JURNAL FASILKOM
Volume 15 No. 1 | April 2025: 108-114

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzkia

Muhammad Thoriq[1], Fajar Maulana[2], Yofhanda Septi Eirlangga[3], Nova Hayati[4], Muhammad Ashim Madani[5]
[125]Informatika, Universitas Adzkia
[34]Sistem Informasi, Universitas Adzkia
[1]thoriq.if@adzkia.ac.id*, [2]fajar@adzkia.ac.id, [3]yofhanda_se@adzkia.ac.id, [4]novahyt@adzkia.ac.id,
[5]ashimmadani1412@gmail.com

Abstract

This study aims to develop a predictive model for the acceptance of applicants for the Smart Indonesia Card (KIP) scholarship at Universitas Adzkia using the Naïve Bayes algorithm. The prediction process was conducted using RapidMiner software through the Knowledge Discovery in Database (KDD) approach, which includes five main stages: data selection, preprocessing, transformation, modeling, and evaluation. The dataset consists of 829 KIP scholarship applicants from the 2024 intake, comprising six key attributes: DTKS Status, P3KE Status, School District/City, School Province, Gender, and Applicant Status (Accepted/Not Accepted). The predictive model was tested using three data split scenarios: 70:30, 80:20, and 90:10. Evaluation results show that the 80:20 data split scenario produced the best performance, with an accuracy of 77.11%, precision of 66.67%, and recall of 5.13%. The low recall value is attributed to the small number of accepted applicants compared to the total number of applicants, as well as the high proportion of applicants who did not meet the eligibility criteria for the KIP scholarship. This class imbalance in the dataset affected the model's ability to detect the minority class (accepted). These findings suggest that the Naïve Bayes algorithm can serve as an initial approach for data-driven scholarship selection, although further development is required, including data balancing techniques and exploration of alternative algorithms. The results of this study can be utilized by the Universitas Adzkia admissions team to streamline and optimize the KIP scholarship selection process in an objective manner.

Keywords: naïve bayes, kartu indonesia pintar, classification, data mining, student selection.

Abstrak

Penelitian ini bertujuan untuk mengembangkan model prediksi penerimaan pendaftar beasiswa Kartu Indonesia Pintar (KIP) di Universitas Adzkia menggunakan algoritma Naïve Bayes. Proses prediksi dilakukan dengan memanfaatkan perangkat lunak RapidMiner melalui pendekatan *Knowledge Discovery in Database* (KDD) yang mencakup lima tahapan utama: seleksi data, pra-pemrosesan, transformasi, pemodelan, dan evaluasi. Dataset yang digunakan terdiri dari 829 data pendaftar beasiswa KIP tahun 2024, dengan enam atribut utama: Status DTKS, Status P3KE, Kabupaten/Kota Sekolah, Provinsi Sekolah, Jenis Kelamin, dan Status Pendaftar (Diterima/Tidak Diterima). Model prediksi diuji melalui tiga skenario pembagian data, yaitu rasio 70:30, 80:20, dan 90:10. Hasil evaluasi menunjukkan bahwa skenario pembagian data 80:20 memberikan performa terbaik dengan akurasi 77,11%, *precision* 66,67%, dan *recall* 5,13%. Nilai *recall* yang rendah disebabkan oleh jumlah pendaftar yang diterima beasiswa sangat sedikit dibandingkan dengan total pendaftar, serta banyaknya peserta yang tidak memenuhi syarat penerimaan sesuai kriteria KIP Kuliah. Ketidakseimbangan kelas dalam dataset ini berdampak pada kemampuan model dalam mengenali kelas *minoritas* (diterima). Temuan ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan sebagai pendekatan awal untuk seleksi beasiswa berbasis data, meskipun diperlukan pengembangan lebih lanjut, baik dari segi teknik penyeimbangan data maupun eksplorasi algoritma lain. Hasil penelitian ini dapat dimanfaatkan oleh Tim PMB Universitas Adzkia untuk mempercepat dan mengefisienkan proses seleksi penerima beasiswa KIP secara objektif.

