

BAB IV

HASIL DAN ANALISIS

4.1 MODEL ALGORITMA *NAÏVE BAYES*

Algoritma yang digunakan untuk membuat model penelitian adalah algoritma *Naïve Bayes*. Algoritma *Naïve Bayes* digunakan untuk mengklasifikasi serta memprediksi penerimaan mahasiswa pada *dataset* Data Mahasiswa Baru. Sebelum melakukan proses pengolahan *dataset* menggunakan *tools*, akan dilakukan proses perhitungan secara manual terlebih dahulu. Perhitungan secara manual dilakukan untuk memahami proses pengolahan data menggunakan algoritma *Naïve Bayes*. Proses perhitungan manual akan dilakukan menggunakan bantuan *software Microsoft Excel*. Selanjutnya akan dilakukan proses pengolahan menggunakan bantuan *Tools WEKA*. *Tools WEKA* dipilih karena *tools* tersebut memiliki kumpulan fitur yang dapat membantu dalam mengolah data dengan algoritma yang diinginkan, dan mengolah data dalam jumlah banyak. Beberapa tahapan yang terdapat pada proses pengolahan *dataset* Data Mahasiswa Baru menggunakan *tools WEKA* yaitu visualisasi atribut, dan hasil klasifikasi menggunakan algoritma *Naïve Bayes*.

4.2 PERHITUNGAN MANUAL *NAÏVE BAYES*

Dalam melakukan proses perhitungan dibutuhkan *data training* dan *data testing*. *Data training* dan *data testing* yang akan digunakan pada Proses perhitungan manual akan diperoleh dari cuplikan data yang diambil dari *dataset*

Data Mahasiswa Baru. Cuplikan *dataset* yang akan digunakan adalah sebanyak 100 data dengan sistem pembagian 70% *data training* dan 30% *data testing* dan didapatkan *data training* sebanyak 70 data dan *data testing* sebanyak 30 data. Tahapan pertama yang dikerjakan adalah menghitung nilai probabilitas tiap kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut, untuk data dengan tipe numerik akan dilakukan perhitungan nilai rata-rata dan standar deviasi dari masing-masing parameter yang menggambarkan angka. Cuplikan *data training* dapat dilihat pada tabel 4.1

Tabel 4.1 Cuplikan *Data Training*

Nilai Mapel UN	Ranking Sekolah	Pilihan 1	Pilihan 2	Lulus Pilihan
82.75	37	TEKNIK INFORMATIK A	MANAJEMEN	Tidak Lulus
81.03	50	ILMU POLITIK	PENDIDIKAN LUAR SEKOLAH	pil_2
81.28	182	PENDIDIKAN GEOGRAFI	PENDIDIKAN LUAR SEKOLAH	pil_1
81.25	152	KESEHATAN MASYARAKAT	AGRIBISNIS	Tidak Lulus
80.58	187	KESEHATAN MASYARAKAT	MANAJEMEN	Tidak Lulus
...
83.95	105	AKUNTANSI	EKONOMI PEMBANGUNAN	Tidak Lulus
81.38	143	MANAJEMEN	PENDIDIKAN EKONOMI	Tidak Lulus
81.64	183	TEKNIK SIPIL	KESEHATAN MASYARAKAT	Tidak Lulus
82.86	234	AKUNTANSI	MANAJEMEN	Tidak

				Lulus
83.44	64	KESEHATAN MASYARAKA T	GIZI	Tidak Lulus

Data training yang digunakan pada proses perhitungan manual algoritma *Naïve Bayes* berjumlah 70 data. Berikut adalah proses perhitungan Manual algoritma *Naïve Bayes*:

1. Menghitung Jumlah Dan Probabilitas

Tahapan pertama yang dikerjakan adalah menghitung nilai probabilitas tiap kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut. Rumus yang digunakan dalam menghitung probabilitas label adalah:

$$P = \left(\frac{X}{C}\right) \dots \dots \dots (4.1)$$

Keterangan:

P = Probabilitas

X = Jumlah Kelas Label Pada *Data Training*

C = Jumlah *Data Training*

2. Menghitung Nilai Rata-Rata dan Standar Deviasi

Untuk data dengan tipe numerik akan dilakukan perhitungan nilai *mean* dan standar deviasi dari masing-masing parameter yang menggambarkan angka.

Rumus yang digunakan untuk menghitung nilai mean adalah sebagai berikut:

$$\mu = \sum_{i=1}^n xi \text{ atau } \mu = \frac{x1+x2+x3+\dots+xn}{n} \dots \dots \dots (4.2)$$

Keterangan:

μ : rata-rata hitung

x_i : nilai sampel ke-i

n : jumlah sampel

Dan rumus persamaan standar deviasi adalah sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots \dots \dots (4.3)$$

Keterangan:

σ : standar deviasi

μ : rata-rata hitung

x_i : nilai x ke-i

n : jumlah sampel

3. Nilai Distribusi *Gaussian*

Tahapan selanjutnya adalah menghitung nilai probabilitas untuk *data testing* dengan tipe data numerik menggunakan persamaan *Gaussian*.

Rumus persamaan dalam mencari nilai distribusi *Gaussian* adalah sebagai berikut:

$$P = (X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \times e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots \dots \dots (4.4)$$

Keterangan :

P : Peluang

X_i : Atribut ke-i

Y_j : Nilai atribut ke-i

σ : standar deviasi atribut

μ : rata-rata hitung (*mean*) atribut

4. Probabilitas Akhir Setiap Kelas

Selanjutnya adalah menghitung probabilitas akhir setiap kelas dengan mengkalikan semua probabilitas atribut. Untuk atribut dengan data numerik nilai probabilitas diambil dari nilai distribusi *Gaussian*.

