

# KLASIFIKASI WAKTU PENYELESAIAN SKRIPSI MAHASISWA MENGUNAKAN METODE WEIGHTED NAÏVE BAYES (STUDI KASUS: PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS MUHAMMADIYAH GRESIK)

Gilang Atala Panhares<sup>1)</sup>, Indra Gita Anugrah<sup>2)</sup>

<sup>1, 2)</sup>Program Studi Teknik Informatika, Universitas Muhammadiyah Gresik  
Jl. Sumatra 101 Gresik Kota Baru (GKB), Randuagung, Kebomas, Gresik  
e-mail: [gilang.panharesi10@gmail.com](mailto:gilang.panharesi10@gmail.com)<sup>1)</sup>, [indragitaanugrah@umg.ac.id](mailto:indragitaanugrah@umg.ac.id)<sup>2)</sup>

## ABSTRAK

*Skripsi merupakan tahapan akhir bagi mahasiswa dalam rangkaian pembelajaran untuk memperoleh gelar sarjana di pendidikan strata-1. Waktu penyelesaian skripsi mulai dari 1 semester, 2 semester, dan lebih dari 2 semester (hingga masa studi mencapai batas maksimalnya). Sangat penting bagi program studi mengetahui klasifikasi waktu penyelesaian skripsi mahasiswa. Informasi tersebut dapat menjadi dasar dalam membuat rencana strategis dalam rangkaian pembelajaran supaya meningkatkan mahasiswa dengan masa studi tepat waktu. Dalam penelitian ini akan dilakukan klasifikasi waktu penyelesaian skripsi mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik dengan menggunakan metode Weighted Naïve Bayes. Dengan menambahkan bobot terhadap atribut kelas pada algoritma naïve bayes, maka ketepatan klasifikasi pada algoritma weighted naïve bayes tidak didasari hanya dengan probabilitas melainkan juga dipengaruhi oleh pembobotan fitur terhadap kelas. Hasil klasifikasi data latih mahasiswa teknik informatika Universitas Muhammadiyah Gresik dalam waktu penyelesaian skripsi didapatkan nilai probabilitas mahasiswa teknik informatika dalam menyelesaikan skripsi dengan waktu 1 semester, 2 semester, dan > 2 semester adalah 0.14286, 0.53968, dan 0.31746. Dan hasil pengujian evaluasi performa dengan menggunakan confusion matrix, didapatkan nilai accuracy 77,5%, precision kelas 1S 100%, precision kelas 2S 72,3%, precision kelas 3S 82,4%, recall kelas 1S 11,1%, recall kelas 2S 100%, recall kelas 3S 93,3%, specificity kelas 1S 100%, specificity kelas 2S 75%, specificity kelas 3S 88%.*

**Kata Kunci:** Klasifikasi, Waktu Penyelesaian Skripsi, Weighted Naïve Bayes.

## ABSTRACT

*This is the final stage for students in a series of learning to obtain a bachelor's degree in strata-1 education. The thesis completion time starts from 1 semester, 2 semesters, and more than 2 semesters (until the study period reaches the maximum limit). It is very important for the study program to know the classification of student thesis completion time. This information can be used as a basis for making strategic plans in a series of lessons in order to improve students with a timely study period. In this study, the classification of the time for completing the thesis of the Informatics Engineering Study Program students of the University of Muhammadiyah Gresik will be carried out using the Weighted Naïve Bayes method. By adding weight to the class attribute in the naïve Bayes algorithm, the classification accuracy in the weighted naïve Bayes algorithm is not only based on probability but is also influenced by the weighting of features in the class. The results of the classification of training data for informatics engineering students at the University of Muhammadiyah Gresik at the time of completing the thesis, the probability values of informatics engineering students in completing the thesis with a time of 1 semester, 2 semesters, and > 2 semesters are 0.14286, 0.53968, and 0.31746. And the results of the performance evaluation test using the confusion matrix, obtained an accuracy value of 77.5%, 1S class precision 100%, 2S class precision 72.3%, 3S class precision 82.4%, 1S class recall 11.1%, class recall 2S 100%, recall class 3S 93.3%, specificity class 1S 100%, specificity class 2S 75%, specificity class 3S 88%.*

**Keywords:** Classification, Thesis Completion Time, Weighted Naïve Bayes.

## I. PENDAHULUAN

Skripsi merupakan tahapan akhir bagi mahasiswa dalam rangkaian pembelajaran untuk memperoleh gelar sarjana di pendidikan strata-1. Mahasiswa menyelesaikan skripsi dengan waktu yang beraneka ragam. Waktu penyelesaian skripsi mulai dari 1 semester, 2 semester, dan lebih dari 2 semester (hingga masa studi mencapai batas maksimalnya). Adapun mahasiswa yang memiliki waktu penyelesaian skripsi di rentang 1-2 semester akan menentukan ketepatan waktu masa studi mahasiswa. Banyaknya mahasiswa yang masa studinya tepat waktu akan berdampak baik terhadap penilaian akreditasi program studi. Sehingga sangat penting bagi program studi mengetahui klasifikasi waktu penyelesaian skripsi mahasiswa. Informasi tersebut dapat menjadi dasar dalam membuat rencana strategis dalam rangkaian