Kata kunci: naïve bayes, kartu indonesia pintar, klasifikasi, data mining, seleksi mahasiswa

1. Pendahuluan

Pendidikan tinggi merupakan aspek krusial dalam peningkatan sumber daya manusia, dan di Indonesia, Program Kartu Indonesia Pintar (KIP) adalah inisiatif utama pemerintah untuk mendukung akses pendidikan tinggi bagi mahasiswa dari keluarga kurang mampu, meskipun seleksi penerima beasiswa ini masih mengandalkan metode manual yang kurang efisien dan rentan terhadap subjektivitas. Pemerintah Indonesia telah meluncurkan berbagai inisiatif guna mendorong peningkatan partisipasi pendidikan di seluruh jenjang, salah satunya melalui program Kartu Indonesia Pintar (KIP) yang ditujukan untuk memberikan dukungan pendidikan bagi siswa yang berasal dari keluarga kurang mampu, sehingga mereka dapat melanjutkan pendidikan hingga perguruan tinggi tanpa menghadapi kendala biaya yang tinggi.[1]

Sebagai kebijakan strategis, Program Kartu Indonesia Pintar (KIP) berperan penting dalam memperluas akses dan partisipasi masyarakat terhadap pendidikan di

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Indonesia. Program ini memberikan dukungan berupa biaya pendidikan dan tunjangan hidup bagi siswa berprestasi yang berasal dari latar belakang ekonomi kurang mampu. Di tingkat perguruan tinggi, KIP diberikan kepada mahasiswa yang memenuhi syarat-syarat tertentu, seperti kriteria ekonomi dan prestasi akademik. Namun, dengan banyaknya pendaftar yang memenuhi kriteria, proses seleksi penerima KIP di perguruan tinggi menjadi semakin kompleks dan menantang.[2][3]

Universitas Adzkia, sebagai salah satu perguruan tinggi di Indonesia, turut ambil bagian dalam pelaksanaan program KIP. Setiap tahun, universitas ini menerima sejumlah besar pendaftar untuk program KIP, sementara kuota yang tersedia terbatas. Oleh karena itu, diperlukan mekanisme seleksi yang efektif dan efisien untuk memastikan bahwa bantuan KIP diberikan kepada mahasiswa yang benar-benar memenuhi kriteria dan paling membutuhkan. Seleksi ini harus mempertimbangkan berbagai faktor, seperti latar belakang ekonomi, prestasi akademik, dan potensi keberhasilan studi mahasiswa.

Seiring dengan perkembangan teknologi informasi dan analisis data, metode-metode berbasis data mining mulai diterapkan dalam berbagai bidang, termasuk di sektor pendidikan. Proses data mining bertujuan menggali informasi dari data berskala besar dan rumit untuk mengidentifikasi pola yang relevan dalam mendukung keputusan. Dalam konteks seleksi penerima KIP, penggunaan metode data mining dapat membantu universitas dalam membuat prediksi yang lebih akurat tentang siapa yang layak menerima bantuan, sehingga proses seleksi menjadi lebih adil, transparan, dan tepat sasaran.[4][5]

Naïve Bayes adalah salah satu algoritma yang dapat digunakan dalam proses prediksi. Meskipun tergolong sederhana, algoritma ini merupakan metode klasifikasi dalam machine learning yang terbukti efektif dalam berbagai skenario, terutama untuk tugas prediksi dan klasifikasi data. Naïve Bayes bekerja dengan prinsip teorema Bayes yang menghitung probabilitas terjadinya suatu peristiwa berdasarkan kondisi sebelumnya. Algoritma ini unggul dalam memproses volume data yang besar secara cepat, menjadikannya efektif untuk analisis dalam skala besar, serta kemampuannya dalam mengelola atribut yang bersifat independen satu sama lain.[6][7]

Sebagai bagian dari upaya meningkatkan pemahaman dalam bidang literatur, penulis telah melakukan kajian dan analisis terhadap berbagai penelitian sebelumnya. Salah satu penelitian yang dikaji adalah karya Gagan Suganda dan rekan-rekannya. Penelitian tersebut menerapkan algoritma Naïve Bayes untuk mengklasifikasi data penerima bantuan KIP Kuliah berdasarkan sejumlah kriteria, seperti penghasilan orang tua, luas tanah, dan prestasi akademik. Dengan memanfaatkan dataset yang berisi 100 data, penelitian ini mencapai akurasi sistem sebesar 88,21%. [4] Selain itu, penelitian lain yang relevan adalah karya Alvina Felicia Watratan dan timnya yang dilakukan pada tahun 2020. Penelitian ini menggunakan dataset yang terdiri dari 33 data guna memperkirakan laju penyebaran COVID-19 di wilayah Indonesia. Meskipun hasil akurasi yang dicapai hanya sebesar 48,48%, penelitian ini menunjukkan potensi algoritma Naïve Bayes dalam penerapan analisis epidemiologi.[11]

