

PENERAPAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI HEREGISTRASI CALON MAHASISWA BARU

Sarwido¹, Gentur Wahyu Nyipto Wibowo², Mohammad Abdul Manan³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Nahdlatul Ulama Jl. Taman Siswa(Pekeng) Tahunan, Jepara, Jawa Tengah

e-mail: ¹sarwido@gmail.com, ²gentur23@yahoo.com, ³mananbungu@gmail.com

Informasi Artikel

Diterima: 05-01-2022

Direvisi: 28-01-2022

Disetujui: 02-02-2022

Abstrak

Penerimaan mahasiswa baru merupakan aktivitas penting perguruan tinggi untuk memperoleh mahasiswa baru. Dalam penerimaan calon mahasiswa baru sering terjadi tidak dilakukannya heregistrasi oleh pendaftar. Hal ini juga terjadi dalam penerimaan mahasiswa baru di UNISNU Jepara. Universitas belum memiliki cara untuk mengetahui kemungkinan calon mahasiswa baru cenderung akan heregistrasi atau tidak. Berdasarkan permasalahan tersebut, maka diterapkan algoritma Naive Bayes untuk melakukan prediksi apakah calon mahasiswa baru cenderung akan heregistrasi atau tidak. Dataset yang digunakan pada penelitian ini diambil dari data PMB tahun 2019-2020. Dataset awal yang diperoleh sebanyak 3.969 record dengan 18 (delapan belas) atribut. Kemudian dilakukan pre-processing data sehingga dataset yang akan digunakan menjadi 2.853 record dengan 14 (empat belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 12 (dua belas) atribut reguler yaitu tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi. Dataset tersebut diolah menggunakan algoritma Naive Bayes serta dilakukan pengujian menggunakan confusion matrix dan kurva ROC (Receiver Operating Characteristic) menggunakan tools RapidMiner. Diperoleh nilai akurasi sebesar 92,67% dan nilai AUC (Area Under Curve) sebesar 0,841 yang dikategorikan sebagai klasifikasi yang baik.

Kata Kunci: Algoritma Naive Bayes, Heregistrasi, UNISNU Jepara.

Abstract

Admission of new students is an important activity for universities to obtain new students. In the admission of new students, it is often the case that registration is not carried out by the registrant. This also happened in the admission of new students at UNISNU Jepara. Universities don't yet have a way of knowing whether a prospective new student is likely to enroll or not. Based on these problems, the Naive Bayes algorithm is applied to predict whether prospective new students are likely to register or not. The dataset used in this study was taken from PMB data for 2019-2020. The initial dataset obtained was 3,969 records with 18 (eighteen) attributes. Then pre-processing the data is carried out so that the dataset to be used becomes 2,853 records with 14 (fourteen) attributes with details of 1 (one) ID attribute, namely name, 12 (twelve) regular attributes, namely year of registration, class program, gender, age, study program, city of origin, father's occupation, mother's occupation, parent's income, home school major, National Examination scores, registration information, and 1 (one) class attribute, namely the status of hereditary. The dataset is processed using the Naive Bayes algorithm and tested using a confusion matrix and ROC (Receiver Operating Characteristic) curve using RapidMiner tools. Obtained an accuracy value of 92.67% and an AUC (Area Under Curve) value of 0.841 which is categorized as a good classification.

Keywords: Naive Bayes Algorithm, Hereregistration, UNISNU Jepara.

1. Pendahuluan

Universitas Islam Nahdlatul Ulama (UNISNU) Jepara menjadi bagian dari persaingan untuk

mendapatkan calon mahasiswa baru pada proses Penerimaan Mahasiswa Baru (PMB) setiap tahunnya, setiap penerimaan calon mahasiswa baru



sering terjadi tidak dilakukannya heregistrasi oleh calon mahasiswa sudah mendaftar. Jika kemungkinan calon mahasiswa baru yang tidak melakukan heregistrasi dapat diketahui lebih awal maka pihak pengelola dapat melakukan tindakan untuk mempertahankan calon mahasiswa baru.

Salah satu cara untuk mengetahui calon mahasiswa baru yang tidak heregistrasi adalah dengan melakukan prediksi apakah calon mahasiswa baru cenderung akan heregistrasi atau tidak menggunakan data mining.

