

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1. ANALISIS

Mahasiswa baru adalah faktor penting dalam sebuah universitas, mengingat banyak dan sedikitnya mahasiswa akan sangat berpengaruh dalam peningkatan kualitas sebuah universitas. Di Universitas Muhammadiyah Gresik sendiri setelah penerimaan mahasiswa baru masih banyak mahasiswa yang telat dalam waktu kelulusannya, hal ini dapat mempengaruhi akreditasi dari sebuah jurusan atau pun kampus tersebut, ini bisa terjadi diakibatkan dari sistem penerimaan mahasiswa baru yang kurang selektif dalam menerima mahasiswa baru, dikarenakan penerimaan tersebut tidak diukur sesuai kemampuan atau hasil tes yang didapatkan mahasiswa tersebut, yang pada akhirnya dapat membuat mahasiswa kurang tepat atau tidak sesuai dengan kemampuannya dan mengakibatkan keterlambatan pada kelulusan mahasiswa.

3.2. HASIL ANALISIS

Dari analisis proses seleksi dalam penerimaan mahasiswa baru di Universitas Muhammadiyah Gresik, didapatkan hasil yaitu diperlukan sebuah metode untuk mengatasi masalah calon mahasiswa dalam melihat ketepatan waktu dalam masa kuliah. Salah satu metode yang dapat diterapkan untuk memecahkan masalah ini adalah *data mining*. *Data mining* bisa digunakan untuk mengatasi masalah ini, dikarenakan *data mining* sendiri paling banyak digunakan untuk mengetahui pola dari suatu data, dan sesuai dengan studi kasus yang ada pada Universitas Muhammadiyah Gresik. Maka dari itu perlu adanya sistem yang dapat mengklasifikasikan mahasiswa sesuai program studi berdasarkan hasil dari tes CBT yang mereka kerjakan di Universitas Muhammadiyah Gresik, agar para calon mahasiswa baru dapat lulus dengan tepat waktu. Salah satu metode untuk memprediksi kelulusan calon mahasiswa baru, yaitu menggunakan *data mining* di mana akan dicari pola yang terdapat pada *database* lulusan untuk memprediksi kelulusan calon mahasiswa baru. Data dari penelitian ini merupakan hasil dari

penerimaan mahasiswa baru di Biro Admisi dan Humas Universitas Muhammadiyah Gresik periode 2022/2023, yaitu program studi manajemen tahun 2022/2023. Dikarenakan data dari program studi tersebut cukup untuk menjadi bahan dari penelitian ini, dan dari data tersebut nantinya akan di lakukan perbandingan pengujian dengan berbagai metode yang ada pada *data mining*. Seperti *K-Nearest Neighbor* dan *Naïve Bayes*. Ketiga metode tersebut adalah metode *data mining* yang digunakan untuk mengklasifikasikan data, sehingga akan dilakukan perbandingan hasil dari ketiga metode tersebut untuk mendapatkan metode yang terbaik dalam studi kasus ini.

3.3. REPRESENTASI DATA

Data yang digunakan didapatkan dari biro admisi humas dan biro administrasi akademik, data yang diambil ini adalah data penerimaan mahasiswa baru Universitas Muhammadiyah Gresik tahun 2022 dan data kelulusan dari program studi Teknik informatika, elektro, dan industri, dengan periode gelombang yaitu gelombang 1 sampai 3, dengan jalur tes CBT (*Computer Based Test*).

Data yang digunakan dalam penelitian ini berupa data yang berkaitan dengan nilai tes pada saat calon mahasiswa baru melakukan tes CBT. Data ini digunakan untuk melakukan klasifikasi terhadap calon mahasiswa baru dengan jumlah data 125 yang akan dibagi menjadi 2, yaitu data latih dan data uji, kedua data ini akan diuji dengan 2 metode yang di mana nantinya akan diambil hasil metode terbaik. Data yang digunakan disini ada 8 atribut terdiri dari gelombang pendaftaran, jenis kelamin, nilai bhs Indonesia, nilai bhs Inggris, nilai Matematika, nilai jurusan atau mata pelajaran pilihan, nilai CBT, dan lama studi yang di mana lama studi ini akan menjadi *class* untuk di tentukan tepat waktu atau tidaknya seorang calon mahasiswa tersebut dalam menempuh studinya . Data tersebut bisa dilihat pada tabel berikut :

Tabel 3.1. Atribut Data

No.	Atribut	Tipe Atribut	Keterangan
1	Gelombang	Kategorial	Periode gelombang yang