Penelitian lainnya yang dilakukan oleh Dedi Darwis, dkk pada tahun 2023. Studi ini dilakukan untuk menelaah tanggapan masyarakat mengenai pelayanan BMKG melalui pendekatan algoritma Naïve Bayes. Hasilnya menunjukkan tingkat akurasi sebesar 69,97%, yang membuktikan efektivitas metode ini dalam klasifikasi sentimen.[12] Penelitian serupa oleh Muhammad Romadloni Putra, dkk pada tahun 2024, menggunakan fitur MFCC untuk mengenali suara burung. Penelitian ini mencapai akurasi 90%, menunjukkan potensi algoritma Naïve Bayes dalam konservasi burung melalui pengenalan suara.[13]

Penelitian lain yang relevan dilakukan oleh Widdi Djatmiko dan rekan-rekannya pada tahun 2023. Berdasarkan hasil penelitian, Random Forest terbukti lebih unggul dalam hal performa, khususnya dari segi akurasi mencapai 97,52%, dibandingkan dengan algoritma Naïve Bayes.[14] Penelitian lainnya adalah studi berjudul yang dilakukan oleh Hidayatunnisa'i dan rekan-rekannya pada tahun 2023. Penelitian ini menemukan bahwa metode Support Vector Machine menunjukkan kinerja yang lebih baik dengan akurasi mencapai 97%, dibandingkan dengan Naïve Bayes yang memperoleh akurasi 95%. Kedua studi tersebut menyoroti peran algoritma dalam meningkatkan efisiensi dan akurasi prediksi serta analisis data.[15]

Penelitian lain oleh Rizky Aziz, dkk pada tahun 2024, menggunakan algoritma Multinomial Naïve Bayes untuk menganalisis ulasan pengguna aplikasi OYO. Dengan akurasi mencapai 87%, hasil penelitian mengonfirmasi bahwa Naïve Bayes mampu memberikan performa yang baik dalam penerapan analisis sentimen.[16] Sementara itu, penelitian oleh Marsani Asfi, dkk pada tahun 2020 menunjukkan hasil akurasi 90% dalam pemilihan dosen pembimbing berdasarkan data historis dan kriteria kompetensi.[17]

Penelitian terakhir mencakup studi tentang kesehatan, oleh Amellia Veronica Agustin, dkk pada tahun 2023. Penelitian ini menggunakan dataset dari Kaggle dan mencapai akurasi 78,50% dalam diagnosis diabetes.[6] Selain itu, penelitian oleh Deny Novianti pada tahun 2019, memprediksi penyakit hepatitis dengan akurasi sebesar 76,77%. Kedua penelitian ini menunjukkan potensi besar algoritma Naïve Bayes dalam membantu diagnosis dini dan pengelolaan data kesehatan.[18]

Dari beberapa uraian di atas, penulis tertarik mengajukan penelitian menggunakan algoritma Naïve Bayes pada Pendaftar Beasiswa KIP di Universitas Adzkia sebagai objek penelitian. Hasil penelitian ini diharapkan memberikan gambaran kepada Tim PMB

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Universitas Adzkia untuk mengetahui pola Pendaftar Beasiswa KIP sehingga dapat membangun pengetahuan dan memberikan penilaian akurasi performa model algortima Naïve Bayes tersebut.

2. Metode Penelitian

2.1 Naïve Bayes

Algoritma Naive Bayes merupakan sebuah proses klasifikasi dengan metode pemeriksaan pada tingkat kemunculan data yang sering ditemukan dalam data pelatihan yang menghitung sekumpulan kemungkinan untuk mencari peluang terbesar dari probalitias klasifikasi. Algoritma Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan data yang besar. Model persamaan Naive Bayes yaitu sebagai berikut:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (1)$$

Keterangan:

Y : Data dengan kelas yang belum diketahui

X : Hipotesis data y merupakan suatu kelas spesifik

P(x|y) :Probabilitas hipotesis x berdasarkan kondisi y (posteriori probability)

P(x) : Probabilitas hipotesis x (*prior probability*)

P(y|x) : Probabilitas y berdasarkan kondisi pada hipotesis x

p(y) : Probabilitas dari y

2.2 *Dataset*

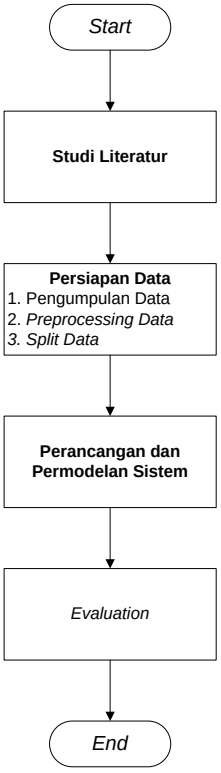
Dataset yang digunakan pada penelitian ini bersumber dari tim PMB Universitas Adzkia tahun 2024 sesuai dengan topik yang telah ditentukan yaitu pendaftar beasiswa KIP Universitas Adzkia tahun 2024. Selain melakukan pengambilan *dataset* pendaftar beasiswa KIP Universitas Adzkia, penulis melakukan metode wawancara yang digunakan untuk memperoleh keterangan, informasi dan penjelasan lebih mendalam terkait dengan permasalahan yang akan dibahas dalam penelitian ini sehingga diperoleh data yang akurat dan terpecaya yang bersumber dari tim PMB Universitas Adzkia tahun 2024. Jumlah dataset yang terkumpul yaitu sebanyak 829 Data yang terdiri dari 6 atribut.

