

## **KLASIFIKASI DATA *MINING* UNTUK MEMPREDIKSI STATUS PENERIMAAN DI PERGURUAN TINGGI NEGERI BAGI LULUSAN BIMBEL NF DENGAN ALGORITME *NAIVE BAYES***

Syafiq Abdurrohman<sup>1</sup>, Arief Wibowo<sup>2\*</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1\*</sup>syafiqabdurrahman12@gmail.com, <sup>2</sup>arief.wibowo@budiluhur.ac.id  
(\*: corresponding author)

**Abstrak-** Semakin pesatnya pertumbuhan teknologi informasi di dunia, sehingga semakin hari semakin banyak pula orang yang memanfaatkan kemajuan teknologi informasi ini dalam berbagai aktivitas kehidupan ataupun untuk menyelesaikan berbagai permasalahan. Salah satunya adalah dalam bidang pendidikan. Implementasinya seperti bagaimana membuat sistem untuk memprediksi status kelulusan siswa/i NF di Perguruan Tinggi Negeri dengan metode *naive bayes* berdasarkan dataset dan atribut yang telah didapatkan. Metode ini mengklasifikasikan data training yang nantinya menghasilkan suatu hasil kelulusan “LULUS” atau “TIDAK LULUS” nya siswa/i tersebut di fakultas dan jurusan yang ingin dia ambil. Sehingga apabila seorang siswa/i menginputkan data yang ada dalam aplikasi kemudian data tersebut diproses selanjutnya menghasilkan hasil kelulusan “LULUS” atau “TIDAK LULUS” nya siswa/i tersebut di fakultas dan jurusan yang ingin dia ambil. *Naive bayes* merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi. Metode yang dapat dipakai untuk mendukung pengambilan (hipotesa) dalam kelulusan siswa/i Nurul Fikri di Perguruan Tinggi Negeri. Dalam tugas akhir ini dibangun Web Aplikasi sistem klasifikasi untuk memprediksi status kelulusan siswa/i NF di Perguruan Tinggi Negeri. Pada proses training data diolah sebanyak 80% dan sisanya akan diolah pada proses calon mahasiswa uji sebanyak 20%. Dari proses testing tersebut menghasilkan kecocokan antara data asli dengan hasil prediksi program, dari algoritma *naive bayes* menghasilkan nilai *accuracy* 80,77 %, *recall* 65,52 dan *precision* 100%.

**Kata Kunci:** *data mining, naive bayes, klasifikasi, prediksi kelulusan siswa di perguruan tinggi negeri.*

## **CLASSIFICATION OF DATA MINING TO PREDICT THE STATUS OF ADMISSION IN STATE UNIVERSITIES FOR NF TUTORING GRADUATES WITH THE NAIVE BAYES ALGORITHM**

**Abstract-** The increasingly rapid growth of information technology in the world, so that more and more people are taking advantage of this advance in information technology in various life activities or to solve various problems. One of them is in the field of education. The implementation is like how to create a system to predict the graduation status of NF students in State Universities with the naive bayes method based on datasets and attributes that have been obtained. This method classifies training data that will produce a result of graduating "PASS" or "NOT PASS" the student in the faculty and department he wants to take. So that if a student inputs the data in the application then the data is processed then produces the result of graduating "PASS" or "NOT PASS" the student in the faculty and department he wants to take. Naive bayes is a statistical approach to induction inference on classification issues. A method that can be used to support the taking (hypothesis) in the graduation of NF students in State Universities. In this final project, a Web Application classification system was built to predict the graduation status of NF students in State Universities. In the training process, 80% of the data is processed and the rest will be processed in the process of prospective test students as much as 20%. From the testing process, it produces a match between the original data and the program's prediction results, from the naive bayes algorithm produces an accuracy value of 80.77%, recall of 65.52 and a precision of 100%.