Menurut [1] data mining adalah sebuah kegiatan yang berhubungan dengan pengumpulan data atau penggunaan data historis dengan tujuan menemukan informasi, pengetahuan, keteraturan, pola atau hubungan data yang berukuran besar. Hasil dari data mining dapat digunakan sebagai alternatif dalam penentuan atau penentuan keputusan pada masa yang akan datang.

Terdapat banyak teknik data mining yang bisa digunakan untuk prediksi, namun salah satu teknik data mining yang mempunyai akurasi tinggi untuk melakukan prediksi adalah metode klasifikasi algoritma Naive Bayes. Menurut [2] metode klasifikasi algoritma naive bayes adalah salah satu teknik data mining yang paling populer untuk mengklasifikasikan data dalam jumlah besar dan dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class dengan akurasi yang tinggi.

Beberapa penelitian sebelumnya yang dijadikan sebagai acuan yaitu [3] yang berjudul "Prediksi Masa Studi Mahasiswa Dengan Menggunakan Algoritma Naive Bayes" yang bertujuan untuk memprediksi ketepatan masa studi mahasiswa. Hasil penelitian tersebut menunjukkan bahwa algoritma Naive Bayes dapat menentukan prediksi masa studi mahasiswa dengan tingkat akurasi sebesar 85,17%.

Selanjutnya [4] yang berjudul "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis" yang bertujuan membandingkan dua algoritma dalam memprediksi penyakit Hepatitis. Hasil penelitian ini menunjukkan bahwa algoritma Naive Bayes memiliki nilai akurasi yang lebih tinggi untuk prediksi penyakit hepatitis yaitu 83,71%, sedangkan nilai akurasi algoritma C4.5 adalah 77,29%.

Selanjutnya [5] yang berjudul "Perbandingan Algoritma C4.5, KNN, dan Naive Bayes untuk Penentuan Model Klasifikasi Penanggung jawab BSI Entrepreneur Center" yang membahas perbandingan dari algoritma C.45, K-Nearest Neighbor, dan Naive Bayes untuk

mengklasifikasi penanggung jawab BSI entrepreneur center. Dari hasil pengujian masing-masing algoritma menunjukkan bahwa algoritma Naive Bayes merupakan algoritma yang paling tepat untuk klasifikasi penanggung jawab BSI entrepreneur center karena memiliki nilai akurasi yang paling tinggi yaitu 80%, sedangkan nilai akurasi algoritma C.45 adalah 73,33% dan algoritma K-Nearest Neighbor memiliki nilai akurasi 70%

Selanjutnya [6] yang berjudul "Sistem Prediksi Pengunduran Diri Calon Mahasiswa Baru Menggunakan Algoritma C45" yang membahas prediksi heregistrasi calon mahasiswa baru menggunakan algoritma C45. Hasil penelitian tersebut menunjukkan bahwa algoritma C4.5 dapat menentukan prediksi heregistrasi calon mahasiswa baru dengan tingkat akurasi sebesar 68,93%.

Berdasarkan hal tersebut, maka peneliti memilih menggunakan metode klasifikasi algoritma Naive Bayes. Algoritma ini akan diterapkan dengan permasalahan dan data dengan atribut yang berbeda. Peneliti akan menerapkan algoritma Naive Bayes untuk melakukan prediksi heregistrasi calon mahasiswa baru di UNISNU Jepara. Dataset yang akan digunakan adalah data calon mahasiswa baru UNISNU Jepara tahun 2019-2020 dengan 14 (empat belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 12 (dua belas) atribut reguler yaitu tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi. Melalui metode klasifikasi algoritma Naive Bayes diharapkan mampu memprediksi heregistrasi calon mahasiswa baru dengan lebih akurat.