5. Probabilitas Akhir

Probabilitas akhir didapatkan dengan menghitung nilai probabilitas akhir kelas menggunakan *Naïve Bayes Classifier*. Rumus perhitungan probabilitas akhir adalah sebagai berikut:

$$P(\text{Kelas}|X) = P(\text{Kelas}) * P(X) \dots \dots \dots (4.5)$$

4.2.1 Hasil Perhitungan Manual Algoritma *Naïve Bayes*

1. Menghitung Jumlah Dan Probabilitas

Terdapat 4 atribut yang akan dihitung pada cuplikan *dataset* yang diperoleh dari *dataset* Data Mahasiswa Baru yaitu Nilai Mapel UN, Ranking, Pilihan 1, dan Pilihan 2 dengan atribut Lulus Pilihan sebagai label. Tahapan pertama adalah menghitung probabilitas atribut dengan tipe data *text* yaitu atribut Pilihan 1 dan Pilihan 2. Rumus yang digunakan dalam menghitung probabilitas atribut Pilihan 1 dan Pilihan 2 adalah:

$$P = \frac{\text{frekuensi kelas atribut}}{\text{jumlah data kelas atribut}}$$

Hasil perhitungan probabilitas atribut Pilihan 1 dapat dilihat pada gambar 4.1, dan Pilihan 2 dapat dilihat pada gambar 4.2.

Pilihan 1	Tidak Lulus	pil_1	pil_2	Tidak Lulus	pil_1	pil_2
TEKNIK INFORMATIKA	6	0	0	0.115385	0	0
ILMU POLITIK	0	0	1	0	0	0.2
PENDIDIKAN GEOGRAFI	0	3	0	0	0.230769231	0
KESEHATAN MASYARAKAT	11	0	0	0.211538	0	0
TEKNIK SIPIL	6	0	0	0.115385	0	0
MANAJEMEN	4	2	2	0.076923	0.153846154	0.4
AKUNTANSI	8	0	1	0.153846	0	0.2
EKONOMI PEMBANGUNAN	2	0	0	0.038462	0	0
EKONOMI SYARIAH	2	0	0	0.038462	0	0
PENDIDIKAN EKONOMI	3	0	0	0.057692	0	0
AGRIBISNIS	2	6	0	0.038462	0.461538462	0
AGROTEKNOLOGI	1	0	0	0.019231	0	0
PENDIDIKAN MATEMATIKA	1	2	0	0.019231	0.153846154	0
PENDIDIKAN SEJARAH	2	0	0	0.038462	0	0
TEKNIK ELEKTRO	2	0	0	0.038462	0	0
PENDIDIKAN BAHASA INGGRI	1	0	0	0.019231	0	0
GIZI	1	0	1	0.019231	0	0.2
	52	13	5	1	1	1

Gambar 4.1 Probabilitas Atribut Pilihan 1

Pada gambar 4.1 diketahui bahwa terdapat 52 data pada *class* label Tidak lulus, 13 data pada pil_1, dan 5 pada pil_2 dengan jurusan Kesehatan Masyarakat sebagai jurusan dengan total jumlah terpilih terbesar dengan nilai 11.

Pilihan 2	Tidak Lulus	pil_1	pil_2	Tidak Lulus	pil_1	pil_2
AGRIBISNIS	3	0	2	0.057692	0	0.4
AGROTEKNOLOGI	1	2	0	0.019231	0.153846154	0
AKUNTANSI	1	0	0	0.019231	0	0
EKONOMI PEMBANGUNAN	4	0	0	0.076923	0	0
EKONOMI SYARIAH	2	1	0	0.038462	0.076923077	0
GIZI	2	1	0	0.038462	0.076923077	0
ILMU POLITIK	0	2	0	0	0.153846154	0
KESEHATAN MASYARAKAT	4	0	0	0.076923	0	0
MANAJEMEN	10	0	0	0.192308	0	0
PENDIDIKAN BAHASA DAN SA	2	1	0	0.038462	0.076923077	0
PENDIDIKAN BIOLOGI	1	0	0	0.019231	0	0
PENDIDIKAN EKONOMI	2	0	0	0.038462	0	0
PENDIDIKAN FISIKA	0	1	2	0	0.076923077	0.4
PENDIDIKAN GEOGRAFI	2	0	0	0.038462	0	0
PENDIDIKAN LUAR SEKOLAH	0	1	1	0	0.076923077	0.2
PENDIDIKAN MATEMATIKA	3	0	0	0.057692	0	0
PENDIDIKAN SEJARAH	0	1	0	0	0.076923077	0
TEKNIK ELEKTRO	2	0	0	0.038462	0	0
TEKNIK INFORMATIKA	2	0	0	0.038462	0	0
TEKNIK SIPIL	2	1	0	0.038462	0.076923077	0
Tidak Memilih	9	2	0	0.173077	0.153846154	0
	52	13	5	1	1	1

Gambar 4.2 Probabilitas Atribut Pilihan 2

Pada gambar 4.1 diketahui bahwa terdapat 52 data pada *class* label Tidak lulus, 13 data pada pil_1, dan 5 pada pil_2 dengan jurusan Manajemen sebagai jurusan dengan total jumlah terpilih terbesar dengan nilai 10.

2. Menghitung nilai Mean dan Standar Deviasi

Pada cuplikan data yang digunakan atribut Nilai Mapel UN dan Ranking Sekolah memiliki tipe data numerik sehingga akan dilakukan perhitungan nilai *mean* dan standar deviasi pada atribut tersebut. Untuk menghitung nilai *mean* digunakan persamaan (1). Perhitungan nilai *mean* pada atribut Nilai Mapel UN adalah sebagai berikut:

a) Menghitung jumlah dan probabilitas label

$$\text{Tidak Lulus} = 52 \text{ (} 52/70 \text{)} = 0.742857143$$

$$\text{pil}_1 = 13 \text{ (} 13/70 \text{)} = 0.185714286$$

$$\text{pil}_2 = 5 \text{ (} 5/70 \text{)} = 0.071428571$$

b) Menghitung Nilai Rata-Rata

$$\mu(\text{Tidak Lulus}) = \frac{82.75+81.25+80.58+83.94+81.82+89.23+84.39+84.23+84.2+80.22+82.16+80.34+82.96+84.15+84.12+82.5+82.51+85.28+82.56+81.95+83.27+84.7+81.87+81.58+81.1+80.28+83.81+84.75+82.8+80.86+82.02+83.29+81.63+81.78+82.31+80.93+79.92+85.2+80.11+81.13+84.5+84.73+81.92+84.85+80.7+80.65+83.95+81.38+81.64+82.86+83.44}{52}$$

$$\mu(\text{Tidak Lulus}) = \frac{4215.1}{52}$$

$$\mu(\text{Tidak Lulus}) = 82.70961538$$

$$\mu(\text{pil}_1) = \frac{81.28+87.11+85.45+86.51+81.93+88.78+81.08+84.15+84.91+84.32+85.15+88.98+87.36}{13}$$

$$\mu(\text{pil}_1) = \frac{1107.01}{13}$$

$$\mu(\text{pil}_1) = 85.15461538$$

$$\mu(\text{pil}_2) = \frac{81.03+83.86+80.13+81.8+85.25}{5}$$

$$\mu(\text{pil}_2) = \frac{412.07}{5}$$

$$\mu(\text{pil}_2) = 82.414$$

Untuk menghitung nilai rata-rata atribut Ranking Sekolah rumus dan proses perhitungan yang digunakan sama seperti pada atribut Nilai Mapel UN.