Tabel 1. Dataset Calon Mahasiswa KIP

No	Atribut	Deskripsi
1	Status DTKS	Status Data Terpadu Kesejahteraan Sosial pendaftar KIP
2	Status P3KE	Status Data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem Pendaftar KIP
3	Kab/Kota Sekolah	Kabupatan/Kota Sekolah Pendaftar KIP
4	Provinsi Sekolah	Provinsi Sekolah Pendaftar KIP
5	Jenis Kelamin	Jenis Kelamin Pendaftar KIP
6	Status Pendaftar	Status hasil Pendaftar KIP

Data yang telah berhasil dikumpulkan selanjutnya akan ditentukan kelas atau label dari dataset Pendaftar Beasiswa KIP yaitu atribut atau variabel yang menjelaskan status Pendaftar. Label status Pendaftar akan memberikan informasi apakah Pendaftar diterima akan dilabeli dengan atribut "Diterima" dan Peserta yang tidak diterima akan dilabeli dengan atribut "Tidak Diterima". Dataset pendaftar beasiswa KIP yang digunakan berdasarkan atribut yang tersedia akan menjadi dasar faktor penentu hasil dari pendaftar beasiswa KIP Universitas Adzkia.

2.3 Alur Penelitian



Gambar 1. Alur Penelitian

1. Tahap awal akan dilakukan proses penelusuran informasi secara detil dan mendalam melalui studi literatur dengan upaya pencarian referensi dari berbagai sumber antara lain jurnal, buku ilmiah, tesis yang relevan berdasarkan permasalahan yang akan diteliti sebagai bahan rujukan dalam menentukan metode yang tepat dan sesuai untuk penyelesaian permasalahan.
2. Persiapan data akan dilakukan proses tahapan aktifitas yang akan memudahkan penulis dalam melakukan penilitian.

a. Proses pengumpulan dataset yang bersumber dari tim PMB Universitas Adzkia tahun 2024 sesuai dengan topik yang telah ditentukan yaitu Pendaftar Beasiswa KIP Universitas Adzkia. Data yang berhasil terkumpul yaitu sebanyak 829 dataset. Data Pendaftar yang berhasil terkumpul akan dibangun model untuk

P-ISSN : 2089-3353
E-ISSN : 2808-9162

mengklasifikasikan Pendaftar KIP menjadi 2 class yaitu Pendaftar Diterima atau Tidak Diterima.

b. Tahap *preprocessing* data untuk dilakukan pemeriksaan dan perbaikan yang berpotensi tidak sesuai atau tidak dibutuhkan (*data cleansing*) pada data yang akan digunakana pada penelitian.

c. Pembagian data akan dilakukan menjadi beberapa eskperimen untuk mengetahui tingkat performa terbaik antara lain dengan rasio 70%data testing dan 30% data *training*, 80% data *testing* dan 20% data *training*, 90% data *testing* dan 10% data *training*.

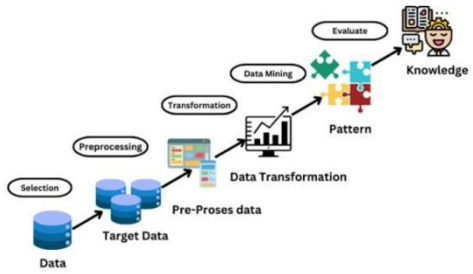
3. Pada tahap perancangan dan permodelan sistem, penulis akan melakukan aktifitas klasifikasi terhadap dataset yang telah dikumpulkan dengan melakukan pemodelan menggunakan algoritma Naive Bayes.

4. Pada tahap evaluasi akan dilakukan proses pengujian kinerja pada setiap algoritma terhadap *dataset* dan *split* data yang telah dilakukan eksperimen dengan nilai luaran berupa *confusion matrix* antara lain *accuracy*, *precission*, dan *recall* sebagai perbandingan tingkat performa untuk kedua algoritma.