	<p>{Gelombang 1 = 0}</p> <p>{Gelombang 2 = 1}</p> <p>{Gelombang 3 = 2}</p>		<p>diikuti oleh calon mahasiswa baru di Universitas Muhammadiyah Gresik.</p>
2	<p>Jenis Kelamin</p> <p>{L = 0}</p> <p>{P = 1}</p>	Kategorial	<p>Jenis kelamin diambil untuk mengklasifikasikan rasio jumlah calon mahasiswa laki – laki dan perempuan yang akan di terima di jurusan pilihanya.</p>
3	<p>Nilai bahasa Indonesia</p>	Numerik	<p>Nilai bahasa Indonesia ini di ambil dari nilai ujian nasional yang nantinya di gunakan sebagai atribut dalam penelitian ini.</p>
4	<p>Nilai bahasa Inggris</p>	Numerik	<p>Nilai bahasa Inggris ini di ambil dari nilai ujian nasional yang nantinya di gunakan sebagai atribut dalam penelitian ini.</p>
5	<p>Nilai matematika</p>	Numerik	<p>Nilai bahasa matematika ini di ambil dari nilai ujian nasional yang nantinya di gunakan sebagai atribut dalam penelitian ini.</p>
6	<p>Nilai kompetensi keahlian atau mata pelajaran pilihan</p>	Numerik	<p>Nilai bahasa Indonesia ini di ambil dari nilai ujian nasional yang nantinya di</p>

			gunakan sebagai atribut dalam penelitian ini.
7	Nilai Tes CBT (<i>Computer Based Test</i>)	Numerik	Nilai tes CBT ini diambil dari tes calon mahasiswa baru pada saat melakukan tes CBT, baik dari gelombang 1 sampai 3, yang nantinya akan di gunakan untuk mengklasifikasikan antara lulus tidaknya calon mahasiswa baru.
8	Lama Studi	Kategorial	Lama Studi ini akan di gunakan sebagai class yang akan menentukan lama atau tidaknya seorang calon mahasiswa dalam menempuh studinya.

Setelah mengetahui beberapa atribut yang akan di gunakan pada penelitian ini, berikut data latih dan data uji yang akan di lakukan pada beberapa metode. Keterangan pada data ini akan menjadi 2 kelas yaitu LULUS TEPAT WAKTU dan LULUS TIDAK TEPAT WAKTU.

Tabel 3.2.Data Latih

No.	Gelombang	Jenis Kelamin	Kategorial Nilai Bhs Indonesia	Kategorial Nilai Bhs Inggris	Kategorial Nilai Matematika	Kategorial Nilai jurusan / mapel pilihan	Kategorial Nilai Cbt	Keterangan
1	Gelombang I	L	90	68	66	88	46	TIDAK TEPAT WAKTU
2	Gelombang II	L	74	66	52	72	47	TIDAK TEPAT WAKTU
3	Gelombang I	L	98	80	78	96	42	TEPAT WAKTU
4	Gelombang I	L	85	66,5	64	82,5	42	TIDAK TEPAT WAKTU
5	Gelombang II	P	45	26,5	24	34,5	42	TIDAK TEPAT WAKTU
6	Gelombang II	L	88,5	70	67,5	86	47	TIDAK TEPAT WAKTU
7	Gelombang II	L	90	72	70	80	25	TIDAK TEPAT WAKTU
8	Gelombang I	L	86	68	66	84	59	TIDAK TEPAT WAKTU
9	Gelombang III	L	47,5	28	26	45,5	38	TEPAT WAKTU
10	Gelombang II	P	90	68	66	88	46	TIDAK TEPAT WAKTU
...
116	Gelombang II	L	88	73	68	83	53	TIDAK TEPAT WAKTU