**Keywords:** *data mining, naive bayes, b classification, prediction of student graduation in state universities.*

---

### **1. PENDAHULUAN**

Perkembangan Teknologi Informasi sekarang ini mengalami kemajuan yang sangat pesat dalam berbagai macam bidang seperti: kesehatan, industri, ekonomi, pendidikan dan lain-lain[1]. Dengan adanya Teknologi Informasi memudahkan segala pekerjaan manusia yang berkaitan dengan mengolah data (*data processing*) secara akurat cepat dan efektif.

Paradigma orang tua terhadap pendidikan juga semakin membaik, kepedulian orang tua terhadap pendidikan anaknya sangat tinggi ini terbukti dari banyaknya orang tua yang mendorong anaknya untuk mengikuti pendidikan non formal seperti bimbingan belajar (bimbel), les private maupun *homeschooling* guna memudahkan anak mendapatkan pendidikan terbaik di jenjang selanjutnya.

Masyarakat juga sekarang ini sangat memperhatikan kualitas dari pendidikan terutama di sekolah [2]. Mereka selalu mencari tempat pendidikan formal (sekolah atau perguruan tinggi) yang terbaik dengan berusaha memenuhi persyaratan atau seleksi masuk sekolah atau perguruan tinggi. Dengan demikian Sekolah Menengah Atas (SMA) pun berlomba-lomba memperbaiki kualitas pendidikan mereka dengan meraih angka terbanyak lulusan yang diterima di Perguruan Tinggi Negeri (PTN).

Salah satu indikator baiknya kualitas sekolah menengah atas dan kejuruan adalah tingkat diterimanya siswa di Perguruan Tinggi Negeri (PTN) [3]. Data siswa dari beberapa Sekolah Menengah Atas dan kejuruan dianalisis untuk mengetahui tingkat probabilitasnya diterimanya siswa di Perguruan Tinggi Negeri berdasarkan dataset yang telah didapatkan beserta atribut nilai TO UTBK yang dilakukan secara berkala.

Dengan memprediksi siswa yang masuk Perguruan Tinggi Negeri, dapat dimanfaatkan oleh pihak Bimbingan Belajar Nurul Fikri untuk bahan promosi[4]. Proses analisis data siswa tersebut menggunakan teknik data *mining* dan proses klasifikasinya menggunakan aplikasi spreadsheet seperti Microsoft Excel dalam bentuk pola kategori “LULUS” dan pola untuk kategori “TIDAK LULUS”.

Dengan tujuan penelitian ini untuk memprediksi siswa angkatan berikutnya yang masuk di Perguruan Tinggi Negeri (PTN), menggunakan hasil model klasifikasi yang terbentuk[5]. Pada penelitian ini data yang digunakan adalah data siswa angkatan 2021 Bimbingan Belajar Nurul Fikri. Yang berjumlah 260 data, Proses data *mining* menggunakan klasifikasi *naive bayes*. *Naive bayes* merupakan metode pengklasifikasian yang memanfaatkan probabilitas dan statistik untuk memprediksi peluang di masa depan dengan memanfaatkan pengalaman di masa sebelumnya[6].

Berdasarkan uraian di atas maka diperlukan sebuah aplikasi untuk memprediksi siswa yang diterima di Perguruan Tinggi Negeri (PTN) menggunakan klasifikasi *naive bayes* yang dapat diakses dengan mudah oleh siapapun dan dimanapun maka penulis membuat penelitian dengan judul “Klasifikasi Data *mining* Untuk Memprediksi Status Penerimaan Di Perguruan Tinggi Negeri Bagi Lulusan Bimbel NF Dengan Algoritma *Naive Bayes*”.

## 2. METODE PENELITIAN

### 2.1 Data Penelitian

Data penelitian didapatkan dari bimbingan belajar Nurul Fikri, Dalam melakukan penelitian ini dilakukan pengumpulan data siswa/I yang lulus di PTN mengambil dari Tahun sebelumnya yaitu TA 2021. Peneliti melakukan penelitian di tempat tersebut dengan waktu hampir 1 minggu, dimulai dari tanggal 21 Maret 2022 sampai 25 Maret 2022.

Didapatkan data awal berjumlah 260 data, diambil dari data asli 12 kali TO lalu di rata-rata skor nya dan jurusan diubah menjadi SAINTEK dan SOSHUM berikut data asli sebelum diubah.