pembelajaran supaya meningkatkan mahasiswa dengan masa studi tepat waktu. Adapun parameter fitur yang digunakan adalah mata kuliah praktikum yang terdiri dari mata kuliah pemrograman terstruktur, mata kuliah pemrograman berorientasi objek, mata kuliah basis data, mata kuliah pemrograman web. Fitur indeks prestasi semestara semester 5 dan 6, mata kuliah pilihan (sesuai topik skripsi), dan semester awal pengambilan skripsi. Dengan label kelas 1 semester, 2 semester, dan > 2 semester. Metode yang digunakan adalah Algoritma Weighted Naïve Bayes. Metode Naïve Bayes terbukti memiliki performa yang sangat baik dalam melakukan klasifikasi[1].Klasifikasi merupakan salah satu metode data mining dengan jenis fungsi pembelajaran supervised. Fungsi pembelajaran supervised digunakan untuk menentukan suatu nilai yang memerlukan class atau label dan dalam proses pembelajarannya di bawah pengawasan guru atau supervisor.

Keuntungan dari penerapan metode ini yaitu pengklasifikasian akan mendapatkan nilai error yang lebih kecil ketika data set berjumlah besar [2]. Selain itu menurut Han & Kamber (2006), klasifikasi Naïve Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar. Pembobotan fitur kelas dapat mempengaruhi hasil klasifikasi [10]. Dengan menambahkan bobot terhadap atribut kelas pada algoritma naïve bayes, maka ketepatan klasifikasi pada algoritma weighted naïve bayes tidak didasari hanya dengan probabilitas melainkan juga dipengaruhi oleh pembobotan fitur terhadap kelas.

## II. LANDASAN TEORI

### 2.1 Klasifikasi Data Mining

Menurut Eko Prasetyo (2013), klasifikasi merupakan suatu teknik pelatihan/pembelajaran suatu objek data kedalam kelas tertentu dari jumlah kelas yang tersedia. Dan kemudian hasil dari pelatihan tersebut menjadi suatu model pembelajaran yang disimpan sebagai data memori. Model klasifikasi memiliki suatu model pembelajaran yang dapat mengambil suatu masukan, kemudian memikirkan masukan tersebut dan memberikan jawaban sebagai keluaran dari pemikiran tersebut [7]. Pada Gambar 2.1 menunjukkan kerangka kerja pada klasifikasi. Pada tahap awal, dibutuhkan sejumlah data latih (x,y) yang digunakan sebagai data pembangun model menggunakan proses algoritma klasifikasi tertentu. Dari hasil model klasifikasi yang terbentuk digunakan sebagai pola untuk memprediksi kelas dari data uji (x,?) sehingga akan diketahui kelas y yang seharusnya.

### 2.2 Algoritma Naïve Bayes

Algoritma *naïve bayes* adalah algoritma klasifikasi berdasarkan pengklasifikasian statistik. Metode ini setiap atribut yang dimiliki pada data set. Klasifikasi *naïve bayes* didasarkan pada *bayes theorem*. Klasifikasi *naïve bayes* juga memiliki kecepatan pemrosesan data yang cepat dengan tingkat akurasi yang tinggi bila diterapkan pada database yang besar. Metode ini mengenali setiap atribut sebagai atribut yang *independent*. Sehingga dengan karakteristik atribut tersebut, metode klasifikasi ini disebut sebagai algoritma yang *naïve*.

### 2.3 Algoritma Weighted Naïve Bayes

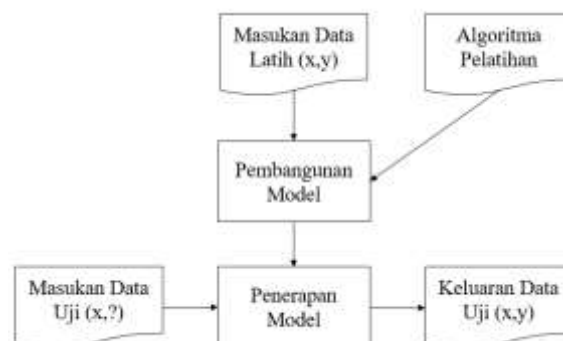
Dengan menambahkan bobot pada setiap atribut dalam data set, maka yang menjadi dasar ketepatan klasifikasi bukan hanya probabilitas tiap atribut melainkan juga dari bobot setiap atributnya.

Berikut ini merupakan langkah-langkah dalam menentukan klasifikasi menggunakan metode Weighted Naïve Bayes:

1. Menghitung nilai probabilitas tiap kelas.

Rumus:

$$P(C_i) = \frac{\sum C_i}{n} \quad (1)$$



**Gambar 2.1** Proses Klasifikasi

Keterangan:

$P(C_i)$  : Probabilitas label kelas  $C_i$

$\sum C_i$  : Jumlah data dengan label kelas  $C_i$

$n$  : Jumlah total data latih

2. Menghitung nilai probabilitas tiap fitur.

$$P(x_k|C_i) = \frac{\sum x_k|C_i}{\sum C_i} \quad (2)$$

Keterangan:

$P(x_k|C_i)$  : Probabilitas fitur  $x_k$  dengan label kelas  $C_i$

$\sum x_k|C_i$  : Jumlah data fitur  $x_k$  dengan label kelas  $C_i$

$\sum C_i$  : Jumlah data dengan label kelas  $C_i$

3. Menghitung nilai probabilitas pada tiap data.

$$P(C_i|X) = P(C_i) \prod_{k=1}^n P(x_k|C_i)^{w_k} \quad (3)$$

Keterangan:

$P(C_i|X)$  : Probabilitas kelas  $C_i$  pada data  $X$

$P(C_i)$  : Probabilitas label kelas  $C_i$

$P(x_k|C_i)$  : Probabilitas fitur  $x_k$  dengan label kelas  $C_i$

$w_k$  : Bobot atribut

#### 2.4 Rank Order Centroid

*Rank Order Centroid* (ROC) merupakan metode pembobotan dengan merujuk pada tingkat kepentingan kriteria. ROC akan membuat model prioritas, bahwa kriteria pertama lebih penting dari pada kriteria kedua, kriteria kedua lebih penting dari pada kriteria ketiga, dan begitu juga seterusnya. Sehingga dapat digambarkan sebagai berikut:

$$C_1 > C_2 > C_3 > C_m \quad (4)$$

Keterangan:

$C_1$  : Kriteria prioritas pertama

$C_2$  : Kriteria prioritas kedua

$C_3$  : Kriteria prioritas ketiga

$C_m$  : Kriteria prioritas ke- $m$

Dan nilai bobot didapatkan dengan rumus perhitungan berikut:

$$W_m = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{i}\right) \quad (5)$$

Keterangan:

$W_m$  : Bobot kriteria

$m$  : Jumlah kriteria

$i$  : Kriteria

#### 2.5 Confusion Matrix

Pengujian pada penelitian ini menggunakan *confusion matrix*. Menurut Han & Kember (2012), pengujian ini digunakan untuk mengukur kemampuan atau sebagai tingkat pengakuan algoritma dalam mengklasifikasi dataset. Dengan demikian parameter benar dan salah dalam pengklasifikasian data uji pada kelas yang berbeda merupakan proses *confusion matrix*.

**Tabel 2.1** Confusion Matrix

Two-Class Prediction	Predicted Class	
	Yes	No
Yes	True Positive	False Negative

<i>Actual Class</i>	<i>No</i>	<i>False Positive</i>	<i>True Negative</i>
---------------------	-----------	-----------------------	----------------------

Jika dalam *actual class* data termasuk dalam kelas yes atau no dan pada saat dilakukan proses prediksi menghasilkan kelas yang sama dengan *actual class* maka disebut *true positive* atau *true negative*. Sedangkan prediksi yang salah dalam klasifikasi yang menghasilkan prediksi yes disebut *false positive* dan prediksi yang salah dalam klasifikasi yang menghasilkan prediksi no disebut *false negative*. Dalam menentukan nilai akurasi tersebut, dapat dihitung menggunakan rumus:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$Specificity = \frac{TN}{TN+FP} \quad (9)$$

Keterangan:

TP : Jumlah data *true positive*

TN : Jumlah data *true negative*

FP : Jumlah data *false positive*

FN : Jumlah data *false negative*

### III. ANALISIS DAN PERANCANGAN SISTEM

#### 3.1 Analisis Sistem

Nilai mata kuliah Pemrograman Terstruktur, Pemrograman Berorientasi Objek, Basis Data dan Pemrograman Web menjadi variabel prediktor pertama yaitu Nilai Mata Kuliah Praktikum. Penentuan mata kuliah tersebut didasari oleh mata kuliah dengan terdapat kewajiban bagi mahasiswa mengikuti praktikum dan mata kuliah tersebut memiliki 4 sks pada kurikulum akademik 2014 yang tercantum pada buku “Manual Akademik Universitas Muhammadiyah Gresik 2018-2019”.

Indeks prestasi semester 5 dan 6 dihitung nilai rata-rata dan menjadi variabel prediktor kedua yaitu Indeks Prestasi Sementara. Indeks prestasi pada semester 5 dan 6 menjadi peranan yang penting dalam rangkaian pembelajaran. Variabel prediktor ketiga dan keempat yaitu nilai mata kuliah pilihan yang sesuai dengan topik skripsi mahasiswa dan semester awal mahasiswa mengambil skripsi pada rencana studi.

#### 3.2 Representasi Data

Data variabel prediktor dan variabel respons didapat dari data kelulusan mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik. Data kelulusan mahasiswa diambil dari tahun lulusan 2019-2021. Data variabel prediktor yang diperoleh dalam penelitian ini berupa data yang berkaitan dengan atribut nilai mata kuliah pemrograman terstruktur, pemrograman berorientasi objek, basis data, pemrograman web dan mata kuliah pilihan yang sesuai dengan topik skripsi mahasiswa. Atribut lainnya adalah semester awal mahasiswa mengambil skripsi, indeks prestasi sementara semester 5 dan semester 6. Data variabel respons waktu penyelesaian skripsi diklasifikasikan menjadi 3 label, yaitu 1 semester, 2 semester dan lebih dari 2 semester. Data tersebut akan dibagi menjadi 4 atribut seperti pada **Tabel 3.1**.