5. Pada tahap akhir akan dilakukan diskusi dan analisa perbandingan Naïve bayes menggunakan *dataset* dan *split* data yang telah dilakukan. Pada proses ini akan disampaikan rasio dataset yang memiliki tingkat performa terbaik prediksi pola perilaku Pendaftar Beasiswa KIP Universitas Adzkia.

3. Hasil dan Pembahasan

KDD merupakan proses komputasi yang melibatkan perhitungan matematis untuk mengekstraksi informasi serta melakukan estimasi probabilistik berdasarkan kemungkinan kejadian di masa mendatang[19] Tahapan dalam metode KDD meliputi pemilihan data (*data selection*), pra-pemrosesan data (*data preprocessing)*, transformasi data (*data transformation*), penambangan data (*data mining*), dan evaluasi hasil (*evaluation*).



Gambar 2. *Knowledge Discovery in Database Process* (KDD)

3.1 Data *selection*

Dataset yang dibutuhkan dalam penilitian ini yaitu sekumpulan data yang diperoleh dari Tim PMB Universitas Adzkia tahun 2024 yang memuat informasi berupa riwayat catatan pendaftar yang terdaftar dalam *web* SPMB Universitas Adzkia. Terdapat 6 atribut dalam dataset yang akan digunakan dalam penelitian ini. Penggunaan atribut yang relevan dalam penelitian akan bermanfaat dan memberikan kontribusi dalam analisis data. Seluruh atribut memiliki hubungan keterkaitan yang akan menentukan kondisi perilaku pendaftar Beasiswa KIP dalam mendaftar di *web* SPMB Universitas Adzkia.

3.2 *Preprocessing*

Pada tahapan *Preprocessing* akan dilakukan proses analisa jumlah data yang tidak tersedia pada setiap atribut sehingga dapat dilakukan penanganan data yang hilang (*missing values*). Data yang hilang (*missing values*) akan mengakibatkan dapat tidak dapat diproses karena tidak memiliki nilai yang dapat dilakukan proeses perhitungan. Penanganan terhadap data yang kosong dapat dilakukan dengan menghapus baris tidak relevan dan data yang dianggap anomali sehingga menyisakan data yang valid untuk dijadikan sebagai data penentu yang akan digunakan dalam perhitungan. Hasil proses *missing values* sebagaimana tercantum pada tabel 2.

Tabel 2. Dataset Calon Mahasiswa KIP

No	Atribut	*Missing Vaue*
1	Status DTKS	0%
2	Status P3KE	0%
3	Kab/Kota Sekolah	0%
4	Provinsi Sekolah	0%
5	Jenis Kelamin	0%
6	Status Pendaftar	0%

3.3 *Transformation*

Selanjutnya dilakukan proses transformasi data untuk mengubah data menjadi format yang sesuai untuk analisis. Tranformasi data bertujuan untuk menjadikan data dapat lebih berkualitas dan memudahkan dalam proses perhitungan. Setelah seluruh tahapan tranformation telah dilakukan, maka akan didapatkan data yang layak untuk digunakan sebagai varibel dalam penelitian ini. Contoh dataset yang telah berhasil dilakukan dikumpulkan dan tranformasi data yaitu sebagaimana yang tercantum dalam tabel 3.

Tabel 3. Pembagian data *traning* dan data *testing*

No	Atribut	P1	P2	…	P829
1	Status DTKS	Terdata	Belum Terdata	…	Belum Terdata
2	Status P3KE	Belum Terdata	Terdata: Desil 4	…	Terdata: Desil 4
3	Kab/Kota Sekolah	Kab. Pesisir Selatan	Kab. Pasaman	…	Kab. Lima Puluh Koto
4	Provinsi Sekolah	Sumatera Barat	Sumatera Barat	…	Sumatera Barat

P-ISSN : 2089-3353
E-ISSN : 2808-9162

5	Jenis Kelamin	P	P	…	P
6	Diterima/Tidak Diterima	Tidak Diterima	Diterima	…	Tidak Diterima

Berdasarkan data pada tabel 3 di atas, diketahui terdapat data sebanyak 829 yang berhasil dikumpulkan dari tim PMB Universitas Adzkia. Dataset tersebut mencakup 6 atribut yang terdiri dari data Pendaftar yang terdaftar dalam web SPMB Universitas Adzkia yang memuat Status DTKS, Status P3KE, Kabupaten/Kota Sekolah pendaftar, Jenis Kelamin, dan Status Penerimaan (Diterima/Tidak Diterima) akan ditentukan sebagai class pada proses prediksi pendaftar beasiswa KIP Universitas Adzkia. Dalam melakukan proses klasifikasi, dibutuhkan informasi yang cukup memadai dan mendukung dalam konteks pembuatan model pembelajaran mesin. Pemilihan atribut di atas merupakan informasi data Pendaftar yang dapat bermanfaat untuk membantu memahami interaksi dan variasi antara atribut dalam menentukan label dari dataset dengan jumlah besar yang kompleks dan beragam.