**Tabel 1.** Data Asli

No	Nama PTN	Jurusan	Skor	Label
1	Universitas Padjajaran	Psikologi	568,005	Lulus
2	Universitas Indonesia	Farmasi	588,045	Lulus
3	Universitas Bwawijaya	Perpajakan	579,850	Lulus
4	Universitas Gajah Mada	Akuntansi	458,900	Tidak Lulus
...	...	...	...	...
260	Universitas Diponegoro	Arsitektur	397,285	Tidak Lulus

Data setelah diubah, untuk mempermudah proses klasifikasi.

**Tabel 2.** Data Setelah diubah

No	Nama PTN	Jurusan	Skor	Label
1	Universitas Padjajaran	SAINTEK	>500	Lulus
2	Universitas Indonesia	SAINTEK	>500	Lulus
3	Universitas Bwawijaya	SOSHUM	>500	Lulus
4	Universitas Gajah Mada	SOSHUM	401-500	Tidak Lulus
...	...	...	...	...
260	Universitas Diponegoro	SAINTEK	301-400	Tidak Lulus

## 2.2 Jenis Data

Adapun jenis data yang digunakan dalam penelitian ini adalah jenis data primer dan sekunder:

- Data primer, yaitu data yang langsung dikumpulkan oleh peneliti (atau petugasnya) dari sumber pertamanya (Sumadi Suryabrata, 1987). Adapun yang menjadi sumber data primer dalam penelitian ini adalah data Try Out UTBK 1-12 Siswa/i Bimbingan Belajar Nurul Fikri TA 2020/2021.
- Data sekunder, yaitu data yang langsung dikumpulkan oleh peneliti sebagai penunjang dari sumber pertama. Dapat juga dikatakan data yang tersusun dalam bentuk dokumen-dokumen. Dalam penelitian ini, dokumentasi dan angket merupakan sumber data sekunder.

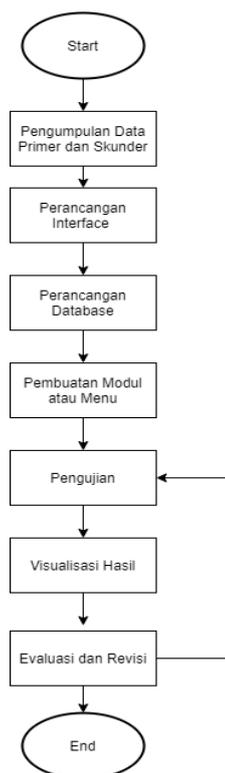
## 2.3 Sumber Data

Yang dimaksud dengan sumber data dalam penelitian adalah subjek dari mana data dapat diperoleh. Untuk Lebih jelas mengenai data dan sumber data yang digunakan dalam penelitian ini, maka peneliti mengumpulkan data dan menyajikan data pada tabel 3.1 dibawah ini:

**Tabel 3.** Jenis dan Sumber Data Penelitian

No	Data Penelitian	Jenis Data	Sumber Data
1	Profil Bimbingan Belajar Nurul Fikri	Sekunder	<a href="https://www.bimbelnurulfikri.id/">https://www.bimbelnurulfikri.id/</a>
2	Data Siswa/i Bimbingan Belajar Nurul Fikri	Primer	Internal Bimbingan Belajar Nurul Fikri
3	Data Try Out UTBK 1-12 Bimbingan Belajar Nurul Fikri	Primer	Internal Bimbingan Belajar Nurul Fikri

Untuk tahap penelitiannya bisa dilihat Di Gambar 1



**Gambar 1.** Tahapan Metode Penelitian

Tahapan penelitian dimulai dengan mengumpulkan dataset yang akan digunakan pada penelitian ini, data ini diperoleh dari data siswa lulusan tahun 2021. kemudian data tersebut 80% digunakan untuk data training dan 20% untuk data uji. Kemudian pembuatan interface yaitu tampilan layar dari aplikasi berbasis web yang user friendly dan mudah dipahami pengguna.