Tabel 3.3.Data Uji

No.	Gelombang	Jenis Kelamin	Kategorial Nilai Bhs Indonesia	Kategorial Nilai Bhs Inggris	Kategorial Nilai Matematika	Kategorial Nilai jurusan / mapel pilihan	Kategorial Nilai Cbt	Keterangan
33	Gelombang II	L	94	84	78	88	32	TEPAT WAKTU
34	Gelombang II	L	97	85	77	89	26	TEPAT WAKTU
35	Gelombang II	L	60	40	22,5	52,5	34	TIDAK TEPAT WAKTU
36	Gelombang I	L	71,6	58	50	63,6	47	TIDAK TEPAT WAKTU
37	Gelombang I	L	84,6	71	63	76,6	54	TIDAK TEPAT WAKTU
38	Gelombang I	L	80	67,2	60	72,8	24	TIDAK TEPAT WAKTU
39	Gelombang I	L	70,8	58	50,8	63,3	68	TIDAK TEPAT WAKTU
40	Gelombang II	L	70	42	22,5	40	41	TIDAK TEPAT WAKTU
76	Gelombang II	L	94	84	78	88	32	TEPAT WAKTU
77	Gelombang II	L	97	85	77	89	26	TIDAK TEPAT WAKTU
.....
125	Gelombang II	L	74	56	46	64	46	TIDAK TEPAT WAKTU

3.3.1. Perhitungan Metode *K-Nearest Neighbor*

Berikut ini merupakan langkah-langkah perhitungan untuk memperoleh nilai jarak *Euclidean* label kelas kelulusan pada Data uji terhadap Data latih, Kecuali data atribut gelombang dan jenis kelamin di ubah menjadi numerikal, 0 = gelombang 1, 1 = gelombang 2, 2 = gelombang 3, dan untuk atribut jenis kelamin yaitu di ubah menjadi numerikal, 0 = laki – laki, dan 1 = perempuan.

Menentukan nilai dari $K = 3, 5$

1. Menghitung jarak *Euclidean* masing-masing data uji terhadap data latih menggunakan persamaan (2.7).

Contoh data latih ke -1 :

Tabel 3.4. Data Latih 1

No.	Gelombang	Jenis Kelamin	Kategorial Nilai Bhs Indonesia	Kategorial Nilai Bhs Inggris	Kategorial Nilai Matematika	Kategorial Nilai jurusan / mapel pilihan	Kategorial Nilai CBT	Keterangan
33	Gelombang	L	94	84	78	88	32	TEPAT WAKTU

Perhitungan data testing 1 terhadap data training 100

$$d_{y_1}d_{x_{116}} = \sqrt{\left(\begin{array}{l} \text{Gelombang 1 (0)} \\ -\text{Gelombang 2 (1)} \end{array} \right)^2 + \left(\begin{array}{l} \text{Jenis Kelamin l (0)} \\ \text{Jenis Kelamin l (0)} \end{array} \right)^2 + (94 - 88)^2 + (84 - 73)^2 + (78 - 68)^2 + (88 - 83)^2 + (32 - 53)^2} = 175,34252194$$

Tabel 3.5 Jarak *Euclidean* Data testing 1 terhadap Data *Training* 1

No.	Jarak Euclidean
1	70,57265759
2	47,88005848
3	89,89160139
4	63,35219018
5	28,27543103
6	70,91896784
7	70,84137209
8	70,49468065
9	19,53202498
...
116	72,3567550405627

2. Mengurutkan jarak euclidean dari yang terkecil

Tabel 3.6 Jarak *Euclidean* Data testing 1 terhadap Data *training* 1 setelah di urutkan ke yang terkecil

No.	Jarak Euclidean
74	12,60952021
55	17,27715254
9	19,53202498
19	19,76107285

95	21,2186239
54	23,43608329
61	24,88473428
24	25,37676102
48	26,56953142
....
108	114,6801639

3. Menentukan label kelas hasil kategori *nearest neighbor* dari label mayoritas terbanyak.

Tabel 3.7. Penentuan Label Kelas

No.	Jarak Euclidean	K = 3	K = 5
52	5,291502622	TIDAK TEPAT WAKTU	TIDAK TEPAT WAKTU
94	6,244997998	TIDAK TEPAT WAKTU	TIDAK TEPAT WAKTU
59	6,32455532	TIDAK TEPAT WAKTU	TIDAK TEPAT WAKTU
93	6,708203932		TIDAK TEPAT WAKTU
46	8,888194417		TEPAT WAKTU

Hasil dari perhitungan kategori *nearest neighbor* K=3 adalah tidak tepat waktu dari hasil mayoritas tersebut. Sedangkan untuk hasil dari K= 5 di dapatkan hasil dengan mayoritas tidak tepat waktu.