3.4 Data *Mining* (Tahap Pemodelan)

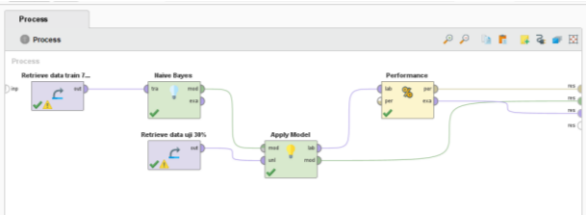
Pada penelitian ini akan dilakukan proses *split* data menjadi 2 bagian yaitu data training dan data testing. Variasi pembagian data akan dilakukan menjadi beberapa bagian antara lain dengan rasio 70% data *training* dan 30% data *testing*, 80% data *training* dan 20% data *testing*, 90% data *training* dan 10% data *testing*. Variasi pembagian data ini merupakan upaya eksperimen untuk mengetahui teknik *split* data yang memiliki perfoma terbaik. Jumlah komposisi pembagian data sebagaimana tercantum pada tabel 4 di bawah ini.

Tabel 4. Pembagian data *traning* dan data *testing*

No	% *Training*	% *Testing*	Jumlah Data *Training*	Jumlah Data *Testing*
1	70%	30%	579	250
2	80%	20%	663	166
3	90%	10%	745	84

Setelah ditentukan data yang telah dilakukan pembagian (*split data*), tahapan selanjutnya yaitu masuk dalam tahap proses pemodelan data dengan menggunakan algoritma Naive Bayes. Pemodelan data menggunakan tools *Rapid Miner* untuk melakukan proses seluruh rangkaian dalam penelitian ini. *Rapid Miner* merupakan alat bantu (*tools*) dalam melakukan proses analisa untuk data *mining*, *machine learning*, dan *predictive analytics*. Pengguna *Rapid Miner* dapat memanfaatkan banyak fitur *machine learning* dengan metode data *mining* yang cukup variatif.

Proses pemodelan *mining* menggunakan Naive Bayes yaitu melalui tahapan *entry dataset melalui tools Rapid Miner*. Dataset tersebut meliputi seluruh atribut yang memiliki hubungan keterkaitan dengan label yang telah ditentukan. Pada penilitian ini label yang akan ditentukan yaitu pada atribut Diterima/tidak diterima. Label tersebut akan menjadi hasil prediksi yang akan menentukan apakah Pendaftar yang terdaftar dalam web SPMB Universitas Adzkia diterima atau tidak diterima beasiswa KIP Universitas Adzkia. Pemodelan data pada algoritma Naive Bayes dapat dilihat pada gambar 3 dibawah ini.



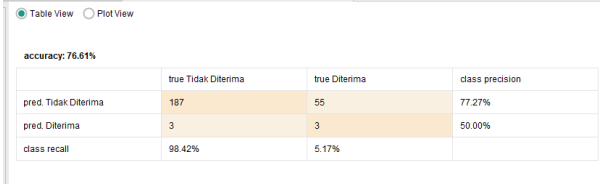
Gambar 3. Permodelan *Naïve Bayes*

3.5 Evaluation

Hasil pengujian performa terhadap algoritma Naïve Bayes menggunakan pendekatan *confusion matrix*. *Confusion matrix* umumnya banyak digunakan untuk melakukan menguji performa suatu model dengan cara membandingkan antara data aktual dengan data prediksi. Pengukuran kinerja suatu model menggunakan *Confusion Matrix* dapat dimanfaatkan untuk menentukan kinerja nilai pemodelan algoritma klasifikasi berdasarkan hasil perhitungan nilai akurasi yang diperoleh dari jumlah prediksi kasus yang benar dan jumlah prediksi yang salah. Pada penelitian ini akan dilakukan skenario pembagian data (*split data*) sebanyak 3 variasi data training (70%, 80%, 90%) untuk mengetahui performa terbaik pada setiap model yang akan dilakukan pengujian.

3.5.1 Naïve Bayes *Split Data* 70:30

Pemodelan data menggunakan Naïve Bayes pada split data 70:30 menghasilkan Confusion Matrix yang terbagi menjadi 3 pengujian (*Accuracy, Precision, Recall*) seperti terlihat pada gambar 4 di bawah ini.