c) Menghitung Standar Deviasi

Tahapan selanjutnya adalah menghitung standar deviasi atribut Nilai Mapel UN dan Ranking Sekolah dengan menggunakan rumus persamaan (2). Berikut adalah proses perhitungan nilai standar deviasi atribut Nilai Mapel UN:

$$\begin{aligned} \sigma(\text{Tidak Lulus}) = & \sqrt{(82.75 - 82.70961538)^2 + (81.25 - 82.70961538)^2 + (80.58 - 82.70961538)^2 + (83.94 - 82.70961538)^2 + (81.82 - 82.70961538)^2 + (89.23 - 82.70961538)^2 + (84.39 - 82.70961538)^2 + (84.23 - 82.70961538)^2 + (84.2 - 82.70961538)^2 + (80.22 - 82.70961538)^2 + (82.16 - 82.70961538)^2 + (80.34 - 82.70961538)^2 + (82.96 - 82.70961538)^2 + (84.15 - 82.70961538)^2 + (84.12 - 82.70961538)^2 + (82.5 - 82.70961538)^2 + (82.51 - 82.70961538)^2 + (85.28 - 82.70961538)^2 + (82.56 - 82.70961538)^2 + (81.95 - 82.70961538)^2 + (83.27 - 82.70961538)^2 + (84.7 - 82.70961538)^2 + (81.87 - 82.70961538)^2 + (81.58 - 82.70961538)^2 + (81.1 - 82.70961538)^2 + (80.28 - 82.70961538)^2 + (83.81 - 82.70961538)^2 + (84.75 - 82.70961538)^2 + (82.8 - 82.70961538)^2 + (80.86 - 82.70961538)^2} \end{aligned}$$

$$\begin{aligned}
& 82.70961538)^2 + (82.02 - 82.70961538)^2 + (83.29 - \\
& 82.70961538)^2 + (81.63 - 82.70961538)^2 + (81.78 - \\
& 82.70961538)^2 + (82.31 - 82.70961538)^2 + (80.93 - \\
& 82.70961538)^2 + (79.92 - 82.70961538)^2 + (85.2 - \\
& 82.70961538)^2 + (80.11 - 82.70961538)^2 + (81.13 - \\
& 82.70961538)^2 + (84.5 - 82.70961538)^2 + (84.73 - \\
& 82.70961538)^2 + (81.92 - 82.70961538)^2 + (84.85 - \\
& 82.70961538)^2 + (80.7 - 82.70961538)^2 + (80.65 - \\
& 82.70961538)^2 + (83.95 - 82.70961538)^2 + (81.38 - \\
& 82.70961538)^2 + (81.64 - 82.70961538)^2 + (82.86 - \\
& 82.70961538)^2 + (83.44 - 82.70961538)^2 / (52 - 1)
\end{aligned}$$

$$\sigma(\text{Tidak Lulus}) = \sqrt{\frac{160.0059173}{51}}$$

$$\sigma(\text{Tidak Lulus}) = \sqrt{3.137370927}$$

$$\sigma(\text{Tidak Lulus}) = 1.771262523$$

$$\sigma(\text{pil}_1) =$$

$$\begin{aligned}
& \sqrt{(81.28 - 85.15461538)^2 + (87.11 - 85.15461538)^2 + \\
& (85.45 - 85.15461538)^2 + (86.51 - 85.15461538)^2 + \\
& (81.93 - 85.15461538)^2 + (88.78 - 85.15461538)^2 + \\
& (81.08 - 85.15461538)^2 + (84.15 - 85.15461538)^2 + \\
& (84.91 - 85.15461538)^2 + (84.32 - 85.15461538)^2 + \\
& (85.15 - 85.15461538)^2 + (88.98 - 85.15461538)^2 + \\
& (86.51 - 85.15461538)^2 / (13 - 1)
\end{aligned}$$

$$\sigma(pil_1) = \sqrt{\frac{79.14086922}{12}}$$

$$\sigma(pil_1) = \sqrt{6.595072435}$$

$$\sigma(pil_1) = 2.568087311$$

$$\sigma(pil_2) = \sqrt{\frac{(81.03 - 82.414)^2 + (83.86 - 82.414)^2 + (80.13 - 82.414)^2 + (81.8 - 82.414)^2 + (85.25 - 82.414)^2}{(5 - 1)}}$$

$$\sigma(pil_2) = \sqrt{\frac{17.64292}{4}}$$

$$\sigma(pil_2) = \sqrt{4.41073}$$

$$\sigma(pil_2) = 2.100173802$$

Untuk menghitung nilai standar deviasi Ranking Sekolah rumus dan proses perhitungan yang digunakan sama seperti pada atribut Nilai Mapel UN.

3. Menghitung Probabilitas Kelas Label

Setelah menghitung nilai *mean* dan standar deviasi atribut dengan tipe data numerik tahapan selanjutnya adalah menghitung probabilitas kelas pada label. Pada *dataset* yang digunakan terdapat 30 jumlah data dengan rincian 23 data kelas Tidak lulus, 5 data kelas Pil_1, dan 2 data kelas Pil_2. Untuk mendapatkan nilai probabilitas setiap kelas pada label akan dibagi dengan total jumlah data pada label. Proses perhitungan nilai probabilitas kelas label adalah sebagai berikut:

Tidak Lulus : $52/70 = 0.742857143$

Pil_1 : $13/70 = 0.185714286$

Pil_2 : $5/70 = 0.071428571$

4. Menghitung Nilai Distribusi *Gaussian*

Langkah selanjutnya adalah menghitung nilai distribusi *Gaussian* untuk *data testing* yang mempunyai atribut dengan tipe nilai numerik, yaitu atribut Nilai Mapel UN dan Ranking Sekolah. *Data testing* yang akan digunakan berjumlah 30 data. Cuplikan *data testing* dapat dilihat pada tabel 4.2.