3.3.2. Perhitungan Metode *Naive Bayes*

1. Menghitung nilai probabilitas kelas menggunakan persamaan (2.1)

Tabel 3.8. Probabilitas Kelas

	TEPAT WAKTU	TIDAK TEPAT WAKTU
Keterangan	38%	62%

2. Menghitung probabilitas fitur menggunakan rumus (2.2)

- a. Fitur gelombang

Tabel 3.9. Probabilitas Fitur Gelombang

Gelombang	TEPAT WAKTU	TIDAK TEPAT WAKTU
-----------	-------------	-------------------

Gelombang I	53%	50%
Gelombang II	42%	45%
Gelombang III	5%	5%

b. Fitur jenis kelamin

Tabel 3.10. Tabel Fitur Jenis Kelamin

Jenis Kelamin	TEPAT WAKTU	TIDAK TEPAT WAKTU
L	82%	90%
P	18%	10%

3. Menghitung Mean data numerikal dengan menggunakan rumus

(2.4)

Tabel 3.11 Mean dari data Numerikal

	TEPAT WAKTU	TIDAK TEPAT WAKTU
Nilai bhs Indonesia	87,05789474	81,46193548
Nilai bhs Inggris	75,35789474	69,00080645
Nilai Matematika	66,61842105	61,32096774
Nilai Jurusan / mapel pilihan	78,22105263	73,83854839
Nilai Tes	36,26315789	40,66129032

4. Menghitung standar *devisiasi* data *numerikal* menggunakan rumus (2.5)

Tabel 3.12 Standar devisiasi dari data Numerikal

	TEPAT WAKTU	TIDAK TEPAT WAKTU
Nilai bhs Indonesia	12,46444005	12,81045153
Nilai bhs Inggris	16,22790709	15,85773972
Nilai Matematika	16,56508629	17,73720862
Nilai Jurusan / mapel pilihan	13,36128247	15,681431
Nilai Tes	10,94626732	11,98488604

5. Menghitung normal *distribution* data numerikal menggunakan rumus (2.6)

Tabel 3.13 Data testing ke 1

No.	Gelombang	Jenis Kelamin	Nilai Bhs Indonesia	Nilai Bhs Inggris	Nilai Matematika	Nilai jurusan / mapel pilihan	Nilai Cbt	Keterangan
33	Gelombang II	L	94	84	78	88	32	TEPAT WAKTU

P(Nilai Bhs Indonesia

$$= 94 | \text{keterangan} = \text{tepat waktu} = (1 / (12,46444005$$

$$* (2.506))) * (2.718)^{-((94 - 87,05789474)^2 / (2 * 12,46444005^2))} = 0,027408104$$

P(Nilai Bhs Indonesia

$$= 94 | \text{keterangan} = \text{tidak tepat waktu} = (1 / (15,85773972$$

$$* (2.506))) * (2.718)^{-((94 - 81,46193548)^2 / (2 * 15,85773972^2))} = 0,019290104$$

P(Nilai Bhs Inggris

$$= 84 | \text{keterangan} = \text{tepat waktu} = (1 / (16,22790709$$

$$* (2.506))) * (2.718)^{-((84 - 75,35789474)^2 / (2 * 16,22790709^2))} = 0,021333567$$

P(Nilai Bhs Inggris

$$= 84 | \text{keterangan} = \text{tidak tepat waktu} = (1 / (15,85773972$$

$$* (2.506))) * (2.718)^{-((84 - 69,00080645)^2 / (2 * 15,85773972^2))} = 0,016084144$$

P(Nilai Matematika

$$= 78 | \text{keterangan} = \text{tepat waktu} = (1 / (16,56508629$$

$$* (2.506))) * (2.718)^{-((78 - 66,61842105)^2 / (2 * 16,56508629^2))} = 0,019019759$$

P(Nilai Matematika

$$= 78 | \text{keterangan} = \text{tidak tepat waktu} = (1 / (17,73720862$$

$$* (2.506))) * (2.718)^{-((78 - 61,32096774)^2 / (2 * 17,73720862^2))} = 0,014454866$$

P(Nilai Jurusan / mapel pilihan

$$= 88 | \text{keterangan} = \text{tepat waktu} = (1 / (13,36128247$$

$$* (2.506))) * (2.718)^{-((88 - 78,22105263)^2 / (2 * 13,36128247^2))} = 0,022842581$$

P(Nilai Jurusan / mapel pilihan

= 88|keterangan = tidak tepat waktu) = $(1/(15,681431$

$*(2.506))) * (2.718)^{-((88 - 73,83854839)^2 / (2 * 15,681431^2))} = 0,020945016$

P(Nilai cbt

= 32|keterangan = tepat waktu) = $(1/(10,94626732$

$*(2.506))) * (2.718)^{-((32 - 36,26315789)^2 / (2 * 10,94626732^2))} = 0,033783679$