Gambar 4. Nilai *Confussion Matrix Split Data* 70:30

Berdasarkan hasil pengujian yang tercantum pada gambar 4, performa model Naïve Bayes pada split data 70:30 menghasilkan nilai *Accuracy* = 76.61%, *Precision* = 50%, *Recall* = 5.17%.

3.5.2 Naïve Bayes *Split Data* 80:20

Pemodelan data menggunakan Naïve Bayes pada split data 80:20 menghasilkan Confusion Matrix yang terbagi menjadi 3 pengujian (*Accuracy, Precision, Recall*) seperti terlihat pada gambar 5 di bawah ini.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

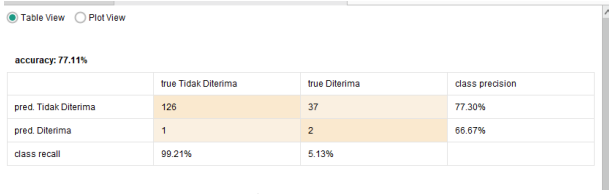


Table View Plot View

accuracy: 77.11%

	true Tidak Diterima	true Diterima	class precision
pred. Tidak Diterima	126	37	77.30%
pred. Diterima	1	2	66.67%
class recall	99.21%	5.13%	

Gambar 5. Nilai *Confussion Matrix Split Data* 80:20

Berdasarkan hasil pengujian yang tercantum pada gambar 5, performa model Naïve Bayes pada split data 80:20 menghasilkan nilai *Accuracy* = 77.11%, *Precision* = 66.67%, *Recall* = 5.13%.

3.5.3 Naïve Bayes *Split Data* 90:10

Pemodelan data menggunakan Naïve Bayes pada split data 90:10 menghasilkan Confusion Matrix yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 6 di bawah ini.

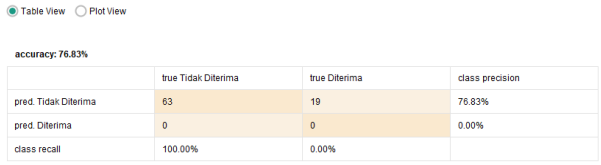


Table View Plot View

accuracy: 76.83%

	true Tidak Diterima	true Diterima	class precision
pred. Tidak Diterima	63	19	76.83%
pred. Diterima	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 6. Nilai *Confussion Matrix Split Data* 70:30

Berdasarkan hasil pengujian yang tercantum pada gambar 6, performa model Naïve Bayes pada split data 90:10 menghasilkan nilai *Accuracy* = 76.83%, *Precision* = 0%, *Recall* = 0%.

Berikut hasil rekap keseluruhan pengujian performa confusion matrix dari model Naive Bayes dan Random Forest menggunakan Rapid Miner sebagaimana tercantum pada tabel 5 di bawah ini.

Tabel 5. Pembagian data *traning* dan data *testing*

Model	*Split Data*	*Accuracy*	*Precision*	*Recall*
Naïve Bayes	70:30	76.61%	50%	5.17%
Naïve Bayes	80:20	77.11%	66.67%	5.13%
Naïve Bayes	90:10	76.83%	0%	0%

Berdasarkan proses pengujian menggunakan *confusion matrix*, performa Naïve Bayes pada *split data* 80:20 memiliki keunggulan dalam pemodelan data dibandingkan dengan rasio 70:30 dan 90:10.

4. Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi penerimaan pendaftar beasiswa KIP di Universitas Adzkia menggunakan algoritma Naïve Bayes. Dengan menerapkan pendekatan *Knowledge Discovery in Database* (KDD), model dibangun berdasarkan enam atribut utama dari 829 data pendaftar tahun 2024. Proses klasifikasi dilakukan melalui tiga skenario pembagian data, yakni rasio 70:30, 80:20, dan 90:10, untuk mengevaluasi performa model pada variasi proporsi data pelatihan dan pengujian.

Dari ketiga skenario yang diuji, pembagian data 80:20 memberikan hasil terbaik dengan nilai akurasi sebesar 77,11%, precision 66,67%, dan recall 5,13%. Hasil ini menunjukkan bahwa model cukup baik dalam memprediksi mayoritas data, namun belum optimal dalam mengidentifikasi kelas minoritas, yaitu pendaftar yang diterima beasiswa. Hal ini menunjukkan bahwa algoritma Naïve Bayes memiliki potensi sebagai alat bantu seleksi awal dalam proses penerimaan beasiswa berbasis *data mining*.

Nilai recall yang rendah disebabkan oleh ketidakseimbangan kelas dalam dataset, di mana jumlah pendaftar yang diterima sangat sedikit dibandingkan dengan total pendaftar. Hal ini diperparah oleh banyaknya pendaftar yang tidak memenuhi kriteria kelayakan sesuai dengan persyaratan KIP Kuliah. Ketidakseimbangan ini menghambat kemampuan model dalam mengenali pola pada kelompok pendaftar yang diterima, sehingga diperlukan upaya untuk menyeimbangkan data guna meningkatkan performa model, khususnya pada aspek *recall*.

