengkapan berkas dari para mahasiswa baru, yang dimana syarat – syarat berkas yang diajukan oleh mahasiswa baru tersebut sudah disepakati bersama oleh pihak kampus dan harus dimiliki oleh mahasiswa yang akan mengajukan sebagai tahap seleksi administrasi. Dan sebaliknya akan ditolak apabila berkasnya kurang lengkap atau tidak memenuhi persyaratan.
2. Pemahaman data (*data understanding*)

Dalam tahap ini dilakukan pengumpulan data, mengenali lebih lanjut data yang akan digunakan. Tahap ini merupakan proses tindak lanjut setelah proses pemahaman bisnis. Dimana peneliti menerapkan metode observasi, wawancara dan studi pustaka untuk mendapatkan data-data yang diperlukan. Data yang digunakan adalah data sekunder, yang didapat dari jumlah mahasiswa yang mengajukan beasiswa sebanyak 386 orang mahasiswa yang mengajukan beasiswa pada tahun 2021 dan ditetapkan lulus seleksi sebanyak 116 orang.

### 3. Pengolahan data atau persiapan (Data preparation)

Pada tahap ini akan ditentukan variabel apa saja yang ingin di analisis, melakukan perubahan pada beberapa variabel jika diperlukan sehingga data siap untuk dimodelkan. Jumlah data yang diperoleh pada penelitian ini sebanyak 496 record, baik mahasiswa yang lulus seleksi atau diterima maupun yang tidak lulus seleksi (ditolak), akan tetapi data tersebut masih mengandung duplikasi dan anomali atau inkonsisten data. Untuk mendapatkan data yang berkualitas, ada

### 4. Pemodelan (Modeling) .

Pada tahap ini peneliti menetapkan pemodelan dengan menggunakan metode klasifikasi, dengan algoritma naive bayes berbasis particle swarm optimization (PSO).

### 5. Evaluasi (evaluation)

Pada tahap ini dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi menggunakan confusion matrik yaitu Akurasi dan AUC atau grafik ROC

### 6. Penyebaran (Deployment)

Tahap ini merupakan tahap akhir, yaitu penerapan proses data mining sebagai solusi dalam memecahkan permasalahan dan sebagai bahan evaluasi dalam mengambil keputusan serta menerapkan suatu kebijakan

## 3. Metode Penelitian

### 3.1. Teknik pengumpulan data.

Dalam penelitian ini teknik pengumpulan data yang digunakan adalah teknik observasi, wawancara dan studi pustaka[3].

#### 1. Teknik Observasi

Teknik observasi merupakan teknik pengamatan secara langsung terhadap kegiatan yang sedang dilakukan. Teknik observasi yang dilakukan adalah mengumpulkan data dengan melakukan observasi atau pengamatan langsung ke lokasi penelitian yaitu Universitas Hamzanwadi untuk melakukan penelitian dan mencari informasi-informasi yang dibutuhkan dalam menyelesaikan penelitian ini.

#### 2. Teknik Wawancara

Melakukan komunikasi langsung dengan wakil rektor dan staff kemahasiswaan universitas

hamzanwadi dan ketua Tim PMB beserta staf nya. Teknik ini dilakukan untuk mendapatkan izin penelitian serta memperoleh data-data yang diperlukan dalam penelitian ini.

### 3. Studi pustaka

Melakukan metode pengumpulan data dengan cara mencari informasi dan referensi yang sesuai dengan topik penelitian melalui buku-buku, jurnal dan browsing-browsing melauai situs-situs internet.

### 2. Teknik Analisis Data

Penelitian ini menggunakan teknik analisis data sekunder, berupa data mahasiswa yang mengajukan beasiswa bidikmisi pada universitas hamzanwadi pada tahun anggaran 2021, baik yang diterima maupun di tolak seleksi yang kemudian akan digunakan sebagai instrumen guna memperoleh data dalam proses menentukan calon penerima beasiswa. Adapun atribut yang digunakan dalam penelitian ini adalah sebanyak tujuh (7) atribut, yang terdiri dari Status DTKS, prestasi (akademik / non akademik), pekerjaan orang Tua, jumlah tanggungan keluarga, jumlah penghasilan, dan kepemilikan Rumah. [15]

### 3. Teknik pengolahan Data

Perangkat lunak yang digunakan untuk mengolah datanya adalah Rapidminer dengan metode Klasifikasi algoritma Naive Bayes[16].