**Tabel 3.1** Atribut Data

No	Atribut	Tipe Atribut
1	Mata Kuliah Praktikum a. Pemrograman Terstruktur b. Pemrograman Berorientasi Objek c. Basis Data d. Pemrograman Web	Kategorial
2	Indeks Prestasi Sementara a. Indeks Prestasi Sementara Semester 5	Numerik

	b. Indeks Prestasi Sementara Semester 6	
3	Mata Kuliah Pilihan	Kategorial
4	Semester Awal Mahasiswa Mengambil Skripsi	Kategorial

Jumlah data yang digunakan sebanyak 103 data. Data tersebut dibagi menjadi 63 data latih dengan pembagian 21 data pada tiap kelas dan 40 data uji. Data latih berfungsi sebagai data pembentuk nilai probabilitas tiap nilai fitur dan tiap label kelas, sedangkan data uji berfungsi sebagai pengujian label kelas prediksi dengan label kelas data aktual.

### 3.3 Preprocessing Data

Data latih dan data uji terlebih dahulu dilakukan proses *preprocessing* data untuk mengubah tipe atribut agar sesuai dengan tipe kategorial atribut Tabel 3.2 untuk Data Latih, Tabel 3.3 untuk Data Uji, Tabel 3.4 Tabel Atribut dan hasil *preprocessing* pada data latih ditunjukkan pada Tabel 3.5.

**Tabel 3.2** Data Latih

No	Mata Kuliah Dasar				IPS 5	IPS 6	MKP	Mulai Skripsi	Waktu Penyelesaian (Semester)
	PTR	PBO	BD	P.WEB					
1	A	B	B	A	3,13	3,65	C	7	1
2	AB	AB	A	AB	3,7	3,8	BC	7	1
3	B	B	B	B	3,43	3,67	A	7	1
4	B	B	BC	C	3,55	3,48	AB	8	1
5	B	B	C	A	3,43	3,93	A	7	1
...	...	...	...	...	...	...	...	...	...
63	BC	B	BC	B	1,98	2,17	AB	10	4

**Tabel 3.3** Data Uji

No	Mata Kuliah Dasar				IPS 5	IPS 6	MKP	Mulai Skripsi	Waktu Penyelesaian (Semester)
	PTR	PBO	BD	P.WEB					
1	BC	B	AB	B	3	1,18	A	10	1
2	C	A	AB	AB	3,63	3,87	AB	7	1
3	A	A	A	B	3,63	3,67	BC	10	1
4	AB	BC	A	B	2,91	3,17	BC	10	1
5	AB	AB	C	B	1,89	3,03	BC	10	1
...	...	...	...	...	...	...	...	...	...
40	BC	B	AB	A	3,48	2,96	AB	9	4

**Tabel 3.4** Tipe Atribut

No	Atribut	Tipe Atribut Awal	Tipe Atribut Setelah Preprocessing
1	Mata Kuliah Praktikum(MK)	Kategorial	Kategorial (Sangat Baik, Baik, Cukup, Kurang, Sangat Kurang)
2	Indeks Prestasi Semester (IPS)	Numerik	Kategorial (Sangat Baik, Baik, Cukup, Kurang, Sangat Kurang)
3	Mata Kuliah Pilihan (MKP)	Kategorial	Kategorial (Sangat Baik, Baik, Cukup, Kurang, Sangat Kurang)
4	Semester Awal Mahasiswa Mengambil Skripsi (SAS)	Kategorial	Kategorial (Sangat Baik, Baik, Kurang, Sangat Kurang)

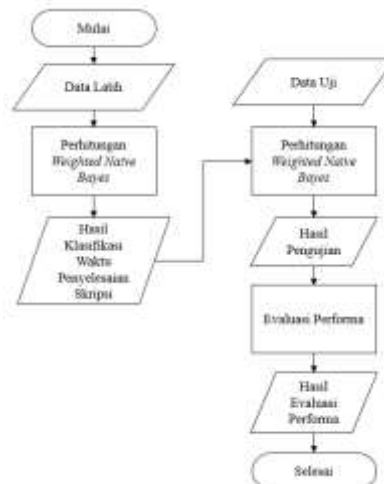
**Tabel 3.5** Hasil *Preprocessing* Data Latih

No	MK	IPS	MKP	SAS	WP
1	SANGAT BAIK	SANGAT BAIK	KURANG	SANGAT BAIK	1S
2	SANGAT BAIK	BAIK	SANGAT KURANG	SANGAT BAIK	1S
3	BAIK	SANGAT BAIK	SANGAT BAIK	SANGAT BAIK	1S
4	KURANG	SANGAT BAIK	BAIK	BAIK	1S
5	BAIK	SANGAT BAIK	SANGAT BAIK	SANGAT BAIK	1S
...	...	...	...	...	...
63	CUKUP	KURANG	BAIK	SANGAT KURANG	3S