2. Metode Penelitian

Tahapan yang dilalui dalam penelitian ini terdiri dari :

1. Pengumpulan Data

Data yang dikumpulkan pada penelitian ini adalah data calon mahasiswa baru yang diperoleh dari UPT Pangkalan Data dan Pengembangan Sistem Informasi UNISNU Jepara yang berupa data calon mahasiswa baru tahun 2019-2020 yang digunakan sebagai dasar untuk memprediksi heregistrasi calon mahasiswa baru

2. *Pre-processing* Data

Pre-processing data tersebut meliputi penyusunan dan pengkategorian data, seleksi atribut, dan pembersihan data agar memberikan hasil yang optimal

3. Penerapan Algoritma Naive Bayes

Setelah data di *pre-processing* maka data akan diolah dengan algoritma Naive Bayes menggunakan *tools* RapidMiner..

4. Evaluasi dan Validasi

Setelah dilakukan penerapan algoritma Naive Bayes maka akan dilakukan pengujian menggunakan *confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*)

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Tahapan pengumpulan data pada penelitian ini dilakukan di Universitas Islam Nahdlatul Ulama (UNISNU) Jepara di bagian UPT Pangkalan Data dan Pengembangan Sistem Informasi. Tahapan ini dilakukan dengan cara observasi dan wawancara perihal Penerimaan Mahasiswa Baru (PMB). Data yang diperoleh sebanyak 3.969 record dengan 18 (delapan belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 16 (enam belas) atribut reguler yaitu tahun pendaftaran, gelombang pendaftaran, jalur pendaftaran, program kelas, status awal pendaftaran, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jenis sekolah asal, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi dalam bentuk Microsoft Excel sehingga dapat mempermudah pengolahan data. Data tersebut diambil dari data pendaftaran calon mahasiswa baru pada tahun 2019-2020.

3.2 *Pre-processing* Data

a. Penyusunan Data

Setelah data didapatkan data kemudian dilakukan penyusunan dan pengkategorian data. Data dikategorikan agar mempermudah dan menghasilkan hasil yang maksimal dalam penerapan algoritma. Atribut dikategorikan berdasarkan kriteria atribut yang sudah ditentukan menggunakan software Microsoft Excel.

Sedangkan tipe data masing-masing atribut dapat dilihat pada tabel 1 sebagai berikut:

Tabel 1. Tipe Data

NO.	Nama	Tipe Data	Role
1.	Nama	Polynomial	ID
2.	Tahun Pendaftaran	Polynomial	Reguler
3.	Program Kelas	Binominal	Reguler
4.	Jenis Kelamin	Binominal	Reguler
5.	Usia	Polynomial	Reguler
6.	Prodi	Polynomial	Reguler
7.	Kota Asal	Polynomial	Reguler
8.	Pekerjaan Ayah	Polynomial	Reguler
9.	Pekerjaan Ibu	Polynomial	Reguler
10.	Penghasilan Orangtua	Polynomial	Reguler
11.	Jurusan Sekolah Asal	Polynomial	Reguler
12.	Nilai UN	Polynomial	Reguler
13.	Informasi Pendaftaran	Polynomial	Reguler
14.	Status Heregistrasi	Binominal	Label

b. Seleksi Atribut

Atribut diseleksi dengan melakukan pembobotan menggunakan operator *information gain*. Pada tahap ini akan dilakukan penyeleksian terhadap 16 (enam belas) atribut reguler dengan menghitung pembobotan/pengaruh atribut menggunakan operator *information gain* dengan software RapidMiner untuk menentukan atribut reguler apa saja yang berpengaruh.

attribute	weight
Gelombang Pendaftaran	0.017
Jalur Pendaftaran	0.001
Program Kelas	0.092
Status Awal Pendaftaran	0
Tahun Pendaftaran	1
Jenis Kelamin	0.127
Usia	0.052
Prodi	0.330
Kota Asal	0.489
Pekerjaan Ayah	0.144
Pekerjaan Ibu	0.057
Penghasilan Orangtua	0.041
Jenis Sekolah Asal	0.009
Jurusan Sekolah Asal	0.505
Nilai UN	0.138
Informasi Pendaftaran	0.052

Gambar 2. Hasil *Information Gain*

Dari gambar diatas dapat dilihat nilai pembobotan terhadap 16 (enam belas) atribut reguler. Berdasarkan nilai pembobotan tersebut, terdapat 4 (empat) atribut reguler yang nilai pembobotannya kecil/tidak berpengaruh yaitu gelombang pendaftaran, jalur pendaftaran, status awal pendaftaran, dan jenis asal sekolah. Sedangkan atribut reguler yang nilai pembobotannya besar/berpengaruh terdapat 12 (dua belas) atribut yaitu tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, informasi pendaftaran.