Tahapan selanjutnya adalah perancangan database di dalam aplikasi database mysql dengan memperhatikan tipe data dan ukuran datanya. Setelah itu pembuatan modul atau menu yang dibutuhkan untuk mengarahkan (navigasi) dari satu halaman ke halaman lainnya.

Tahapan selanjutnya adalah pengujian untuk menguji sejauh mana sistem berjalan sesuai dengan tujuan perancangan dengan melihat visual dari hasil akhir yang diharapkan. Tahapan terakhir adalah evaluasi jika ada kekurangan maka harus diperbaiki dan dilakukan pengujian lagi.

## 2.4 Rancangan Basis Data

Dalam proses pembuatan aplikasi ini, dibutuhkan basis data yang berisikan semua data untuk menjalankan aplikasi sebagai berikut :

**Tabel 4.** Tabel *Users*

No	Nama Kolom	Tipe	Lebar	Keterangan
1	Id_user	Int	20	Primary Key
2	nama	Varchar	255	
3	username	Varchar	255	
4	password	Varchar	255	
5	akses	Varchar	40	

**Tabel 5.** Tabel siswa

No	Nama Kolom	Tipe	Lebar	Keterangan
1	Id	Int	3	Primary Key
2	Nama_PTNI	Varchar	38	
3	Jurusan_PTNI	Varchar	7	
4	Rata_TO	Varchar	9	
5	Kelulusan	Varchar	5	

**Tabel 6.** Tabel Uji

No	Nama Kolom	Tipe	Lebar	Keterangan
1	Id_klasifikasi	Int	50	Primary Key
2	Nama_PTNI	Varchar	38	
3	Jurusan_PTNI	Varchar	7	
4	Rata_TO	Varchar	9	
5	Kelas_asli	Varchar	5	

**Tabel 7.** Tabel Klasifikasi

No	Nama Kolom	Tipe	Lebar	Keterangan
1	Id_klasifikasi	Int	50	Primary Key
2	Id_uji	Int	50	
3	Nama_PTNI	Varchar	38	
4	Jurusan_PTNI	Varchar	7	
5	Rata_TO	Varchar	9	
6	Kelas_asli	Varchar	5	

**Tabel 8.** Tabel Universitas

No	Nama Kolom	Tipe	Lebar	Keterangan
1	Id	Int	2	Primary Key
2	Nama_Universitas	Varchar	38	

## 2.5 Metode Naive Bayes

*Naive bayes* adalah konsep dari probabilitas yang dipakai sebagai cara untuk menentukan kelompok kelas dokumen teks dan juga dapat mengolah data dalam jumlah besar serta menghasilkan akurasi yang cukup tinggi (Hearst, 1999) dan *naive bayes* memiliki tingkat akurasi yang lebih baik dibanding model classifier lainnya (Xhemali et al., 2009).

Pada algoritma *naive bayes* setiap dokumen dipresentasikan dengan memasukkan “a1, a2, a3, ..., an” di mana a1 adalah kata pertama dan berikutnya sampai, sedangkan V yaitu label katagori, Selanjutnya yaitu mencari nilai tertinggi dari kategori teks yang diujikan (V\_MAP) persamaan (V\_MAP) yaitu pada rumus 1 sebagai berikut:

$$V_{AP} = \operatorname{argmax}_{v \in V} P(V_j) \cap_i P(a_i | V_j) \quad (1)$$

Nilai  $P(V_j)$  dihitung pada saat data latih, dengan rumus 2 sebagai berikut:

$$V_j = \frac{|Docs_j|}{|training|} \quad (2)$$

$|docs_j|$  adalah dokumen yang memiliki kategori j pada dokumen latih dan  $|training|$  adalah jumlah dokumen latih.

$$P(a_i | V_j) = \frac{|n+1|}{|n+kosakata|} \quad (3)$$

$$P(a_i | V_j) = \frac{|n+1|}{|n+kosakata|}$$

- Ni: jumlah kemunculan kata ai pada dokumen yang berkategori vj.
- n: jumlah seluruh kata pada dokumen yang berkategori vj.
- kosakata: jumlah kata pada seluruh dokumen latih.