P(Nilai cbt

= 32|keterangan = tidak tepat waktu) = $(1/(11,98488604$

$*(2.506))) * (2.718)^{-((32 - 40,66129032)^2 / (2 * 11,98488604^2))} = 0,025636933$

6. Menghitung nilai probabilitas akhir masing masing data uji menggunakan persamaan (2.3)

Contoh data uji ke 1

Tabel 3.14. Data Uji Ke 1

No.	Gelombang	Jenis Kelamin	Nilai Bhs Indonesia	Nilai Bhs Inggris	Nilai Matematika	Nilai jurusan / mapel pilihan	Nilai Cbt	Keterangan
33	Gelombang II	L	94	84	78	88	32	TEPAT WAKTU

$p(33|tepat waktu)$

= $p(\text{gelombang} = \text{gelombang 2})$

* $p(\text{jenis kelamin} = L)$

* $p(\text{nilai bahasa indonesia} = \text{sangat baik})$

* $p(\text{nilai bahasa inggris} = \text{sangat baik})$

* $p(\text{nilai matematika}$

= baik)

* $p(\text{nilai jurusan mata pelajaran pilihan}$

= sangat baik) * $p(\text{nilai cbt} = \text{baik})$

* $p(\text{tepat waktu})$

$p(33|tepat waktu)$

= $42,1053\% * 81,5789 * \%2,7408 * \%2,1334 * \%1,9020$

$* \%2,2843 * \%3,3784\% * 68\% = 0,0000001120205444\%$

$$p(33|tidak\ tepat\ waktu) = p(gelombang = gelombang\ 2) * p(jenis\ kelamin = L) * p(nilai\ bahasa\ indonesia = sangat\ baik) * p(nilai\ bahasa\ inggris = sangat\ baik) * p(nilai\ matematika = baik) * p(nilai\ jurusan\ mata\ pelajaran\ pilihan = sangat\ baik) * p(nilai\ cbt = baik) * p(tidak\ tepat\ waktu)$$

$$p(33|tidak\ tepat\ waktu) = 45\% * 90\% * 2\% * 2\% * 1\% * 2\% * 3\% * 21\% = 0,000000060904269\%$$

Hasil dari perhitungan probabilitas data uji dapat dilihat pada tabel berikut :

Tabel 3.15. Hasil Data Uji

No	Probabilitas Atribut							Probabilitas Kesimpulan
	Gelombang	Jenis Kelamin	Kategorial Nilai Bhs Indonesia	Kategorial Nilai Bhs Inggris	Kategorial Nilai Matematika	Kategorial Nilai jurusan / mapel pilihan	Kategorial Nilai Cbt	TEPAT WAKTU
33	42%	82%	3%	2%	2%	2%	3%	0,0000001120205444%
34	42%	82%	2%	2%	2%	2%	2%	0,0000000627607211%
35	42%	82%	0%	0%	0%	0%	4%	0,0000000000105091%
36	53%	82%	1%	1%	1%	2%	2%	0,0000000180722389%
37	53%	82%	3%	2%	2%	3%	1%	0,0000000830086186%
38	53%	82%	3%	2%	2%	3%	2%	0,0000001146746381%
39	53%	82%	1%	1%	2%	2%	0%	0,0000000004119620%
40	42%	82%	1%	0%	0%	0%	3%	0,000000000055964%
76	42%	82%	3%	2%	2%	3%	3%	0,0000001889187676%
77	42%	82%	3%	2%	2%	3%	1%	0,0000000447379245%
....
125	42%	82%	2%	1%	1%	2%	2%	0,0000000134423219%

Tabel 3.16. Hasil Data Uji

No	Probabilitas Atribut							Probabilitas Kesimpulan
	Gelombang	Jenis Kelamin	Nilai Bhs Indonesia	Nilai Bhs Inggris	Nilai Matematika	Nilai jurusan / mapel pilihan	Nilai Cbt	TIDAK TEPAT WAKTU
33	45%	90%	2%	2%	1%	2%	3%	0,000000060904269%
34	45%	90%	1%	2%	2%	2%	2%	0,000000027482666%