**Tabel 3.6** Hasil *Preprocessing* Data Uji

No	MK	IPS	MKP	SAS	WP
1	BAIK	KURANG	SANGAT BAIK	SANGAT KURANG	1S
2	SANGAT BAIK	SANGAT BAIK	BAIK	SANGAT BAIK	1S
3	SANGAT BAIK	SANGAT BAIK	KURANG	SANGAT KURANG	1S
4	SANGAT BAIK	BAIK	KURANG	SANGAT KURANG	1S
5	BAIK	KURANG	KURANG	SANGAT KURANG	1S
...	...	...	...	...	...
40	SANGAT BAIK	BAIK	BAIK	KURANG	3S

### 3.4 Flowchart Sistem



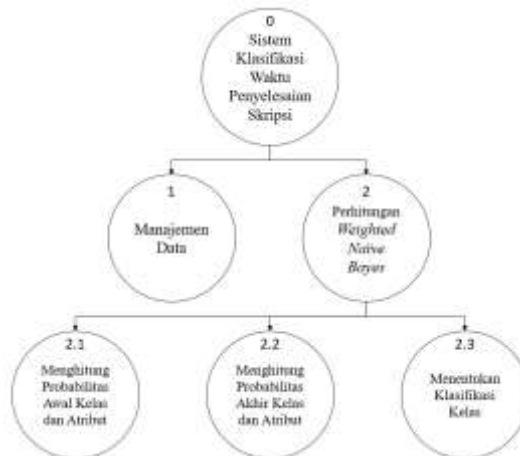
**Gambar 3.1** Flowchart Sistem Klasifikasi

### 3.5 Diagram Konteks



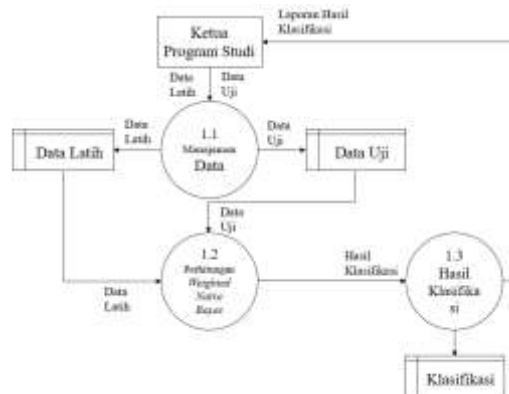
Gambar 3.2 Diagram Konteks

### 3.6 Diagram Berjenjang



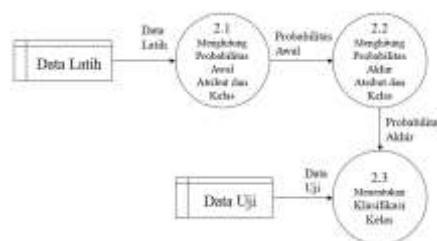
Gambar 3.3 Diagram Berjenjang

### 3.7 Data Flow Diagram Level 0



Gambar 3.4 Data Flow Diagram Level 0

### 3.8 Data Flow Diagram Level 1



Gambar 3.5 Data Flow Diagram Level 1

### 3.9 Perhitungan Weighted Naive Bayes

Berikut ini merupakan langkah-langkah perhitungan untuk memperoleh nilai probabilitas label kelas waktu penyelesaian skripsi pada Hasil *Preprocessing* Data Latih (**Tabel 3.2**):

1. Menghitung nilai probabilitas awal tiap kelas dengan menggunakan **Persamaan (1)**.

Label Kelas 1:

$$P(1S) = \frac{\sum 1S}{n} = \frac{21}{63} = 0,333$$

Label Kelas 2:

$$P(2S) = \frac{\sum 2S}{n} = \frac{21}{63} = 0,333$$

Label Kelas 3:

$$P(3S) = \frac{\sum 3S}{n} = \frac{21}{63} = 0,333$$

2. Menghitung nilai probabilitas awal tiap fitur dengan menggunakan **Persamaan (2)**.

**Tabel 3.7** Probabilitas Fitur MK

$P(x_k C_i)$	WP Class		
	1	2	3
<b>MK</b>			
SANGAT BAIK	0,57143	0,95238	0,47619
BAIK	0,33333	0,04762	0,28571
CUKUP	0	0	0,09524
KURANG	0,09524	0	0,09524
SANGAT KURANG	0	0	0,04762

**Tabel 3.8** Probabilitas Fitur IPS

$P(x_k C_i)$	WP Class		
	1	2	3
<b>IPS</b>			
SANGAT BAIK	0,38095	0,57143	0
BAIK	0,23810	0,42857	0,42857
CUKUP	0,04762	0	0,28571
KURANG	0,14286	0	0,14286
SANGAT KURANG	0,19048	0	0,14286

**Tabel 3.9** Probabilitas Fitur MKP

$P(x_k C_i)$	WP Class		
	1	2	3
<b>MKP</b>			
SANGAT BAIK	0,28571	0,47619	0,14286
BAIK	0,14286	0,2381	0,42857
CUKUP	0,33333	0,2381	0
KURANG	0,09524	0,04762	0,2381
SANGAT KURANG	0,14286	0	0,19048

**Tabel 3.10** Probabilitas Fitur SAS

$P(x_k C_i)$	WP Class		
	1	2	3
<b>SAS</b>			
SANGAT BAIK	0,47619	0,47619	0,04762
BAIK	0,19048	0,52381	0,28571
KURANG	0,04762	0	0,33333
SANGAT KURANG	0,28571	0	0,33333