## 2.6 Confusion Matrix Dan Pengukuran Performa

Tujuan *Confusion Matrix* menganalisa kualitas kinerja model klasifikasi dalam mengenali variabel dari seluruh kelas. *Confusion Matrix* berisi informasi mengenai kelas sebenarnya dan kelas prediksi dari suatu proses klasifikasi (Anam & Santoso, 2018). Tabel matrix digunakan untuk mempresentasikan hasil evaluasi model klasifikasi. Misalnya data set terbagi menjadi kelas LULUS dan kelas TIDAK LULUS, maka kelas LULUS diasumsikan sebagai variabel positif dan kelas TIDAK LULUS diasumsikan sebagai variabel negatif. Nilai *accuracy*, *reccal* dan *precision* dapat diperoleh dari hasil evaluasi menggunakan *Confusion Matrix*. Tabel 9 merupakan contoh *Confusion Matrix*:

**Tabel 9. Confusion Matrix**

		Lulus	Tidak Lulus	Jumlah
Kelas	Lulus	TP	FN	P
	Tidak Lulus	FP	TN	N
Aktual	Jumlah	P	N	P + N

Perhitungan nilai akurasi, precision dan recall dinyatakan dalam persamaan berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$precision = \frac{TP}{TP+FP} \quad (5)$$

$$call = \frac{TP}{TP+FN} \quad (6)$$

Keterangan:

TP (*True Positive*) : Jumlah variabel positif yang dilabeli dengan benar oleh *classifier*, sebagai contoh variabel dengan label status kelulusan = tepat waktu.

TN (*True Negative*) : Jumlah variabel negatif yang dilabeli dengan benar oleh *classifier*.

FP (*False Positive*) : Jumlah variabel negatif yang salah dilabeli oleh *classifier*.

FN (*False Negative*) : Jumlah variabel positif yang salah dilabeli oleh *classifier*.

P : Jumlah sampel positif.

N : Jumlah sampel negative.

### 3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi analisis, hasil implementasi ataupun pengujian serta pembahasan dari topik penelitian, yang bisa dibuat terlebih dahulu metodologi penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, tabel dan lainnya.

#### 3.1 Hitung Probabilitas

Adapun nilai probabilitas didapatkan dari data training, berikut perhitungan probabilitas jumlah class, mencari peluang nama di PTN, mencari peluang jurusan di PTN, mencari peluang rata TO bisa dilihat pada penjelasan dibawah.

##### a. Menghitung Jumlah Class/label

Pada Label Data Training terdapat 208 data. Jumlah siswa/i yang lulus di PTN

**Tabel 10. Menghitung Jumlah Class/Label**

Menghitung Jumlah class/label			
Label	Jumlah Data	Jumlah Seluruh Data	Hasil
Lulus	127	208	0,6105769
Tidak Lulus	81	208	0,3894231

##### b. Mencari Probabilitas Peluang Nama PTN

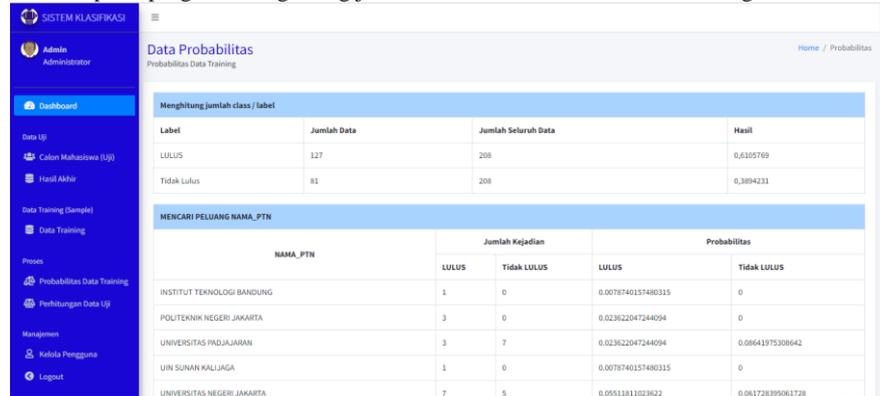
Ada 45 universitas dicari jumlah kejadian dan probabilitas masing-masing universitas.