Tabel 4.2 *Data testing*

Nilai Mapel UN	Ranking Sekolah	Pilihan 1	Pilihan 2
80.86	187	GIZI	MANAJEMEN
83.1	138	MANAJEMEN	AKUNTANSI
85.29	38	PENDIDIKAN EKONOMI	PENDIDIKAN SEJARAH
85.42	236	AGRIBISNIS	AGROTEKNOLOGI
80.48	187	AKUNTANSI	EKONOMI SYARIAH
...
83.71	40	MANAJEMEN	AKUNTANSI
84.16	156	AKUNTANSI	PENDIDIKAN EKONOMI
84.78	113	PENDIDIKAN BIOLOGI	PENDIDIKAN MATEMATIKA
82.1	138	TEKNIK INFORMATIKA	TEKNIK ELEKTRO
82.54	95	MANAJEMEN	AKUNTANSI

Setelah *data testing* didapatkan, perhitungan nilai distribusi *Gaussian* dapat dilakukan, data yang akan digunakan sebagai contoh dalam proses

perhitungan nilai distribusi gaussian adalah data pertama pada *data testing*.

Berikut adalah proses perhitungan nilai distribusi *Gaussian*:

$$P(\text{Nilai Mapel UN} = 80.86 | \text{Tidak Lulus}) = \frac{1}{\sqrt{2\pi \cdot 1.771262523}} * e^{-\frac{(80.86 - 82.70961538)^2}{2 \cdot 3.137370927}}$$

$$P(\text{Nilai Mapel UN} = 80.86 | \text{Tidak Lulus}) = \frac{1}{3.336041166} * 2.718^{0.5452139918}$$

$$P(\text{Nilai Mapel UN} = 80.86 | \text{Tidak Lulus}) = \frac{1}{3.336041166} * 1.724879963$$

$$P(\text{Nilai Mapel UN}) = 0.5170439683$$

Proses perhitungan pada atribut numerik lainnya akan menggunakan proses yang sama seperti perhitungan yang telah dilakukan pada atribut Nilai Mapel UN.

5. Menghitung Probabilitas Akhir setiap Atribut

Tahapan selanjutnya adalah menghitung probabilitas akhir setiap atribut dengan menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut. Berikut adalah proses perhitungan nilai probabilitas akhir setiap atribut:

$$P(X | \text{Tidak Lulus}) = P(\text{Nilai Mapel UN} | \text{Tidak Lulus}) * P(\text{Ranking Sekolah} | \text{Tidak Lulus}) * P(\text{Pilihan 1} | \text{Tidak Lulus}) * P(\text{Pilihan 2} | \text{Tidak Lulus})$$

$$P(X|\text{Tidak Lulus}) = 0.5170439683 * 0.026676122 * 0.019230769 * 0.192307692$$

$$P(X|\text{Tidak Lulus}) = 1.75154E-05$$

$$P(X|\text{Pil}_1) = P(\text{Nilai Mapel UN} | \text{Pil}_1) * P(\text{Ranking Sekolah} | \text{Pil}_1) * P(\text{Pilihan 1} | \text{Pil}_1) * P(\text{Pilihan 2} | \text{Pil}_1)$$

$$P(X|\text{Pil}_1) = 0.06415601 * 0.034319799 * 0 * 0$$

$$P(X|\text{Pil}_1) = 0$$

$$P(X|\text{Pil}_2) = P(\text{Nilai Mapel UN} | \text{Pil}_2) * P(\text{Ranking Sekolah} | \text{Pil}_2) * P(\text{Pilihan 1} | \text{Pil}_2) * P(\text{Pilihan 2} | \text{Pil}_2)$$

$$P(X|\text{Pil}_2) = 0.209412408 * 0.03860444 * 0.2 * 0$$

$$P(X|\text{Pil}_2) = 0$$

Proses perhitungan pada data lainnya akan menggunakan rumus perhitungan yang sama dengan data pertama. Probabilitas akhir setiap data pada atribut dapat dilihat pada gambar 4.3

Nilai Mape	Ranking Se	Pilihan 1	Pilihan 2		Tidak Lulus	pil_1	pil_2
80.86	187	GIZI	MANAJEMEN		1.3E-05	0	0
83.1	138	MANAJEM	AKUNTANSI		1.57E-05	0	0
85.29	38	PENDIDIK/	PENDIDIKAN SEJARAH		0	0	0
85.42	236	AGRIBISNI/	AGROTEKNOLOGI		3.89E-07	5.63E-05	0
80.48	187	AKUNTAN/	EKONOMI SYARIAH	P	1.66E-05	0	0
85	113	KESEHATA	PENDIDIKAN MATEMATIKA	R	6.38E-05	0	0
82.51	187	KESEHATA	PENDIDIKAN BAHASA INGGRIS	O	0	0	0
83.43	23	EKONOMI/	MANAJEMEN	B	2.02E-05	0	0
82.88	40	TEKNIK INI	GIZI	A	2.01E-05	0	0
84.7	23	AGRIBISNI/	TEKNIK SIPIL	B	2.43E-06	2.81E-05	0
83.88	62	PENDIDIK/	PENDIDIKAN EKONOMI	I	4.27E-06	0	0
83.26	259	AGRIBISNI/	Tidak Memilih	L	4.19E-06	2.75E-05	0
86.08	161	AGRIBISNI/	EKONOMI SYARIAH	I	2.45E-06	6.38E-05	0
84.55	238	AGRIBISNI/	KESEHATAN MASYARAKAT	T	2.61E-06	0	0
81.68	93	TEKNIK SIF	Tidak Memilih	A	0.000176	0	0
83.65	12	AGROTEK/	AGRIBISNIS	S	2.07E-06	0	0
80.98	67	PENDIDIK/	MANAJEMEN		1.82E-05	0	0
80.78	187	PENDIDIK/	AGROTEKNOLOGI	A	1.24E-06	9.2E-06	0
86	66	PENDIDIK/	PENDIDIKAN FISIKA	K	0	0	0
77.13	208	PENDIDIK/	AGROTEKNOLOGI	H	0	0	0
88.07	7	PENDIDIK/	Tidak Memilih	I	8.69E-08	0	0
86.08	157	AKUNTAN/	PENDIDIKAN EKONOMI	R	1.02E-05	0	0
81.15	191	EKONOMI/	EKONOMI PEMBANGUNAN		1.11E-05	0	0
82.96	12	KESEHATA	AGROTEKNOLOGI	A	8.57E-06	0	0
87.67	42	AGRIBISNI/	KESEHATAN MASYARAKAT	T	3.74E-07	0	0
83.71	40	MANAJEM	AKUNTANSI	R	5.82E-06	0	0
84.16	156	AKUNTAN/	PENDIDIKAN EKONOMI	I	4.06E-05	0	0
84.78	113	PENDIDIK/	PENDIDIKAN MATEMATIKA	B	0	0	0
82.1	138	TEKNIK INI	TEKNIK ELEKTRO	U	4.55E-05	0	0
82.54	95	MANAJEM	AKUNTANSI	T	1.54E-05	0	0

Gambar 4.3 Probabilitas Akhir Setiap Atribut

6. Probabilitas Akhir

Tahapan selanjutnya adalah menghitung probabilitas akhir data dengan cara mengkalikan probabilitas label dengan probabilitas akhir atribut.