35	45%	90%	1%	0%	0%	1%	3%	0,0000000035446 2%
36	50%	90%	2%	2%	2%	2%	3%	0,00000011215777 5%
37	50%	90%	3%	2%	2%	3%	2%	0,00000021442413 2%
38	50%	90%	3%	2%	2%	2%	1%	0,00000014741724 9%
39	50%	90%	2%	2%	2%	2%	0%	0,00000000918699 2%
40	45%	90%	2%	1%	0%	0%	3%	0,0000000027736 2%
76	45%	90%	3%	2%	2%	2%	3%	0,00000023797430 3%
77	45%	90%	3%	2%	2%	3%	2%	0,00000018959332 2%
....
125	45%	90%	3%	2%	2%	2%	3%	0,00000009407532 2%

Tabel 3.17 Tabel Prediksi

No	PREDICTED CLASS		
	PREDIKSI TEPAT WAKTU	PREDIKSI TIDAK TEPAT WAKTU	PREDIKSI CLASS
33	0,000000060904269%	0,0000001120205444%	TEPAT WAKTU
34	0,000000027482666%	0,0000000627607211%	TEPAT WAKTU
35	0,00000000354462%	0,000000000105091%	TIDAK TEPAT WAKTU
36	0,000000112157775%	0,0000000180722389%	TIDAK TEPAT WAKTU
37	0,000000214424132%	0,0000000830086186%	TIDAK TEPAT WAKTU
38	0,000000147417249%	0,0000001146746381%	TIDAK TEPAT WAKTU
39	0,000000009186992%	0,0000000004119620%	TIDAK TEPAT WAKTU
40	0,00000000277362%	0,000000000055964%	TIDAK TEPAT WAKTU
76	0,000000237974303%	0,0000001889187676%	TIDAK TEPAT WAKTU
77	0,000000189593322%	0,0000000447379245%	TIDAK TEPAT WAKTU
....
125	0,000000094075322%	0,0000000134423219%	TIDAK TEPAT WAKTU