3. Menghitung sebaran klasifikasi data latih dengan metode *Weighted Naive Bayes* dengan menggunakan **Persamaan (3)**.



Bobot preferensi atribut data didapat dari perhitungan menggunakan metode *Rank Order Centroid* dengan penentuan prioritas atribut oleh Kepala Program Studi Teknik Informatika Universitas Muhammadiyah Gresik. Urutan prioritas atribut sebagai berikut:

C1 : Indeks Prestasi Sementara

C2 : Semester Awal Skripsi

C3 : Mata Kuliah Praktikum

C4 : Mata Kuliah Pilihan

Dari prioritas atribut tersebut, dilakukan penentuan bobot preferensi dengan menggunakan Persamaan (5):

$$W_1 = \frac{1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}}{4} = 0,520833$$

$$W_2 = \frac{0 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}}{4} = 0,270833$$

$$W_3 = \frac{0 + 0 + \frac{1}{3} + \frac{1}{4}}{4} = 0,145833$$

$$W_4 = \frac{0 + 0 + 0 + \frac{1}{4}}{4} = 0,0625$$

Contoh perhitungan pada data latih ke-1:

$$P(1S|X) = 0,333 \times 0,57143^{0,145833} \times 0,38095^{0,520833} \times 0,09524^{0,0625} \times 0,47619^{0,270833}$$

$$P(1S|X) = 0,13123$$

$$P(2S|X) = 0,333 \times 0,95238^{0,145833} \times 0,57143^{0,520833} \times 0,04762^{0,0625} \times 0,47619^{0,270833}$$

$$P(2S|X) = 0,16722$$

$$P(3S|X) = 0,333 \times 0,47619^{0,145833} \times 0^{0,520833} \times 0,2381^{0,0625} \times 0,04762^{0,270833}$$

$$P(3S|X) = 0$$

**Tabel 3.11** Hasil Klasifikasi Data Latih

No	1S	2S	3S	WP
1	0,13123	0,16722	0	2S
2	0,10538	0,15919	0,08001	2S
3	0,12994	0,12476	0	1S
4	0,08087	0	0	1S
5	0,12994	0,12476	0	1S
...	...	...	...	...
63	0	0	0,0357	3S

4. Menghitung probabilitas akhir tiap kelas dengan menggunakan Persamaan (1).

Label Kelas 1:

$$P(1S) = \frac{\sum 1S}{n} = \frac{9}{63} = 0,14286$$

Label Kelas 2:

$$P(2S) = \frac{\sum 2S}{n} = \frac{34}{63} = 0,53968$$

Label Kelas 3:

$$P(3S) = \frac{\sum 3S}{n} = \frac{20}{63} = 0,31746$$

5. Menghitung nilai probabilitas akhir tiap fitur dengan menggunakan Persamaan (2).

**Tabel 3.12** Probabilitas Akhir Fitur MK

$P(x_k C_i)$	WP Class		
MKD	1	2	3
SANGAT BAIK	0,1111	0,97059	0,4
BAIK	0,5556	0,02941	0,4
CUKUP	0	0	0,1
KURANG	0,3333	0	0,05
SANGAT KURANG	0	0	0,05

**Tabel 3.13** Probabilitas Akhir Fitur IPS

$P(x_k C_i)$	WP Class		
IPS	1	2	3
SANGAT BAIK	0,3333	0,5	0
BAIK	0	0,47059	0,35
CUKUP	0	0	0,35
KURANG	0,3333	0	0,15
SANGAT KURANG	0,3333	0,02941	0,15

**Tabel 3.14** Probabilitas Akhir Fitur MKP

$P(x_k C_i)$	WP Class		
MKD	1	2	3
SANGAT BAIK	0,2222	0,35294	0,5
BAIK	0,1111	0,26471	0,35
CUKUP	0,4444	0,23529	0
KURANG	0,1111	0,08824	0,2
SANGAT KURANG	0,1111	0,05882	0,2

### 3.10 Instrument Pengujian

Nilai probabilitas akhir tiap kelas dan tiap fitur menjadi nilai ukur pada pengujian menggunakan algoritma *Weighted Naïve Bayes*. Data uji dilakukan proses pengujian data dengan perhitungan *Weighted Naïve Bayes*.

1. Menghitung Pengujian Sebaran Klasifikasi Data Uji pada tiap kelas dengan algoritma *Weighted Naïve Bayes*.

**Tabel 3.15** Hasil Pengujian Klasifikasi Data Uji

No	1S	2S	3S	WP		AKURASI
				P	A	
1	0	0	0,07398	3S	1S	SALAH
2	0,04293	0,29012	0	2S	1S	SALAH
3	0,04293	0,27087	0	2S	1S	SALAH
4	0	0	0,11343	3S	1S	SALAH
5	0,05429	0	0,04154	1S	1S	BENAR
...	...	...	...	...	...	...
40	0	0	0,11747	3S	3S	BENAR