Cuplikan proses perhitungan probabilitas akhir adalah sebagai berikut:

$$P(\text{Tidak Lulus}|X) = P(\text{Tidak Lulus}) * P(X)$$

$$P(\text{Tidak Lulus}|X) = 0.742857143 * 1.30114E-05$$

$$P(\text{Tidak Lulus}|X) = 9.66564E-06$$

$$P(\text{Pil}_1|X) = P(\text{Pil}_1) * P(X)$$

$$P(\text{Pil}_1|X) = 0.185714286 * 0$$

$$P(\text{Pil}_1|X) = 0$$

$$P(\text{Pil}_2|X) = P(\text{Pil}_2) * P(X)$$

$$P(\text{Pil}_2|X) = 0.071428571 * 0$$

$$P(\text{Pil}_2|X) = 0$$

Proses perhitungan pada data lainnya menggunakan rumus perhitungan yang sama seperti data pertama. Hasil probabilitas akhir dapat dilihat pada gambar 4.4.

Nilai Mape	Ranking Se	Pilihan 1	Pilihan 2	P	Tidak Lulus	pil_1	pil_2
80.86	187	GIZI	MANAJEMEN	R	9.67E-06	0	0
83.1	138	MANAJEM	AKUNTANSI	O	1.16E-05	0	0
85.29	38	PENDIDIK/	PENDIDIKAN SEJARAH	B	0	0	0
85.42	236	AGRIBISNI	AGROTEKNOLOGI	A	2.89E-07	1.04E-05	0
80.48	187	AKUNTAN/	EKONOMI SYARIAH	B	1.23E-05	0	0
85	113	KESEHATA	PENDIDIKAN MATEMATIKA	I	4.74E-05	0	0
82.51	187	KESEHATA	PENDIDIKAN BAHASA INGGRIS	L	0	0	0
83.43	23	EKONOMI	MANAJEMEN	I	1.5E-05	0	0
82.88	40	TEKNIK IN	GIZI	T	1.5E-05	0	0
84.7	23	AGRIBISNI	TEKNIK SIPIL	A	1.81E-06	5.22E-06	0
83.88	62	PENDIDIK/	PENDIDIKAN EKONOMI	S	3.17E-06	0	0
83.26	259	AGRIBISNI	Tidak Memilih		3.11E-06	5.11E-06	0
86.08	161	AGRIBISNI	EKONOMI SYARIAH	A	1.82E-06	1.19E-05	0
84.55	238	AGRIBISNI	KESEHATAN MASYARAKAT	K	1.94E-06	0	0
81.68	93	TEKNIK SIF	Tidak Memilih	H	0.000131	0	0
83.65	12	AGROTEK/	AGRIBISNIS	I	1.54E-06	0	0
80.98	67	PENDIDIK/	MANAJEMEN	R	1.35E-05	0	0
80.78	187	PENDIDIK/	AGROTEKNOLOGI		9.24E-07	1.71E-06	0
86	66	PENDIDIK/	PENDIDIKAN FISIKA	D	0	0	0
77.13	208	PENDIDIK/	AGROTEKNOLOGI	A	0	0	0
88.07	7	PENDIDIK/	Tidak Memilih	T	6.46E-08	0	0
86.08	157	AKUNTAN/	PENDIDIKAN EKONOMI	A	7.61E-06	0	0
81.15	191	EKONOMI	EKONOMI PEMBANGUNAN		8.25E-06	0	0
82.96	12	KESEHATA	AGROTEKNOLOGI	T	6.37E-06	0	0
87.67	42	AGRIBISNI	KESEHATAN MASYARAKAT	E	2.78E-07	0	0
83.71	40	MANAJEM	AKUNTANSI	S	4.32E-06	0	0
84.16	156	AKUNTAN/	PENDIDIKAN EKONOMI	T	3.01E-05	0	0
84.78	113	PENDIDIK/	PENDIDIKAN MATEMATIKA	I	0	0	0
82.1	138	TEKNIK IN	TEKNIK ELEKTRO	N	3.38E-05	0	0
82.54	95	MANAJEM	AKUNTANSI	G	1.14E-05	0	0

Gambar 4.4 Probabilitas Akhir Data Testing

4.3 PROSES PENGOLAHAN DATASET MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA TOOLS WEKA

Pada proses ini *tools WEKA* akan digunakan untuk mengolah *dataset* Data Mahasiswa Baru menggunakan algoritma *Naïve Bayes* dengan metode *10-Cross Fold Validation*. Pada *10-Fold Cross Validation* data dibagi menjadi *10 Fold* pada perulangan. Setiap perulangan data dibagi menjadi *9 Fold* untuk *training model*

dan *1 Fold* untuk *testing model*. Hasil yang akan didapat pada proses ini berupa akurasi, presisi, *TP Rate*, *FP Rate*, *Recall*, *F-Measure*, *MCC*, *ROC Area*, *PRC Area*, dan *Confusion Matrix*. Nilai akurasi dan *Confusion Matrix* akan digunakan sebagai evaluasi model *Naïve Bayes* pada *dataset* Data Mahasiswa Baru.