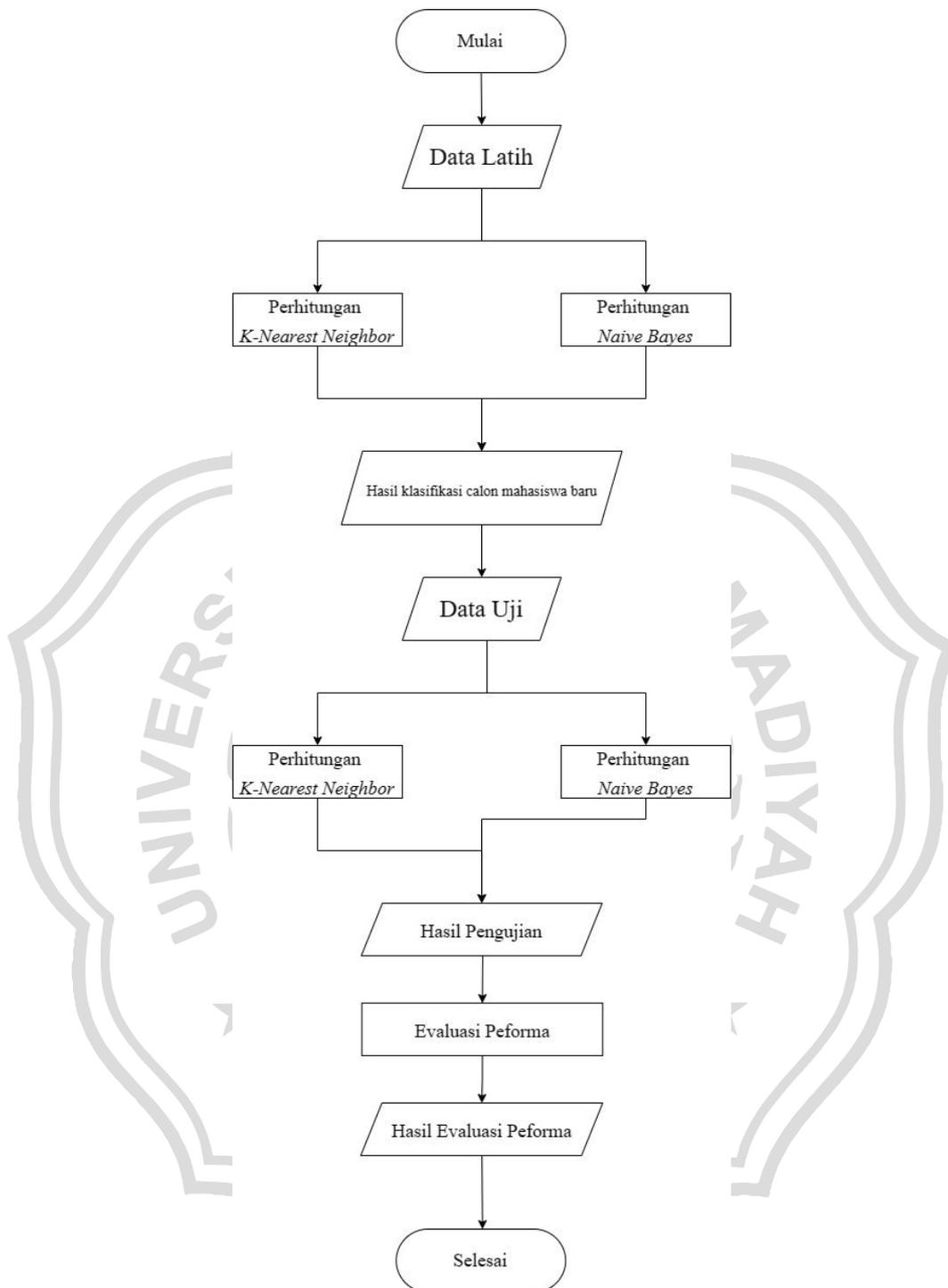
3.4. PERANCANGAN SISTEM

Dari hasil analisis sistem dan perhitungan metode, dilanjutkan dengan proses perancangan sistem perbandingan metode untuk mendapatkan klasifikasi metode yang terbaik pada penerimaan calon mahasiswa baru di Universitas Muhammadiyah Gresik. Perancangan sistem dimuat dalam bentuk *flowchart*, diagram konteks, diagram berjenjang, dan data *flow diagram* seperti pada **Gambar 3.1. Flowchart Sistem.**

3.4.1. Flowchart

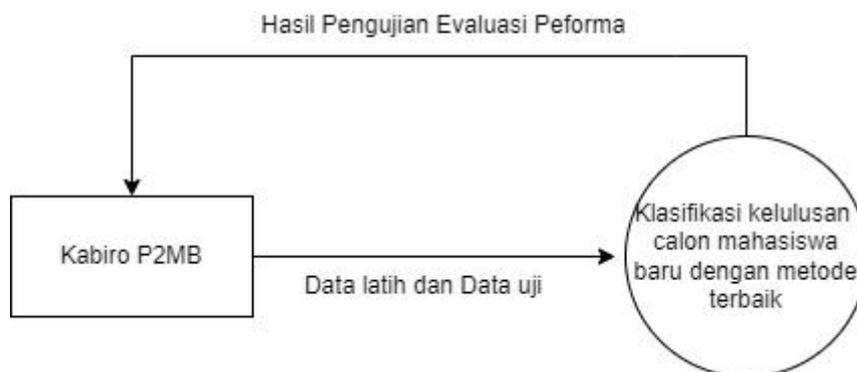
Pada sistem kelulusan calon mahasiswa baru ini dimulai dengan memberikan *input* pada data latih yang sebelumnya sudah dilakukan proses *preprocessing*. Dari data latih akan dihitung probabilitas tiap kelas dengan memilih metode yang akan digunakan sebagai penentuan kelulusan dari calon mahasiswa baru. Nilai dari probabilitas tersebut akan digunakan dalam perhitungan metode yang dipilih. Yaitu pada perhitungan metode *K-Nearest Neighbor* dan *Naïve Bayes*.

Setelah klasifikasi didapatkan dilanjutkan proses selanjutnya adalah melakukan pengujian dari hasil klasifikasi. Data yang digunakan pada proses ini adalah data uji yang sudah dilakukan *preprocessing*. Data uji kemudian dihitung dengan 2 metode yang di pilih pada perhitungan yang ada pada sistem ini, dengan fitur dari klasifikasi data latih. Hasil dari pengujian ini akan dilakukan evaluasi performa menggunakan *confusion matrix*. Evaluasi peforma ini bertujuan untuk mengukur keefektifan dari hasil klasifikasi yang telah didapatkan menggunakan 2 metode yang ada pada sistem ini.