**Tabel 11. Mencari Peluang Nama PTN**

NAMA_PTNI	Jumlah Kejadian		Probabilitas	
	A	B	LULUS	Tidak LULUS
	LULUS	Tidak LULUS		
UNIVERSITAS LAMPUNG	3	0	0.023622047244094	0
UNIVERSITAS SULTAN AGENG TIRTAYASA	13	2	0.10236220472441	0.024691358024691
INSTITUT TEKNOLOGI SUMATERA	8	0	0.062992125984252	0
...	...	...	...	...

NAMA_PTN	Mencari Peluang Nama PTN		Probabilitas	
	Jumlah Kejadian		LULUS	Tidak LULUS
	A LULUS	B Tidak LULUS		
INSTITUT TEKNOLOGI BANDUNG	1	0	0.0078740157480315	0
Jumlah	127	81	1	1

Berikut adalah gambar tampilan program menghitung jumlah class / label dan Mencari Peluang Nama PTN



The screenshot shows a web application interface with a sidebar menu and a main content area. The main content area displays two tables. The first table, titled 'Menghitung jumlah class / label', shows the following data:

Label	Jumlah Data	Jumlah Seluruh Data	Hasil
LULUS	127	208	0,605769
Tidak Lulus	81	208	0,3894231

The second table, titled 'MENCARI PELUANG NAMA\_PTN', shows the following data:

NAMA_PTN	Jumlah Kejadian		Probabilitas	
	LULUS	Tidak LULUS	LULUS	Tidak LULUS
INSTITUT TEKNOLOGI BANDUNG	1	0	0.0078740157480315	0
POLITEKNIK NEGERI JAKARTA	3	0	0.023622047244094	0
UNIVERSITAS PADJARAN	3	7	0.023622047244094	0.08641975308642
UIN SUNAN KALLAJA	1	0	0.0078740157480315	0
UNIVERSITAS NEGERI JAKARTA	7	5	0.05511811023622	0.06172839501728

**Gambar 2.** Menghitung Jumlah Class Dan Mencari Peluang Nama\_PTN

- c. Mencari Probabilitas Peluang Jurusan PTN  
 Ada 2 jurusan SAINTEK dan SOSHUM dicari jumlah kejadian dan probabilitas masing-masing jurusan.

**Tabel 12.** Mencari Peluang Jurusan PTN

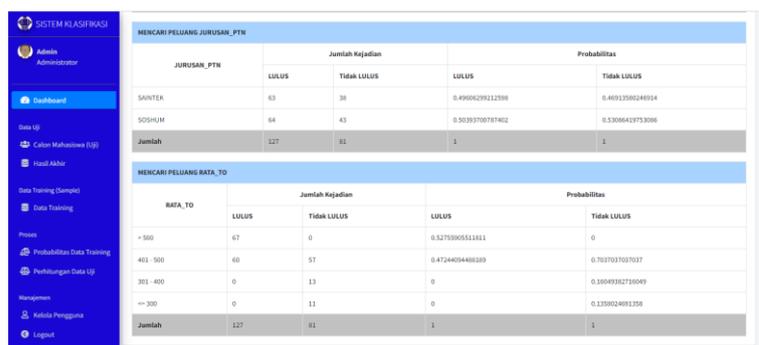
JURUSAN PTN	Mencari Peluang Jurusan PTN		Probabilitas	
	Jumlah Kejadian		LULUS	Tidak LULUS
	A LULUS	B Tidak LULUS		
SOSHUM	64	43	0.50393700787402	0.53086419753086
SAINTEK	63	38	0.49606299212598	0.46913580246914
Jumlah	127	81	1	1

- d. Mencari Probabilitas Peluang Rata TO  
 Ada 4 kategori nilai, setiap nilai dicari jumlah kejadian dan probabilitas masing-masing nilai.