4.3.1 Visualisasi Atribut Menggunakan WEKA

Pada tahap ini akan ditampilkan visualisasi atribut pada *dataset* Data Mahasiswa Baru menggunakan *tools WEKA*. Visualisasi dilakukan untuk lebih memahami *class* yang terdapat pada atribut *dataset*. Adapun beberapa visualisasi atribut yaitu :

1. Visualisasi Atribut Nilai Mapel UN

Atribut Nilai Mapel UN pada *dataset* Data Mahasiswa Baru berisi rentang nilai rata-rata prestasi pendaftar. Nilai prestasi siswa menjadi salah satu faktor yang dapat mempengaruhi penerimaan mahasiswa[47]. Visualisasi atribut Nilai Mapel UN dapat dilihat pada gambar 4.5.

Selected attribute:		
Name: Nilai Mapel UN	Distinct: 1088	Type: Numeric
Missing: 0 (0%)		Unique: 323 (9%)
Statistic	Value	
Minimum	73.75	
Maximum	97.1	
Mean	82.986	
StdDev	2.739	

Gambar 4.5 Visualisasi Atribut Nilai Mapel UN

Pada gambar 4.5 diketahui bahwa atribut Nilai Mapel UN memiliki *class* dalam bentuk numerik dengan nilai terkecil sebesar 73.75, nilai terbesar 97.1, nilai rata-rata sebesar 82.986, dan standar deviasi sebesar 2.739.

2. Visualisasi Atribut Ranking Sekolah

Atribut Ranking Sekolah pada *dataset* Data Mahasiswa Baru berisi nilai yang didapatkan sekolah asal pendaftar berdasarkan nilai total UTBK dengan metode perhitungan 60% nilai TPS (Tes Potensi Skolastik) + 40% nilai TKA (Tes Kompetensi Akademik) [48]. Sebagaimana nilai prestasi siswa, ranking sekolah juga menjadi salah satu faktor yang mempengaruhi penerimaan mahasiswa [47]. Visualisasi atribut Ranking Sekolah dapat dilihat pada gambar 4.6.

Selected attribute		
Name: Ranking Sekolah	Distinct: 272	Type: Numeric
Missing: 0 (0%)		Unique: 29 (1%)
Statistic	Value	
Minimum	1	
Maximum	416	
Mean	132.224	
StdDev	74.858	

Gambar 4.6 Visualisasi Atribut Ranking Sekolah

Pada gambar 4.6 diketahui bahwa atribut Ranking Sekolah Memiliki *class* dalam bentuk numerik dengan nilai terkecil sebesar 1, nilai terbesar 416, nilai rata-rata sebesar 132.224, dan standar deviasi sebesar 74.858.

3. Visualisasi Atribut Pilihan 1

Atribut Pilihan 1 pada *dataset* Data Mahasiswa Baru berisi jurusan utama yang dipilih mahasiswa pendaftar. Pilihan jurusan utama ini memiliki prioritas yang lebih tinggi daripada pilihan kedua. Artinya, jika pendaftar diterima di pilihan pertama, maka pendaftar tidak akan dipertimbangkan lagi untuk pilihan kedua. Dalam melakukan pendaftaran calon mahasiswa

diwajibkan untuk menentukan pilihan utama jurusan yang ingin dituju. Pada *dataset* Data Mahasiswa Baru terdapat 22 jurusan yang telah dipilih. Visualisasi atribut Pilihan 1 dapat dilihat pada gambar 4.7

The screenshot shows a table with the following data:

No.	Label	Count	Weight
1	TEKNIK INFORMATIKA	258	258
2	ILMU POLITIK	103	103
3	PENDIDIKAN GEOGRAFI	86	86
4	KESEHATAN MASYARAKAT	382	382
5	TEKNIK SIPIL	190	190
6	MANAJEMEN	566	566
7	AKUNTANSI	404	404
8	EKONOMI PEMBANGUNAN	146	146
9	EKONOMI SYARIAH	148	148
10	PENDIDIKAN EKONOMI	86	86
11	AGRIBISNIS	160	160

Gambar 4.7 Visualisasi Atribut Pilihan 1

Pada gambar 4.7 diketahui bahwa atribut Pilihan 1 memiliki *class* dengan tipe data nominal adapun penjelasan atribut Pilihan 1 dapat dilihat pada tabel 4.3

Tabel 4.3 Penjelasan Atribut Pilihan 1

Nama Jurusan	Jumlah Pemilih
TEKNIK INFORMATIKA	258
ILMU POLITIK	103
PENDIDIKAN GOGRAFI	86
KESEHATAN MASYARAKAT	382
TEKNIK SIPIL	190
MANAJEMEN	566
AKUNTANSI	404
EKONOMI PEMBANGUNAN	146
EKONOMI SYARIAH	148
PENDIDIKAN EKONOMI	86
AGRIBISNIS	160
AGROTEKNOLOGI	158
PENDIDIKAN MATEMATIKA	197
PENDIDIKAN SEJARAH	69
TEKNIK ELEKTRO	61
PENDIDIKAN BAHASA	142

INGGRIS	
GIZI	85
PENDIDIKAN BIOLOGI	105
PENDIDIKAN FISIKA	40
PENDIDIKAN BAHASA DAN SASTRA INDONESIA	140
PENDIDIKAN LUAR SEKOLAH	21
PENDIDIKAN JASMANI, KESEHATAN DAN REKREASI	40

Pada tabel 4.3 diketahui bahwa pada atribut Pilihan 1 jurusan Manajemen menjadi jurusan yang paling banyak dipilih pendaftar sebagai pilihan utama dengan jumlah pemilih sebanyak 566 sedangkan jurusan Pendidikan Luar Sekolah menjadi jurusan dengan jumlah pemilih paling sedikit yaitu sebanyak 21.

4. Visualisasi Atribut Pilihan 2

Atribut Pilihan 2 pada *dataset* Data Mahasiswa Baru berisi jurusan kedua atau pilihan sekunder yang dipilih mahasiswa pendaftar. Pilihan jurusan kedua ini memiliki prioritas yang lebih rendah daripada pilihan pertama. Artinya, jika pendaftar tidak diterima di pilihan pertama, maka peserta akan dipertimbangkan untuk pilihan kedua. Dalam melakukan pendaftaran calon mahasiswa dapat memilih untuk tidak menentukan pilihan kedua. Pada *dataset* Data Mahasiswa Baru terdapat 22 jurusan yang telah dipilih dengan tambahan opsi untuk tidak memilih jurusan kedua namun dengan menentukan pilihan kedua dengan tepat calon mahasiswa dapat meningkatkan peluang diterima di PTN yang dituju, jika tidak diterima di pilihan pertama. Visualisasi atribut Pilihan 1 dapat dilihat pada gambar 4.8.