2. Evaluasi Performa *Confusion Matrix*

**Tabel 3.16** Evaluasi Performa *Confusion Matrix*

Three-Class Prediction		Predicted Class		
		1S	2S	3S
Actual Class	1S	$P_{11}$	$P_{12}$	$P_{13}$
	2S	$P_{21}$	$P_{22}$	$P_{23}$
	3S	$P_{31}$	$P_{32}$	$P_{33}$

True Positive 1S (TP1) =  $P_{11}$

True Positive 2S (TP2) =  $P_{22}$

$$\begin{aligned} \text{True Positive 3S (TP3)} &= P33 \\ \text{True Negative 1S (TN1)} &= P22 + P23 + P32 + P33 \\ \text{True Negative 2S (TN2)} &= P11 + P13 + P31 + P33 \\ \text{True Negative 3S (TN3)} &= P11 + P12 + P21 + P22 \\ \text{False Positive 1S (FP1)} &= P21 + P31 \\ \text{False Positive 2S (FP2)} &= P12 + P32 \\ \text{False Positive 3S (FP3)} &= P13 + P23 \\ \text{False Negative 1S (FN1)} &= P12 + P13 \\ \text{False Negative 2S (FN2)} &= P21 + P23 \\ \text{False Negative 3S (FN3)} &= P31 + P32 \end{aligned}$$

#### IV. ANALISIS DAN PERANCANGAN SISTEM

##### 4.1 Implementasi

Pengembangan sistem klasifikasi waktu penyelesaian skripsi menggunakan software teks editor “Visual Studio Code” dan software web server “XAMPP”. Pengembangan sistem berbasis website dengan bahasa pemrograman HTML, CSS, PHP, dan JavaScript.

###### 1. Halaman *Login*

Pada halaman *login*, *user* terlebih dahulu memasukan *username* dan *password* yang sesuai dengan *database user* pada sistem. Jika *username* dan/atau *password* tidak sesuai maka *user* tidak berhasil memasuki sistem.



Gambar 4.1 Halaman *Login*

###### 2. Halaman Data Latih

Pada halaman data latih, *user* sistem dapat melihat data latih pada tabel. *User* dapat menambahkan data latih baru ke dalam *database* sistem dan juga dapat melakukan aksi ubah dan hapus data latih. Dari halaman data latih tersaji beberapa informasi aktual yang meliputi jumlah data latih dan probabilitas awal tiap kelas.



Gambar 4.2 Halaman Data latih

###### 3. Halaman Klasifikasi Data Latih

Pada halaman klasifikasi data latih termuat hasil klasifikasi dengan perhitungan metode *weighted naïve bayes* pada data latih. Dan dari hasil tersebut diperoleh hasil probabilitas akhir dari tiap kelas dan akan menjadi dasar pengujian dengan menggunakan metode *weighted naïve bayes* pada data uji.

Gambar 4.3 Halaman Klasifikasi Data Latih

4. Halaman Data Uji

Pada halaman data uji, *user* sistem dapat melihat data uji pada tabel. *User* dapat menambahkan data uji baru ke dalam *database* sistem dan juga dapat melakukan aksi ubah dan hapus data uji. Dari halaman data uji tersaji beberapa informasi aktual yang meliputi jumlah data uji dan probabilitas akhir tiap kelas (dari hasil pengklasifikasian metode *weighted naïve bayes* pada data latih).

Gambar 4.4 Halaman Data Uji

5. Halaman Pengujian Data Uji

Pada halaman pengujian data uji termuat hasil klasifikasi dengan perhitungan metode *weighted naïve bayes* pada data uji. Dan dari hasil tersebut diperoleh nilai akurasi dari pengujian data uji terhadap hasil klasifikasi data latih sebelumnya.

Gambar 4.5 Halaman Pengujian Data Uji

6. Halaman Evaluasi Performa

Selain nilai akurasi, pengujian performa metode juga mendapatkan nilai *precision*, *recall*, dan *specificity* pada tiap kelasnya. Nilai evaluasi performa tersebut dapat dilihat melalui button “Evaluasi Performa” pada halaman pengujian data uji (Gambar 4.5).



Gambar 4.6 Halaman Evaluasi Performa

## 4.2 Pengujian Sistem

Pengujian sistem klasifikasi waktu penyelesaian skripsi mahasiswa menggunakan metode *confusion matrix*. Pengujian ini akan menghitung nilai *accuracy*, *precision*, *recall*, dan *specificity* pada data uji.

Tabel 4.1 Hasil Pengujian *Confusion Matrix*

Three-Class Prediction		Predicted Class		
		1S	2S	3S
Actual Class	1S	$P_{11}$ (1)	$P_{12}$ (5)	$P_{13}$ (3)
	2S	$P_{21}$ (0)	$P_{22}$ (16)	$P_{23}$ (0)
	3S	$P_{31}$ (0)	$P_{32}$ (1)	$P_{33}$ (14)

$$\text{True Positive 1S (TP1)} = P_{11} = 1$$

$$\text{True Positive 2S (TP2)} = P_{22} = 16$$

$$\text{True Positive 3S (TP3)} = P_{33} = 14$$

$$\text{True Negative 1S (TN1)} = P_{22} + P_{23} + P_{32} + P_{33} = 16 + 0 + 1 + 14 = 31$$