**Tabel 13.** Mencari Peluang Rerata TO

RATA_TO	Mencari Peluang Rata TO		Probabilitas	
	Jumlah Kejadian		LULUS	Tidak LULUS
	A LULUS	B Tidak LULUS		
> 500	67	0	0.52755905511811	0
401 - 500	60	57	0.47244094488189	0.7037037037037
301 - 400	0	13	0	0.16049382716049
<= 300	0	11	0	0.1358024691358
Jumlah	127	81	1	1

Berikut adalah gambar tampilan program Mencari Probabilitas Jurusan\_PTN dan Peluang Rata\_TO



The screenshot shows two tables in the application interface. The first table, titled 'MENCARI PELUANG JURUSAN\_PTIN', shows data for two departments: SAINTEK and SOSHUM. The second table, titled 'MENCARI PELUANG RATA\_TO', shows data for three score ranges: <math>\le 500</math>, 401 - 500, and <math>> 500</math>.

JURUSAN_PTIN	Jumlah Kejadian		Probabilitas	
	LULUS	Tidak LULUS	LULUS	Tidak LULUS
SAINTEK	63	30	0.4960299122796	0.4691359024034
SOSHUM	64	43	0.5039700877203	0.5308642975966
Jumlah	127	73	1	1

RATA_TO	Jumlah Kejadian		Probabilitas	
	LULUS	Tidak LULUS	LULUS	Tidak LULUS
<math>\le 500</math>	67	0	0.527590551811	0
401 - 500	60	57	0.4724094481889	0.7037037037037
301 - 400	0	13	0	0.3809523809523809
<math>> 500</math>	0	11	0	0.2590244927536
Jumlah	127	81	1	1

Gambar 3. Mencari Probabilitas Jurusan\_PTIN dan Peluang Rata\_TO

### 3.2 Hasil Pengujian Metode Naive Bayes

Pengujian ini dilakukan untuk memprediksi kelulusan siswa/i NF yang dimasukan dan dihitung dengan algoritma *naive bayes* memiliki beberapa persen tingkat akurasi nya, recallnya, dan presisinya. Dengan data testingnya 52 data siswa (20%) yang diambil dari 260 data.

### 3.3 Perhitungan Confusion Matrix

Berikut ini adalah perhitungan *Confusion Matrix* yang akan mencari Akurasi, Recall, dan Presisi dari data testing, yang membandingkan antara data sebelum masuk kedalam sistem dengan data yang sudah masuk ke dalam sistem. Tabel Perhitungan *Confusion Matrix* dapat dilihat pada tabel 14.

Tabel 14. Perhitungan *Confusion Matrix*

Label	Hasil Prediksi	
	Positive	Negative
LULUS	19	10
TIDAK LULUS	0	23

Dari data table diatas bisa didapatkan nilai akurasi sebagai berikut:

$$\text{True Positive (TP)} = 19$$

$$\text{True Negative (TN)} = 23$$

$$\text{False Positive (FP)} = 0$$

$$\text{False Negative (FN)} = 10$$

a. Perhitungan akurasi:

$$= \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$

$$= \frac{19+23}{19+0+23+10} \times 100\% = \frac{42}{52} \times 100\% = 80,77\%$$

Tingkat akurasi dari sitem ini dengan menggunakan algoritma *Naive Bayes* sebesar 80,77%

b. Perhitungan *Recall*

$$= \frac{TP}{TP+FN} \times 100\%$$

$$= \frac{19}{19+10} \times 100\% = \frac{19}{29} \times 100\% = 65,52\%$$

Perhitungan *Recall* yang didapat sebesar 65,52%

c. Perhitungan *Precision*

$$= \frac{TP}{TP+FP} \times 100\%$$

$$= \frac{19}{19+0} \times 100\% = \frac{19}{19} \times 100\% = 100\%$$

Perhitungan *Precision* yang didapat sebesar 100%

Berikut adalah gambar tampilan program hasil akhir

Label	Positive	Negative
LULUS	19	10
TIDAK LULUS	0	23

Perhitungan Akurasi	
True Positive (TP)	19
True Negative (TN)	23
False Positive (FP)	0
False Negative (FN)	10
<b>Total</b>	<b>52</b>
<b>Akurasi (%)</b>	<b>80,77 %</b>