Selected attribute				
Name: Pilihan 2		Distinct: 23		Type: Nominal
Missing: 0 (0%)				Unique: 0 (0%)
No.	Label	Count	Count	Weight
1	MANAJEMEN	407	407	407
2	PENDIDIKAN LUAR SEKOLAH	48	48	48
3	AGRIBISNIS	232	232	232
4	Tidak Memilih	618	618	618
5	ILMU POLITIK	102	102	102
6	PENDIDIKAN EKONOMI	130	130	130
7	AKUNTANSI	185	185	185
8	PENDIDIKAN GEOGRAFI	78	78	78
9	TEKNIK TEKNIK SIPIL	92	92	92
10	KESEHATAN MASYARAKAT	183	183	183
11	AGROTEKNOLOGI	171	171	171

Gambar 4.8 Visualisasi Atribut Pilihan 2

Pada gambar 4.8 diketahui bahwa atribut Pilihan 2 memiliki *class* dengan tipe data nominal adapun penjelasan atribut Pilihan 2 dapat dilihat pada tabel 4.4

Tabel 4.4 Penjelasan Pada Atribut Pilihan 2

Nama Jurusan	Jumlah
MANAJEMEN	407
PENDIDIKAN LUAR SEKOLAH	48
AGRIBISNIS	232
TIDAK MEMILIH	618
ILMU POLITIK	102
PENDIDIKAN EKONOMI	130
AKUNTANSI	185
PENDIDIKAN GEOGRAFI	78
TEKNIK SIPIL	92
KESEHATAN MASYARAKAT	183
AGROTEKNOLOGI	171
PENDIDIKAN BAHASA DAN SASTRA INDONESIA	153
PENDIDIKAN MATEMATIKA	100
EKONOMI PEMBANGUNAN	230
PENDIDIKAN SEJARAH	87
PENDIDIKAN FISIKA	48
TEKNIK ELEKTRO	59
PENDIDIKAN BIOLOGI	109
TEKNIK INFORMATIKA	119
GIZI	112
EKONOMI SYARIAH	225
PENDIDIKAN BAHASA INGGRIS	95

Pada tabel 4.4 diketahui bahwa pada atribut Pilihan 2 terdapat 618 calon mahasiswa yang memutuskan untuk tidak memilih jurusan kedua sedangkan jurusan Pendidikan Jasmani, Kesehatan Dan Rekreasi menjadi jurusan dengan jumlah pemilih paling sedikit yaitu sebanyak 4.

4.3.2 Hasil Klasifikasi *Naïve Bayes* Menggunakan *WEKA*

Setelah dilakukan proses pengolahan *dataset* Data Mahasiswa Baru menggunakan algoritma *Naïve Bayes* pada *tools WEKA* didapatkan hasil berupa Akurasi, *TP Rate*, *FP Rate*, Presisi, *Recall*, *F-Measure*, *MCC*, *ROC Area*, *PRC Area*, dan *Confusion Matrix* yang dapat dilihat pada gambar 4.9.

```

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2928          81.6281 %
Incorrectly Classified Instances    655          18.3719 %
Kappa statistic                    0.3112
Mean absolute error                0.1631
Root mean squared error            0.2877
Relative absolute error            72.1881 %
Root relative squared error       86.2254 %
Total Number of Instances         3587

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.955    0.704    0.838    0.955    0.893    0.342    0.857    0.959    Tidak Lulus
0.167    0.004    0.381    0.167    0.232    0.245    0.868    0.256    pil_2
0.296    0.042    0.428    0.296    0.402    0.348    0.870    0.571    pil_1
Weighted Avg.:  0.816    0.566    0.791    0.816    0.789    0.342    0.860    0.874

=== Confusion Matrix ===

  a  b  c  <-- classified as
2714 8 121 |  a = Tidak Lulus
  35 8  1 |  b = pil_2
 495 5 206 |  c = pil_1

```

Gambar 4.9 Hasil Klasifikasi Algoritma *Naïve Bayes* Pada *WEKA*

Berdasarkan gambar 4.9 nilai akurasi dan *Confusion Matrix* akan digunakan sebagai evaluasi model algoritma *Naïve Bayes* pada *dataset* Data Mahasiswa Baru. Berikut adalah hasil perhitungan evaluasi model menggunakan *tool WEKA*:

1. Akurasi

Akurasi adalah tolak ukur yang digunakan untuk mengetahui seberapa tepat suatu pola klasifikasi memprediksi kelas data dari data yang akan datang[39]. Nilai akurasi akan digunakan dalam mengevaluasi algoritma *Naïve Bayes* dalam mengolah *dataset* Data Mahasiswa Baru. Nilai akurasi dapat dilihat pada tabel 4.5.

Tabel 4.5 Hasil Akurasi

Keterangan	Jumlah	Persentasi
Correctly Classified Instances	2928	81.6281 %
Incorrectly Classified Instances	659	18.3719 %

Pada tabel 4.5 diketahui bahwa algoritma *Naïve Bayes* berhasil mengklasifikasikan 2928 data dengan tingkat akurasi sebesar 81.6281% pada *dataset* Data Mahasiswa Baru, dengan jumlah data yang gagal di klasifikasikan sebanyak 659 dengan persentase sebesar 18.3719%

2. Presisi

Presisi adalah metrik evaluasi yang mengukur seberapa baik model membuat prediksi yang benar untuk kelas positif dari total prediksi positif yang dilakukan. Nilai presisi adalah nilai sensitifitas atau nilai ketepatan sistem antara informasi yang diberikan oleh sistem untuk menunjukkan secara benar data kelas negatif atau kelas positif[40]. Nilai presisi akan digunakan dalam mengevaluasi algoritma *Naïve Bayes* dalam mengolah *dataset* Data Mahasiswa Baru. Nilai akurasi dapat dilihat pada tabel 4.6.

Tabel 4.6 Hasil Presisi

Kelas	Jumlah	Rata-Rata
Tidak Lulus	0.838	0.791
Pil_2	0.381	
Pil_1	0.628	

Pada tabel 4.6 diketahui bahwa kelas Tidak Lulus Mendapatkan nilai presisi sebesar 0.838, Kelas Pil_2 Mendapatkan nilai presisi sebesar 0.381, Kelas Pil_1 Mendapatkan nilai presisi sebesar 0.628. Dengan Nilai rata-rata dari semua kelas sebesar 0.791.