Sehingga pada penelitian ini atribut yang akan digunakan terdapat 14 (empat belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 12 (dua belas) atribut reguler yaitu tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi.

c. Pembersihan Data

Tahapan ini dilakukan untuk mengidentifikasi dan menghilangkan data yang kosong (missing value) menggunakan operator *replace missing values*.

Rk.	Nama	Status	Pr.	Jenis K.	Usia	Prodi	Kota Asal	Pekerja.	Pekerja.	Pengha.	Jurusa.	Nilai UN	Informa.	
1	DHULES AFREDA MAZINI	Ya	I	R.2	Peremp.	I	Manaje.	Jepara	Wirawa.	Lain-lan	I	Tata Nla.	XX	Temas
2	ERSHA JONATHAN HARWANTO	Ya	I	R.1	Laki-laki	I	Manaje.	Jepara	Pegawa.	Pegawa.	I	Obotoni	X	Internet
3	NOVITA SARU	Ya	I	R.2	Peremp.	I	Manaje.	Jepara	Wirawa.	Lain-lan	I	Adminis.	XX	Temas
4	UMAR SADI	Ya	I	R.1	Laki-laki	I	Manaje.	Jepara	Wahana.	I	IPA	XX	Temas	
5	ISMA ANIN NAWAR	Ya	I	R.1	Peremp.	I	Manaje.	Jepara	Wirawa.	Tidal Be.	II	IPS	XX	Internet
6	MA FLANSA ROSA	Ya	I	R.1	Peremp.	I	Manaje.	Jepara	Wirawa.	Tidal Be.	V	IPA	XX	Internet
7	DANAH NORTA SARU	Ya	I	R.1	Peremp.	II	Manaje.	Jepara	Lain-lan	Lain-lan	II	Teknisi.	XX	Temas
8	MULIAHATUN KHASANAH	Ya	I	R.1	Peremp.	I	Manaje.	Jepara	Pedagan.	Pedagan.	II	IPS	XX	Internet
9	FAUZI ANTONO	Ya	I	R.1	Laki-laki	I	Manaje.	Jepara	Tulang.	Wirawa.	I	IPA	XX	Temas
10	SITI ZAMATUS SAIDAH	Ya	I	R.1	Peremp.	I	Manaje.	Jepara	Wirawa.	Tidal Be.	II	Adminis.	XX	Temas
11	FAHRUR RIZAL	Ya	I	R.2	Laki-laki	I	Manaje.	Jepara	Pegawa.	Tidal Be.	II	Adminis.	XX	Temas
12	NANA NUR SAFITRI	Ya	I	R.1	Peremp.	I	Manaje.	Jepara	Wirawa.	Lain-lan	II	IPS	X	Temas
13	ANAKLA AYU NURHAN	Ya	I	R.2	Peremp.	I	Manaje.	Jepara	Pegawa.	Tidal Be.	II	Adminis.	XX	Temas
14	JAVIER BRYAN THEODORE	Ya	I	R.1	Laki-laki	I	Manaje.	Jepara	Pegawa.	Lain-lan	II	IPA	XX	Pembel.
15	RINA AGUSTINA	Ya	I	R.1	Peremp.	I	Manaje.	Jepara	Tulang.	Tidal Be.	I	Tata Nla.	XX	Radio

Gambar 3. Hasil Pembersihan Data

Dari gambar diatas dapat dilihat pada keterangan filter yaitu terdapat 2.853/3.969 data. Hal ini berarti dari proses pembersihan data terdapat 2.853 data bersih yang bisa digunakan dan terdapat 1.116 data kotor/kosong (missing values) yang tidak bisa digunakan.

3.3 Metode Naive Bayes

Setelah melalui tahapan pre-processing data, langkah selanjutnya adalah mengolah data menggunakan algoritma Naive Bayes. Pada penelitian ini, data dibagi menjadi 2 (dua) bagian, yaitu 70% digunakan sebagai data training dan 30% untuk data testing dari 1.518 data. Pembagian data dilakukan menggunakan operator split data.