Perhitungan Recall dan Precision	
<b>Recall</b>	<b>65,52 %</b>
<b>Precision</b>	<b>100,00 %</b>

Syafiq © 2022 Universitas Budi Luhur

Gambar 4. Hasil Akhir

#### 4. KESIMPULAN

Dari analisis pembahasan yang telah dilakukan pada klasifikasi data mining untuk memprediksi status penerimaan di Perguruan Tinggi Negeri bagi lulusan bimbel NF, maka dapat diambil beberapa kesimpulan sebagai berikut: Penggunaan algoritma klasifikasi *naive bayes* dapat memprediksi diterima siswa NF di perguruan tinggi negeri dengan akurasi 80,77%, recall 65,52 % dan precision 100%. Perancangan aplikasi berbasis web untuk perhitungan algoritma *naive bayes* dapat berjalan dengan baik dan Berdasarkan data siswa yang diperoleh, proses penerapan metode *naive bayes* menghasilkan prediksi dari 2 data testing adalah pertama “LULUS” dan “TIDAK LULUS”.

#### DAFTAR PUSTAKA

- [1] R. Budiman and R. Anto, “Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering),” *ProTekInfo(Pengembangan Ris. dan Obs. Tek. Inform.*, vol. 6, no. 1, p. 6, 2019.
- [2] E. Etriyanti, D. Syamsuar, and N. Kunang, “Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa,” *Telematika*, vol. 13, no. 1, pp. 56–67, 2020.
- [3] A. H. Hailitik, B. S. Djahi, Y. Y. Nabuasa, J. I. Komputer, and U. N. Cendana, “Klasifikasi Jurusan Menggunakan Metode Naive Bayes Pada,” vol. 5, no. 2, pp. 21–27, 2017.
- [4] I. Irmawati, Z. Zainudin, and Y. Yuyun, “Data Mining Untuk Penentuan Model Kelulusan Murid Sma Pada Perguruan Tinggi Negeri; Studi Kasus Di Iain Bone,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 2, pp. 113–118, 2020.
- [5] K. Khoirunnisa, L. Susanti, I. T. Rokhmah, and L. Stianingsih, “Prediksi Siswa Smk Al-Hidayah Yang Masuk Perguruan Tinggi Dengan Metode Klasifikasi,” *J. Inform.*, vol. 8, no. 1, pp. 26–33, 2021.
- [6] O. J. Risza Putri Elburdah, “<http://ejurnal.binawakya.or.id/index.php/MBI> Vol.15 No.11 juni 2021 Open Journal Systems,” vol. 15, no. 11, pp. 5651–5658, 2021.
- [7] A. Saifudin, “Metode Data Mining Untuk Seleksi Calon Mahasiswa,” vol. 10, no. 1, pp. 25–36, 2018.
- [8] H. Y. Sardi and K. Budayawan, “Klasifikasi Tingkat Kelulusan Mahasiswa Elektronika Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus : Pendidikan Teknik Informatika FT-UNP),” *Voteteknika (Vocational Tek. Elektron. dan Inform.*, vol. 8, no. 4, p. 147, 2020, .
- [9] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018.
- [10] B. Yusuf, M. Qalbi, B. Basrul, I. Dwitawati, M. Malahayati, and M. Ellyadi, “Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh,” *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 4, no. 1, p. 50, 2020.
- [11] M. A. Amri, A. P. Windarto, A. Wanto, and I. S. Damanik, “Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas Yang Dimiliki Desa,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 674–679, 2019.
- [12] T. Dwinasty, I. Asror, and y r Murti, “Klasifikasi Program Studi Berdasarkan Nilai Utbk Menggunakan Metode Pseudo Nearest Neighbor Rule,” *eProceedings of Engineering*, vol. 7, no. 3, pp. 9835–9855, 2020.
- [13] Rahmawati and T. Arifin, “Penerapan Algoritma K-Means Untuk Pengelompokan Siswa Lolos Snmptn Di SMAN 8 Bandung,” *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 184–190, 2020.