3. Recall

Dalam *data mining*, *recall* adalah metrik yang digunakan untuk mengevaluasi kinerja model, terutama dalam tugas klasifikasi. Nilai *recall* adalah nilai yang menunjukkan tingkat keberhasilan atau spesifisitas untuk mengetahui kembali sebuah informasi secara benar tentang data yang kelas negatif atau pun konten teks positif[40]. Nilai Recall akan digunakan dalam mengevaluasi algoritma *Naïve Bayes* dalam mengolah *dataset* Data Mahasiswa Baru. Nilai akurasi dapat dilihat pada tabel 4.7.

Tabel 4.7 Hasil Recall

Kelas	Jumlah	Rata-Rata
Tidak Lulus	0.955	0.816
Pil_2	0.167	
Pil_1	0.296	

Pada tabel 4.7 diketahui bahwa kelas Tidak Lulus Mendapatkan nilai presisi sebesar 0.955, Kelas Pil_2 Mendapatkan nilai presisi sebesar 0.167, Kelas Pil_1 Mendapatkan nilai presisi sebesar 0.296. Dengan Nilai rata-rata dari semua kelas sebesar 0.816.

4. *Confusion Matrix*

Confusion matrix adalah metode yang digunakan untuk mengukur kinerja model klasifikasi berdasarkan perhitungan objek *testing*, dimana data hasil prediksi ada diantara dua kelas yaitu menghasilkan kelas positif dan kelas negative[41]. Pada gambar 4.9 diketahui bahwa label *dataset* Data Mahasiswa Baru memiliki 3 kelas, sehingga dalam menentukan total nilai TP, FP, FN-nya harus dilakukan perhitungan pada nilai TP, FP, FN masing-masing kelas. Berikut adalah proses perhitungan TP, FP, FN pada setiap kelas:

1. TP (*True Positive*)

$$TP = TP(\text{Tidak Lulus}) + TP(\text{Pil}_2) + TP(\text{Pil}_1)$$

$$TP = 2714 + 8 + 206$$

$$TP = 2928$$

2. FP (*False Positive*)

$$FP(\text{Tidak Lulus}) = FP \text{ Kelas Pil}_2 + FP \text{ Kelas Pil}_1$$

$$FP(\text{Tidak Lulus}) = 8 + 121$$

$$FP(\text{Tidak Lulus}) = 129$$

$$FP(\text{Pil}_2) = FP \text{ Kelas Tidak Lulus} + FP \text{ Kelas Pil}_1$$

$$FP(\text{Pil}_2) = 39 + 1$$

$$FP(\text{Pil}_2) = 40$$

$$FP(\text{Pil}_1) = FP \text{ Kelas Tidak Lulus} + FP \text{ Kelas Pil}_2$$

$$FP(\text{Pil}_1) = 485 + 5$$

$$FP(\text{Pil}_1) = 490$$

3. FN (*False Negative*)

$$\text{FN (Tidak Lulus)} = \text{FN Kelas Pil}_2 + \text{FN Pil}_1$$

$$\text{FN (Tidak Lulus)} = 39 + 485$$

$$\text{FN (Tidak Lulus)} = 524$$

$$\text{FN (Pil}_2) = \text{FN Kelas Tidak Lulus} + \text{FN Kelas Pil}_1$$

$$\text{FN (Pil}_2) = 8 + 5$$

$$\text{FN (Pil}_2) = 13$$

$$\text{FN (Pil}_1) = \text{FN Kelas Tidak Lulus} + \text{FN Kelas Pil}_2$$

$$\text{FN (Pil}_1) = 121 + 1$$

$$\text{FN (Pil}_1) = 122$$

Hasil perhitungan TP, FP, dan FN masing-masing kelas dapat dilihat pada tabel 4.8.

Tabel 4.8 Hasil Perhitungan *Confusion Matrix*

Metrik	Penjelasan	Jumlah Total	
TP (<i>True Positive</i>)	Jumlah kelas Tidak Lulus, Pil_2, Pil_1 diklasifikasi dengan benar sebagai Tidak Lulus, Pil_2, Pil_1 oleh model	2928	
FP (<i>False Positive</i>)	Jumlah Kelas yang sebenarnya Tidak Lulus tetapi diklasifikasi bukan sebagai Tidak Lulus. Jumlah Kelas yang sebenarnya Pil_2 tetapi diklasifikasi bukan sebagai Pil_2. Jumlah Kelas yang sebenarnya Pil_1 tetapi diklasifikasi bukan sebagai Pil_1	Tidak Lulus	129
		Pil_2	40
		Pil_1	490
FN (<i>False Negative</i>)	Jumlah Kelas yang bukanlah kelas Tidak Lulus tetapi diklasifikasi sebagai Tidak Lulus. Jumlah Kelas yang bukanlah kelas Pil_2 tetapi diklasifikasi sebagai Pil_2. Jumlah Kelas yang bukanlah kelas Pil_1 tetapi diklasifikasi sebagai Pil_1.	Tidak Lulus	524
		Pil_2	13
		Pil_1	122

Tabel 4.6 menampilkan hasil metrik dari *Confusion Matrix* dan jumlah total data masing-masing kelas pada metrik. Berikut adalah penjelasan metrik pada tabel 4.6:

- 1) *True Positive* (TP): Menunjukkan bahwa model klasifikasi algoritma *Naïve Bayes* memprediksi dengan benar bahwa 2928 data adalah kelas Tidak Lulus.
- 2) *False Positive* (FP): Menunjukkan bahwa model klasifikasi algoritma *Naïve Bayes* memprediksi salah bahwa 129 data adalah kelas Tidak Lulus yang mana sebenarnya bukanlah kelas Tidak Lulus, 40 data adalah kelas Pil_2 yang mana sebenarnya bukanlah kelas Pil_2, dan 490 data adalah kelas Pil_1 yang mana bukanlah merupakan kelas Pil_1.
- 3) *False Negative* (FN): Menunjukkan bahwa model klasifikasi algoritma *Naïve Bayes* dapat memprediksi salah bahwa 524 data bukanlah kelas Tidak Lulus yang mana sebenarnya adalah kelas Tidak Lulus, 13 data bukanlah kelas Pil_2 yang mana sebenarnya adalah kelas Pil_2, 122 data bukanlah kelas Pil_1 yang mana sebenarnya adalah kelas Pil_1.