Berikut ini adalah data yang akan diimplementasikan terhadap algoritma Naive Bayes:

Tabel 2. Data Penelitian

No	Nama	Tahun Pendaftaran	Program Kelas	Jenis Kelamin	Usia	Status Heregistrasi
1	Dhal es Afrida Mazi ah	I	R.2	Pere mpuan	I	Ya
2	Ersh a Jonathan Hary anto	I	R.1	Laki-laki	I	Ya
3	Novi ta Sari	I	R.2	Pere mpuan	I	Ya
4	Uma r Said	I	R.1	Laki-laki	I	Ya
5	Isna Anin Nah ar	I	R.1	Pere mpuan	I	Ya
6	Mia Flani a Ros a	I	R.1	Pere mpuan	I	Ya
:	:	:	:	:	:	:
2849	Muh amm ad Nauf al Arre za	II	R.1	Laki-laki	I	Ya
2850	Misy fa Nab eela h	II	R.1	Pere mpuan	I	Ya
2851	M. Auliy a Roh man	II	R.1	Laki-laki	I	Ya
2852	Taris sa Rah mi Ifada	II	R.1	Pere mpuan	I	Ya
28	Riya n	II	R.1	Laki-laki	I	Tida k

53	Afiaf ara					
----	-----------	--	--	--	--	--

a. Menghitung Probabilitas Kelas

Sebelum menentukan prediksi, langkah pertama yang dilakukan adalah menghitung probabilitas masing-masing kategori kelas pada atribut status heregistrasi dengan cara menjumlahkan masing-masing kategori kelas dibagi jumlah semua kategori kelas pada atribut status heregistrasi.

Maka:

$$1. P(C \text{ Pertama}) = P(\text{Class Pertama})$$

$$P(Ya) = 2626/2853$$

$$P(Ya) = 0,920$$

$$2. P(C \text{ Kedua}) = P(\text{Class Kedua})$$

$$P(\text{Tidak}) = 227/2853$$

$$P(\text{Tidak}) = 0,080$$

Sehingga diperoleh hasil sebagai berikut:

Tabel 3. Hasil Probabilitas Kelas

No	Status Heregistrasi	Probabilitas
1	Ya	0,920
2	Tidak	0,080
Jumlah Keseluruhan		1

b. Menghitung Probabilitas Atribut Target Prediksi/Reguler

Setelah menghitung probabilitas atribut kelas, langkah selanjutnya adalah menghitung probabilitas masing-masing kategori atribut target prediksi/reguler yaitu tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, dan informasi pendaftaran. Sampel yang digunakan adalah atribut program kelas.

Maka:

1. $P(X | \text{Program Kelas} = \text{"R.1"} | \text{Status Heregistrasi} =$

$\text{"Tidak"}) / 227$

$= 115/227$

$= 0,507$

2. $P(X | \text{Program Kelas} = \text{"R.2"} | \text{Status Heregistrasi} =$

$\text{"Tidak"}) / 227$

$= 112/227$

$= 0,493$

Sehingga diperoleh hasil sebagai berikut:

Tabel 4. Hasil Probabilitas Atribut Program Kelas

Probabilitas Program Kelas	Status Heregistrasi	
	Ya	Tidak
R.1	0,640	0,507
R.2	0,360	0,493
Jumlah	1	1

c. Menentukan Hasil Prediksi

Setelah menghitung probabilitas pada masing-masing kategori atribut dan kelas, langkah selanjutnya adalah menentukan hasil prediksi terhadap dataset berdasarkan hasil perhitungan probabilitas kategori atribut dan kelas.