### 4. Model yang diusulkan

Model yang diusulkan pada penelitian ini dengan menggunakan metode klasifikasi dengan algoritma naive bayes. Tahap pengujian dilakukan dengan menggunakan k-fold cross validation. Sementara untuk tahap Evaluasi yang digunakan adalah Confussion Matrik dengan nilai akurasi dan nilai AUC yang ditunjukkan dalam grafik ROC[17]

### 4. Hasil dan Pembahasan

Pada penelitian ini telah dilakukan proses pengolahan data menggunakan tool Rapid Miner, dengan metode klasifikasi menggunakan algoritma Naive Bayes, untuk mengklasifikasi Mahasiswa yang berhak menerima Beasiswa Bidikmisi pada Universitas Hamzanwadi tahun anggaran 2021 berdasarkan kriteria yang telah ditentukan pihak Universitas. Diketahui jumlah mahasiswa yang mengajukan beasiswa pada tahun 2021 ini berjumlah 785 orang,, tetapi dari jumlah tersebut banyak yang tidak lulus pada proses administrasi dikarenakan berkas yang kurang lengkap. Adapun jumlah dataset yang digunakan dalam penelitian ini berjumlah 386 sampel dataset. Atribut yang digunakan dalam penelitian ini adalah 7 (tujuh) atribut, yang terdiri dari, yaitu : Nama, Status DTKS, Prestasi, Pekerjaan Orang Tua, Jumlah penghasilan, Jumlah Tanggungan Keluarga dan Kepemilikan Rumah.

Tabel 1. Sampel Dataset

No	Nama	Status DTKS	Prestasi	Pekerjaan org tua	Jmlh tanggungan	Penghasilan org tua	Kepemilikan rumah
1	Lailatun sariro	Belum terdata	Tidak ada	Buruh	2	500000	Tidak
2	M.Lutfi	Belum terdata	Tidak ada	Tidak ada	2	500000	Tidak
3	M. zainul hgasan	Terdata	Tidak ada	Pertani	2	500000	Tidak
4	Miya witrì S	Belum terdata	Tidak ada	Wiraswasta	3	600000	Tidak
5	Selma Azizah	Belum terdata	Tidak ada	Wiraswasta	2	750000	Tidak
6	Setiana W.	Terdata	Tidak ada	Petani	2	500000	Ya
7	Sindi A	Belum terdata	Tidak ada	Nelayan	3	500000	Tidak
8	Ajeng Pitaloka	Belum terdata	Tidak ada	Swast	2	650000	Tidak
9	Amrina Rosada A.	Terdata	Ada	Petani	4	700000	Tidak
10	Andar aditia	Belum terdata	Tidak ada	Buruh	2	700000	Tidak
11	Danil Azmi A.	Belum terdata	Tidak ada	Petani	2	750000	Tidak

Tabel 2. Atribut yang digunakan

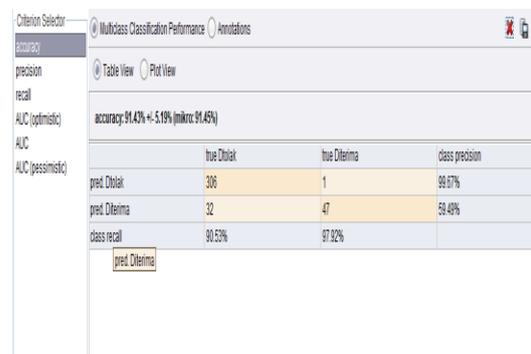
No	Atribut	Nilai
1	Nama	Id
2	Status DTKS	Bynominal
3	Pekerjaan org Tua	Polynomial
4	Prestasi	Bynominal
5	Jmlh Penghasilan org tua	Numerik
6	Jumlah tanggungan	Integer
7	Kepemilikan Rumah	Polynomial

Berikut Hasil pengolahan data yang sudah dilakukan peneliti dengan menggunakan software RapidMiner, Metode Klasifikasi, Algoritma Naive Bayes dengan melakukan k-fold 10 cross Validation.