Ke depan, penelitian ini dapat dikembangkan lebih lanjut dengan menerapkan teknik penyeimbangan data seperti SMOTE, menambahkan atribut yang lebih relevan, serta membandingkan performa dengan algoritma klasifikasi lainnya seperti *Decision Tree*, *Random Forest*, atau *Support Vector Machine*. Dengan demikian, model prediksi penerimaan beasiswa dapat menjadi lebih akurat dan adil, serta dapat diadopsi sebagai sistem pendukung keputusan oleh tim seleksi PMB Universitas Adzkia dalam proses seleksi beasiswa KIP.

Ucapan Terimakasih

Peneliti menyampaikan apresiasi yang sebesar-besarnya kepada Universitas Adzkia atas dukungan yang telah diberikan, sehingga penelitian ini dapat terlaksana dan menghasilkan temuan yang selaras dengan tujuan yang direncanakan.

Daftar Rujukan

[1] F. Nuraeni, D. Kurniadi, and G. Fauzian Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma K-Means++," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 437–443, 2023.
[2] A. Amin, R. N. Sasongko, and A. Yuneti, "Kebijakan Kartu Indonesia Pintar untuk Memerdekakan Mahasiswa Kurang Mampu," *J. Adm. Educ. Manag.*, vol. 5, no. 1, pp. 98–107, 2022.
[3] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering," *J. Focus Action Res. Math. (Factor M)*, vol. 5, no. 1, pp. 127–141, 2022.
[4] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, 2022.
[5] Muhammad Thoriq, A. E. Syaputra, and Y. S. Eirlangga, "Prediksi Peningkatan Kunjungan Pasien Dimasa Mendatang Mengunakan Jaringan Saraf Tiruan Backpropagation," *J. Fasilkom*, vol. 14, no. 1, pp. 34–40, 2024.
[6] A. Veronica Agustin and A. Voutama, "Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Menggunakan Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1002–1007, 2023.
[7] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 2, 2021.
[8] R. I. Borman and M. Wati, "Penerapan Data Maining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *J. Ilm. Fak. Ilmu Komput.*, vol. 09, no. 01, pp. 25–34, 2020.
[9] S. Rokhanah, A. Hermawan, and D. Avianto, "Pengaruh Principal Component Analysis Pada Naïve Bayes dan K-Nearest Neighbor Untuk Prediksi Dini Diabetes Melitus Menggunakan Rapidminer," *EVOLUSI J. Sains dan Manaj.*, vol. 11, no. 1, 2023.
[10] Y. S. Eirlangga, A. E. Syaputra, M. Thoriq, and U. Adzkia, "SPK Penyeleksian Siswa Kelas Unggul Dengan Metode Analytical Hierarchy Process (AHP)," vol. 14, no. 1, pp. 256–262, 2024.
[11] Rayuwati, Husna Gemasih, and Irma Nizar, "IMPLEMENTASI AlGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022.
[12] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021.
[13] M. R. Putra, F. Nurdiyansyah, and A. Y. Rahman, "Klasifikasi Jenis Burung Cucak Berdasarkan Suara Menggunakan MFCC Dan Naive Bayes," vol. 14, no. 2, pp. 463–470, 2024.
[14] W. Djatmiko, Kusrini, and Hanafi, "Perbandingan Naive Bayes dan Random Forest untuk Prediksi Perilaku Peserta Program Rujuk Balik," *J. Fasilkom*, vol. 13, no. 3, pp. 358–367, 2023.
[15] Hidayatunnisa, Kusrini, and Kusnawi, "Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine dalam Analisis Soal," *J. Fasilkom*, vol. 13, no. 2, pp. 173–180, 2023.
[16] R. Aziz, Tresna Maulana Fahrudin, and Wahyu Syaifullah Jauharis Saputra, "Analisis Sentimen Kepuasan Pengguna OYO DiPlaystore Dengan Multinoial Naive Bayes dan Chi-square," *J. Fasilkom*, vol. 14, no. 1, pp. 166–175, 2024.
[17] M. Asfi and N. Fitrianingsih, "Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi," *J. Nas. Inform. dan Teknol. Jar.*, vol. 5, pp. 45–50, 2020.
[18] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019.
[19] A. Nur Kirana, B. Nurhakim, S. Eka Permana, W. Prihartono, and G. Dwilestari, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1637–1642, 2024.