Maka:

1. Dhaes Afrida Maziah =

- $P(X | \text{Status Heregistrasi} = \text{"Ya"})$

$= P(\text{Tahun Pendaftaran} = \text{"I"}) * P(\text{Program Kelas} = \text{"R.2"}) * P(\text{Jenis Kelamin} = \text{"Perempuan"}) * P(\text{Usia} = \text{"I"}) * P(\text{Prodi} = \text{"Manajemen"}) * P(\text{Kota Asal} = \text{"Jejara"}) * P(\text{Pekerjaan Ayah} = \text{"Wiraswasta"}) * P(\text{Pekerjaan Ibu} = \text{"Lain-lain"}) * P(\text{Penghasilan Orangtua} = \text{"II"}) * P(\text{Jurusan Sekolah Asal} =$

$\text{"Tata Niaga"}) * P(\text{Nilai UN} = \text{"XII"}) * P(\text{Informasi Pendaftaran} = \text{"Teman"}) * P(\text{Status Heregistrasi} = \text{"Ya"}) / (P(X = \text{"Ya"}) + P(X = \text{"Tidak"}))$

$= 0,499 * 0,360 * 0,551 * 0,850 * 0,127 * 0,903 * 0,467 * 0,261 * 0,396 * 0,008 * 0,186 * 0,513 * 0,920 / 0,0000003574$

$= 0,874$

- $P(X | \text{Status Heregistrasi} = \text{"Tidak"})$

$$\begin{aligned}
 &= P(\text{Tahun Pendaftaran} = \text{"I"}) * P(\text{Program Kelas} = \text{"R.2"}) * P(\text{Jenis Kelamin} = \text{"Perempuan"}) * P(\text{Usia} = \text{"I"}) * P(\text{Prodi} = \text{"Manajemen"}) * P(\text{Kota Asal} = \text{"Jepara"}) * P(\text{Pekerjaan Ayah} = \text{"Wiraswasta"}) * P(\text{Pekerjaan Ibu} = \text{"Lain-lain"}) * P(\text{Penghasilan Orangtua} = \text{"II"}) * P(\text{Jurusan Sekolah Asal} = \text{"Tata Niaga"}) * P(\text{Nilai UN} = \text{"XII"}) * P(\text{Informasi Pendaftaran} = \text{"Teman"}) * P(\text{Status Heregistrasi} = \text{"Tidak"}) / (P(X = \text{"Ya"}) + P(X = \text{"Tidak"})) \\
 &= 0,912 * 0,493 * 0,392 * 0,780 * 0,084 * 0,859 * 0,546 * 0,273 * 0,379 * 0,009 * 0,194 * 0,595 * 0,080 / 0,0000003574 \\
 &= 0,126
 \end{aligned}$$

Perhitungan diatas juga berlaku untuk record selanjutnya sampai record terakhir pada data. Sehingga diperoleh hasil prediksi sebagai berikut:

Tabel 5. Hasil Perhitungan Prediksi

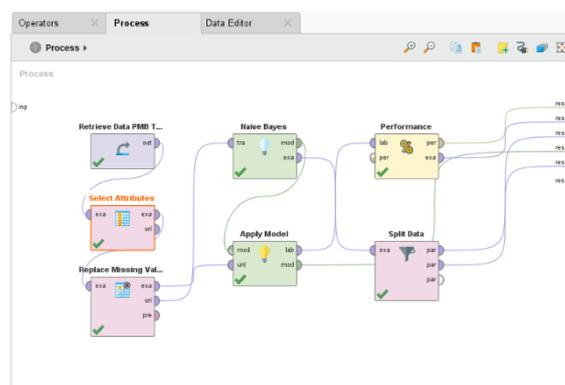
No	Nama	Class Prediction (Status Heregistrasi)	
		Ya	Tidak
1	Dhales Afrida Maziah	0,874	0,126
2	Ersha Jonathan Haryanto	0,833	0,167
3	Novita Sari	0,912	0,088
4	Umar Said	1,000	0,000
5	Isna Anin Nahar	0,943	0,057
6	Mia Flania Rosa	0,934	0,066
:	:	:	:
2849	Muhammad Naufal Arreza	0,973	0,027
2850	Misyfa Nabeelah	0,991	0,009
2851	M. Auliya Rohman	0,978	0,022
2852	Tarissa Rahmi Ifada	0,995	0,005
2853	Riyan Afiafara	0,005	0,995

Sehingga diperoleh hasil prediksi sebagai berikut:

Tabel 6. Hasil Prediksi

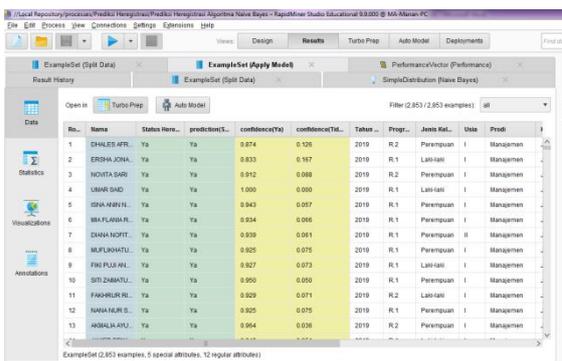
No	Nama	Status Heregistrasi	Nilai Max	Hasil Prediksi
1	Dhales Afrida Maziah	Ya	0,874	Ya
2	Ersha Jonathan Haryanto	Ya	0,833	Ya
3	Novita Sari	Ya	0,912	Ya
4	Umar Said	Ya	1,000	Ya
5	Isna Anin Nahar	Ya	0,943	Ya
6	Mia Flania Rosa	Ya	0,934	Ya
:	:			
2849	Muhammad Naufal Arreza	Ya	0,973	Ya
2850	Misyfa Nabeelah	Ya	0,991	Ya
2851	M. Auliya Rohman	Ya	0,978	Ya
2852	Tarissa Rahmi Ifada	Ya	0,995	Ya
2853	Riyan Afiafara	Tidak	0,995	Tidak

Pemodelan Naive Bayes pada RapidMiner:



Gambar 4. Pemodelan Naiv Bayes RapidMiner

Hasil Prediksi RapidMiner:



Gambar 5. Hasil Prediksi RapidMiner

3.4 Evaluasi Dan Validasi

Evaluasi dan validasi hasil pada penelitian ini dilakukan menggunakan confusion matrix sebagai performance dan kurva ROC (Receiver Operating Characteristic) yang menghasilkan nilai AUC (Area Under Curve). Adapun penjelasan lebih lengkap terkait evaluasi dan validasi akan diuraikan sebagai berikut:

Diketahui:

Tabel 7. Hasil Performance Vector

	True Ya	True Tidak
Prediksi Ya	2599	182
Prediksi Tidak	27	45

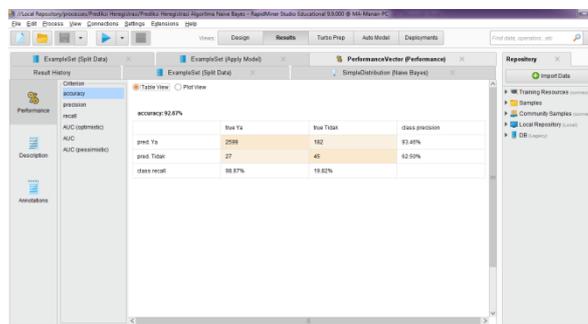
Berdasarkan hasil *performance vector* pada tabel 7, dapat disimpulkan bahwa dari dataset sebanyak 2.853 dengan kelas “Ya” = 2.626 dan kelas “Tidak” = 227 didapatkan hasil prediksi benar sebanyak 2.644 dengan prediksi benar kelas “Ya” = 2.599 dan prediksi benar kelas “Tidak” = 45, kemudian prediksi salah sebanyak 209 dengan prediksi salah kelas “Ya” = 27 dan prediksi salah kelas “Tidak” = 182. Sehingga dihasilkan:

- $$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\%$$

$$Accuracy = \frac{(2599 + 45)}{(2599 + 45 + 182 + 27)} \times 100\%$$

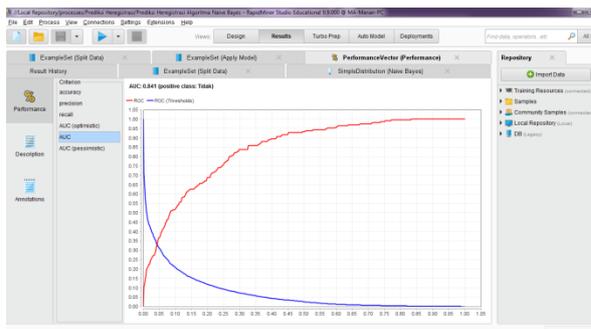
$$Accuracy = 92,67\%$$

Berdasarkan pengujian diatas dapat disimpulkan bahwa persentase *accuracy* yang dihasilkan adalah sebesar 92,67%. Berikut adalah hasil *performance accuracy* pada *software* RapidMiner:



Gambar 6. Hasil Accuracy RapidMiner

Berdasarkan pengujian yang telah dilakukan menggunakan *software* RapidMiner, dihasilkan nilai AUC (*Area Under Curve*) dari kurva ROC (*Receiver Operating Characteristic*) sebesar 0,841 sehingga dapat disimpulkan bahwa algoritma mempunyai kinerja yang baik. Berikut adalah hasil *performance AUC* pada *software* RapidMiner:



Gambar 6. Hasil AUC RapidMiner

4. Kesimpulan

Algoritma Naive Bayes memiliki kinerja yang baik dalam memprediksi heregistrasi calon mahasiswa baru dengan nilai akurasi sebesar 92,67% dan nilai AUC (*Area Under Curve*) sebesar 0,841. Atribut yang mempengaruhi heregistrasi terdapat 14 (empat belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 12 (dua belas) atribut reguler yaitu tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi.

Referensi

- [1] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [2] E. Prasetyo, "Data mining konsep dan aplikasi menggunakan matlab," *Yogyakarta Andi*, 2012.
- [3] M. winny Amelia, A. S. . Lumenta, and A. Jacobus, "Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes," *J. Tek. Inform.*, 2017, doi: 10.35793/jti.11.1.2017.17652.
- [4] W. D. Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *None*, vol. 13, no. 1, pp. 76–84, 2017.
- [5] F. N. Hasan, N. Hikmah, and D. Y. Utami, "Perbandingan Algoritma C4.5, KNN, dan Naive Bayes untuk Penentuan Model Klasifikasi Penanggung jawab BSI Entrepreneur Center," *J. Pilar Nusa Mandiri*, 2018.
- [6] S. Supria, L. Lidyawati, and S. Mawarni, "Sistem Prediksi Pengunduran Diri Calon Mahasiswa Baru Menggunakan Algoritma C45," in *Seminar Nasional Industri dan Teknologi*, 2018, pp. 227–236.
- [7] H. Amalia, A. Pohan, and S. Masripah, "Penerapan Feature Weighting Optimized Pada Naive Bayes Untuk Prediksi Proses Persalinan," *J. Pilar Nusa Mandiri*, vol. 15, no. 1 SE-Articles, Mar. 2019.
- [8] Kusri and L. E. Taufiq, *Algoritma Data Mining*. Yogyakarta: Andi, 2009.
- [9] Y. A. Sari and A. Arwan, "Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naive Bayes Human Detection and Tracking View project Smart Nutrition Box View project," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2018.
- [10] M. R. Maulana and M. A. Al Karomi, "Information Gain untuk Mengetahui Pengaruh Atribut Terhadap Klasifikasi Persetujuan Kredit," *J. PEMERINTAH KOTA PEKALONGAN*, vol. 9, 2015.
- [11] Y. B. Samponu and K. Kusri, "Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu," *J. ELTIKOM J. Tek. Elektro, Teknol. Inf. dan Komput.*, vol. 1, no. 2, pp. 56–63, 2017.
- [12] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016.
- [13] E. Turban, J. E. Aronson, and T. Liang, *Decision Support Systems and Intelligent System*, 7th ed. Yogyakarta: Penerbit : Andi, 2005.
- [14] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naive Bayes," vol. 2, pp. 84–92, 2020.
- [15] A. Siregar and A. Paspabhuana, "Pengolahan Data menjadi Informasi dengan Rapidminer." Surakarta, 2017.
- [1] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [2] E. Prasetyo, "Data mining konsep dan aplikasi menggunakan matlab," *Yogyakarta Andi*, 2012.

- [3] M. winny Amelia, A. S. . Lumenta, and A. Jacobus, "Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes," *J. Tek. Inform.*, 2017, doi: 10.35793/jti.11.1.2017.17652.