Tabel 3. Hasil pengolahan data menggunakan Algoritma Naive Bayes

Cross validation	Hasil Pengolahan algoritma Naive Bayes			
	Accu (%)	Prec (%)	Recall (%)	AUC dan ROC (%)
k-2	90,93	65,92	85,42	0,890
k-3	90,42	61,31	95,83	0,995
k-4	91,43	63,80	97,92	0,996
k-5	88,87	57,91	98,00	0,996
k-6	89,11	55,82	97,92	0,996

Dapat dilihat dari percobaan yang dilakukan menggunakan 10 k-fold cross validation didapatkan hasil akurasi tertinggi pada k-4 yaitu sebesar 91,43 % yang berarti bahwa algoritma naive bayes ini sangat bagus dalam mengklasifikasi mahasiswa yang berhak mendapatkan beasiswa bidikmisi pada Universitas Hamzanwadi.



Gambar 4. Hasil akurasi tertinggi k-fold 4 True ditolak sebesar 306 , ternyata hasil klasifikasi algoritma benar yang berarti True Positif artinya bahwa mahasiswa tersebut faktanya ditolak dan hasil prediksi juga benar. Dan nilai true diterima sebesar 1 dimana hasil prediksi salah, sementara false negatif sebesar 32 artinya antara nilai faktual dan hasil klasifikasi sama-sama salah bahwa mahasiswa tersebut tidak berhak menerima beasiswa dan true negatif sebesar 47 artinya mahasiswa tersebut faktanya diterima tetapi hasil prediksinya salah atau tidak

menerima. Berikut gambar tabel dari confusion matrix tersebut:

Tabel 4. confusion matrix

	True Positif	True Negatif
False positif	306	1
False Negatif	32	47

Perhitungan tabel confusion matrix :

• **Accurac**

$$\begin{aligned}
 &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \% \\
 &= \frac{306 + 47}{306 + 47 + 32 + 1} \times 100 \% \\
 &= \frac{353}{386} \times 100 \% \\
 &= 91,45 \%
 \end{aligned}$$

• **Precision**

$$\begin{aligned}
 &= \frac{TP}{TP + FP} \times 100 \% \\
 &= \frac{306}{306 + 32} \times 100 \% \\
 &= \frac{306}{338} \times 100 \% \\
 &= 90,53 \%
 \end{aligned}$$

• **Recall**

$$\begin{aligned}
 &= \frac{TP}{TP + FN} \times 100 \% \\
 &= \frac{306}{306 + 1} \times 100 \% \\
 &= 99,67 \%
 \end{aligned}$$

Dari hasil perhitungan tabel confusion matrik tersebut, dapat dilihat nilai accurasi yang diperoleh adalah 91,45% yang artinya model tersebut berhasil dalam mengklasifikasi penerima beasiswa dan hasil precission sebesar 90,53% yang berarti sistem berhasil memberikan

informasi yang tepat, sementara Recall nya adalah sebesar 99,67% artinya sistem berhasil dalam menemukan informasi yang benar.



Gambar 5. Grafik ROC dan nilai AUC Algoritma Naive Bayes

Dari grafik ROC tersebut diketahui nilai AUC sebesar 0,996 % yang berarti bahwa klasifikasi tersebut merupakan diagnosa excellent classification[1].

**5. Kesimpulan**

Dari implementasi dan pengujian yang dilakukan peneliti dalam mengklasifikasi kelulusan pemberkasan mahasiswa universitas hamzanwadi yang mendaftarkan diri atau m engusulkan diri untuk mendapatkan beasiswa bidikmisi pada tahun 2021 menggunakan teknik data mining algoritma Naive Bayes, dengan 10 kali pengujian menggunakan cross validation yaitu k-fold 2 sampai k-fold 10, didapatkan nilai akurasi tertinggi sebesar 91,43% dari k-fold 4 validation. Sedangkan AUC didapatkan 0.996% dengan tingkat diagnosa excellent classification.

Dari hasil akurasi dan nilai AUC yang didapatkan tersebut, maka dapat disimpulkan bahwa algoritma Naive Bayes ini lebih akurat dan sangat bagus digunakan dalam menganalisis serta mengklasifikasi penerima beasiswa bidikmisi pada Universitas Hamzanwadi Selong. Serta bisa dijadikan sebagai bahan evaluasi dalam mengambil keputusan untuk proses seleksi beasiswa selanjutnya