$$\text{True Negative 2S (TN2)} = P_{11} + P_{13} + P_{31} + P_{33} = 1 + 3 + 0 + 14 = 18$$

$$\text{True Negative 3S (TN3)} = P_{11} + P_{12} + P_{21} + P_{22} = 1 + 5 + 0 + 16 = 22$$

$$\text{False Positive 1S (FP1)} = P_{21} + P_{31} = 0 + 0 = 0$$

$$\text{False Positive 2S (FP2)} = P_{12} + P_{32} = 5 + 1 = 6$$

$$\text{False Positive 3S (FP3)} = P_{13} + P_{23} = 3 + 0 = 3$$

$$\text{False Negative 1S (FN1)} = P_{12} + P_{13} = 5 + 3 = 8$$

$$\text{False Negative 2S (FN2)} = P_{21} + P_{23} = 0 + 0 = 0$$

$$\text{False Negative 3S (FN3)} = P_{31} + P_{32} = 0 + 1 = 1$$

Tabel 4.2 Hasil Evaluasi Performa *Confusion Matrix*

	Precision	Recall	Specificity	Accuracy
1 Semester	1	0,11111	1	0,775
2 Semester	0,72727	1	0,75	
>2 Semester	0,82353	0,93333	0,88	
Rata-rata	0,850267	0,68148	0,87667	

## V. PENUTUP

### 5.1 Kesimpulan

Hasil klasifikasi data latih mahasiswa teknik informatika Universitas Muhammadiyah Gresik dalam waktu penyelesaian skripsi didapatkan nilai probabilitas mahasiswa teknik informatika dalam menyelesaikan skripsi dengan waktu 1 semester, 2 semester, dan > 2 semester adalah 0.14286, 0.53968, dan 0.31746. Dan hasil pengujian evaluasi performa dengan menggunakan *confusion matrix*, didapatkan nilai *accuracy* 77,5%, *precision* kelas 1S 100%, *precision* kelas 2S 72,3%, *precision* kelas 3S 82,4%, *recall* kelas 1S 11,1%, *recall* kelas 2S 100%, *recall* kelas 3S 93,3%, *specificity* kelas 1S 100%, *specificity* kelas 2S 75%, *specificity* kelas 3S 88%.

## 5.2 Saran

Dalam metode *naïve bayes* atau *weighted naïve bayes*, data yang dijadikan sebagai data latih menjadi faktor penting dalam pengklasifikasiannya. Sehingga dalam data latih sangat lebih baik jika probabilitas fitur dan kelasnya tidak terdapat probabilitas yang bernilai 0. Dengan demikian akan meningkatkan performa klasifikasi dan dibutuhkan teknik *laplace smoothing* dalam mengatasi permasalahan probabilitas 0 dalam metode *naïve bayes* atau *weighted naïve bayes*.

## DAFTAR PUSTAKA

- [1] Putra, Nur Afiq Eka, Agustin, Soffi. Klasifikasi Kematangan Tebu berdasarkan tekstur batang menggunakan metode Naïve Bayes. INDEXIA : Informatic and Computational Intelligent Journal. Vol.2. No.2. 2020.
- [2] Berry, I. H. & Browne, M. 2006. *Lecture Notes in Data Mining*. USA: World Scientific.
- [3] Darmono, A & Hasan, A. 2005. *Menyelesaikan Skripsi dalam Satu Semester*. Gramedia Widiasarana Indonesia, Jakarta, Indonesia.
- [4] Han, J & Kamber, M. 2006. *Data Mining Concepts and Techniques, second edition*. California: Morgan Kaufman.
- [5] Han, J., Kamber, M. & Pei, J. 2012. *Data Mining: Concepts and Techniques, 3rd edition*. San Francisco: Morgan Kaufmann.
- [6] Wijayanti, R. R. & Abdurrazyid. 2021. *Perhitungan Estimasi Waktu Pada Produksi Barang Dengan Menerapkan Algoritma Naïve Bayes Klasifikasi (Studi Kasus PT. Hasil Raya Industries)*. Jurnal Informatika. 109-118.
- [7] Prasetyo, E. 2013. *Data Mining – Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- [8] Prasetyo, E. 2014. *Data Mining – Mengelola Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- [9] Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [10] Sari, P. P. 2020. *Analisis Kinerja Algoritma K-Means Dengan Menentukan Centroid Menggunakan Metode Rank Order Centroid (ROC)*. (Tesis, Universitas Sumatera Utara, 2020). Diakses dari <https://repositori.usu.ac.id>
- [11] Siagian, N. A. 2020. *Pembobotan Fitur Dataset Menggunakan Gain Ratio Guna Meningkatkan Akurasi Metode Naïve Bayesian Classifier*. (Tesis, Universitas Sumatera Utara, 2020). Diakses dari <https://repositori.usu.ac.id>
- [12] Wahyuni, D. T. & Luthfiarta, A. 2014. *Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naïve Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi*. Diakses dari <http://eprints.dinus.ac.id>
- [13] Yudha, B. L., Muflikhah, L. & Wihandika, R. C. 2018. *Klasifikasi Risiko Hipertensi Menggunakan Metode Neighbor Weighted KNearest Neighbor (NWKNN)*. Diakses dari <https://j-ptiik.ub.ac.id>