## 6. Daftar Pustaka

- [1] P. A. Ani and A. Andri, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Mahasiswa Penerima Kip Pada Universitas Bina Darma," *Bina Darma Conf. ...*, pp. 172–180, 2022, [Online]. Available: <https://conference.binadarma.ac.id/index.php/BDCSS/article/view/3045%0Ahttps://conference.binadarma.ac.id/index.php/BDCSS/article/download/3045/1286>
- [2] Y. Farida and N. Ulinnuha, "Klasifikasi Mahasiswa Penerima Program Beasiswa Bidik Misi Menggunakan Naive Bayes," *Syst. Inf. Syst. Informatics J.*, vol. 4, no. 1, pp. 17–22, 2018, doi: 10.29080/systemic.v4i1.317.
- [3] D. Dahri, F. Agus, and D. M. Khairina, "Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 11, no. 2, p. 29, 2016, doi: 10.30872/jim.v11i2.211.
- [4] A. Rahman and A. Suryanto, "Implementasi Sistem Informasi Seleksi Penerima Beasiswa Dengan Metode Naive Bayes Classifier," *J. Penelit. Pendidik. Indones.*, vol. 2, no. 3, pp. 1–8, 2017, [Online]. Available: <http://www.irpp.com/index.php/jpp/article/view/807>
- [5] C. A. P. Dita, P. Chairunisyah, and ..., "Penerapan Naive Bayesian Classifier Dalam Penyeleksian Beasiswa PPA," *J. Comput. ...*, vol. 2, no. 2, pp. 194–198, 2021, [Online]. Available: <https://ejournal.seminar-id.com/index.php/josyc/article/view/649>
- [6] G. Suganda, M. Asfi, R. T. Subagio, and R. P. Kusuma, "Beasiswa Kartu Indonesia Pintar ( Kip ) Kuliah Menggunakan Naive," vol. 9, no. 2, pp. 193–199, 2022, doi: 10.30656/jsii.v9i2.4376.
- [7] S. Iskandar, N. R. Refisis, and B. A. Ginting, "Metode Naive Bayes Classifier Dalam Penentuan Penerima Beasiswa Bidikmisi Di Universitas Negeri Medan," *Karismatika*, vol. 7, no. 1, pp. 10–23, 2021.
- [8] Y. Yahya and H. Bahtiar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20–28, 2021, doi: 10.29408/jit.v4i1.2981.
- [9] I. Pangaribuan, "Penerapan Algoritma Naive Bayes Untuk Seleksi Beasiswa Pemprov Dan Bawaku Di Universitas Komputer Indonesia Application of Naive Bayes Algorithm for Selection of Providing and Bawaku Scholarship in University Computer Indonesia," 2019.
- [10] R. A. Saputra and S. Ayuningtias, "Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerima Beasiswa Pada Smk Pasim Plus Sukabumi," *Swabumi*, vol. IV, no. 2, pp. 114–120, 2016.
- [11] Y. T. U. Heni Sulistiani, "Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa," *Snti*, pp. 300–305, 2018.
- [12] M. Saiful and S. Samsuddin, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 29–38, 2021, doi: 10.29408/jit.v4i1.2982.
- [13] M. Artur, "Review the performance of the Bernoulli Naive Bayes Classifier in

- Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features," *Procedia Comput. Sci.*, vol. 190, no. 2019, pp. 564–570, 2021, doi: 10.1016/j.procs.2021.06.066.
- [14] A. Sudrajat and I. Budi, "Analisis Kinerja Algoritma Support Vector Machine (Svm) Pada Data Seleksi Penerima Beasiswa Menggunakan Particle Swarm Optimization (Pso) (Studi Kasus: Politeknik Tedc Bandung)," *J. TEDC*, vol. 13, no. 1, pp. 1–11, 2019, [Online]. Available: <http://ejournal.poltektedc.ac.id/index.php/tedc/article/view/15>
- [15] Baskoro, Sriyanto, and L. Setya Rini, "Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu," *Semin. Nas. Has. Penelit. dan Pengabd. Masy. Inst. Inform. dan Bisnis Darmajaya*, pp. 87–94, 2021.
- [16] M. Riyyan, H. Firdaus, J. H. Ronggo Waluyo, T. Timur, and J. Barat, "Perbandingan Algoritme Naïve Bayes Dan Knn Terhadap Data Penerimaan Beasiswa (Studi Kasus Lembaga Beasiswa Baznas Jabar)," *J. Inform. Rekayasa Elektron.*, vol. 5, no. 1, pp. 1–10, 2022, [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jireISN.2620-6900>
- [17] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500