Gambar 3.1. *Flowchart Sistem*

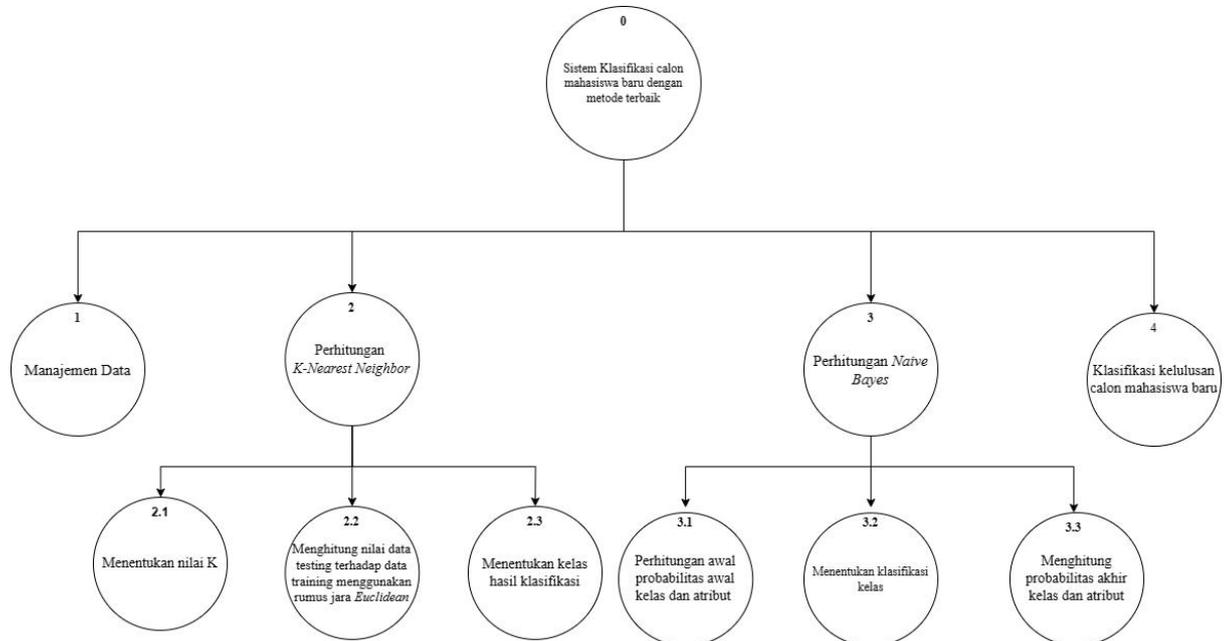
3.4.2. Diagram Konteks



Gambar 3.2. Diagram Konteks

Pada diagram konteks yang ada pada **Gambar 3.2.** Diagram Konteks sistem klasifikasi kelulusan calon mahasiswa, disini Kepala program studi merupakan entitas Kepala program studi memasukkan data latih sebagai *data training* yang akan dihitung dengan ketiga metode yang ada pada sistem ini. Di sini Kepala program studi bisa memilih metode yang sekiranya cocok dengan data yang digunakan, bisa menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*. Pada metode *Naive bayes* data latih digunakan sebagai data *training* yang akan dihitung probabilitas awal tiap kelas, dan data uji sendiri digunakan sebagai data pengujian terhadap hasil klasifikasi data latih yang telah ditetapkan setelah mendapatkan probabilitas awal. Sedangkan pada perhitungan *K-Nearest Neighbor* *Data training* digunakan sebagai data yang akan dihitung, data testing digunakan sebagai data pengujian yang akan dihitung terhadap *data training* untuk mendapatkan hasil nilai jarak antar data menggunakan rumus jarak *Euclidean*. Setelah mendapatkan nilai data dilakukan proses pengurutan peringkat jarak antar data dari jarak terkecil dilakukan pengelompokan hasil klasifikasi. Kemudian menentukan label hasil klasifikasi berdasarkan label mayoritas terbanyak. Setelah Kepala program studi menentukan metode maka akan di dapatkan hasil klasifikasi dari data calon mahasiswa yang akan di kembalikan ke Kepala program studi.

3.4.3. Diagram Berjenjang



Gambar 3.3. Diagram Berjenjang

Top Level : 0 Sistem Klasifikasi calon mahasiswa baru dengan metode terbaik seperti yang ada pada **Gambar 3.3.** Diagram Berjenjang

Level 0 : 1 Manajemen Data

2 Perhitungan menggunakan metode *k-Nearest Neighbor*

3 Perhitungan menggunakan metode *Naive Bayes*

4 Klasifikasi kelulusan calon mahasiswa baru

Level 1 : 2.1 Menentukan nilai K

2.2 Menentukan nilai data menggunakan jarak *Euclidean*

2.3 Menentukan kelas hasil klasifikasi

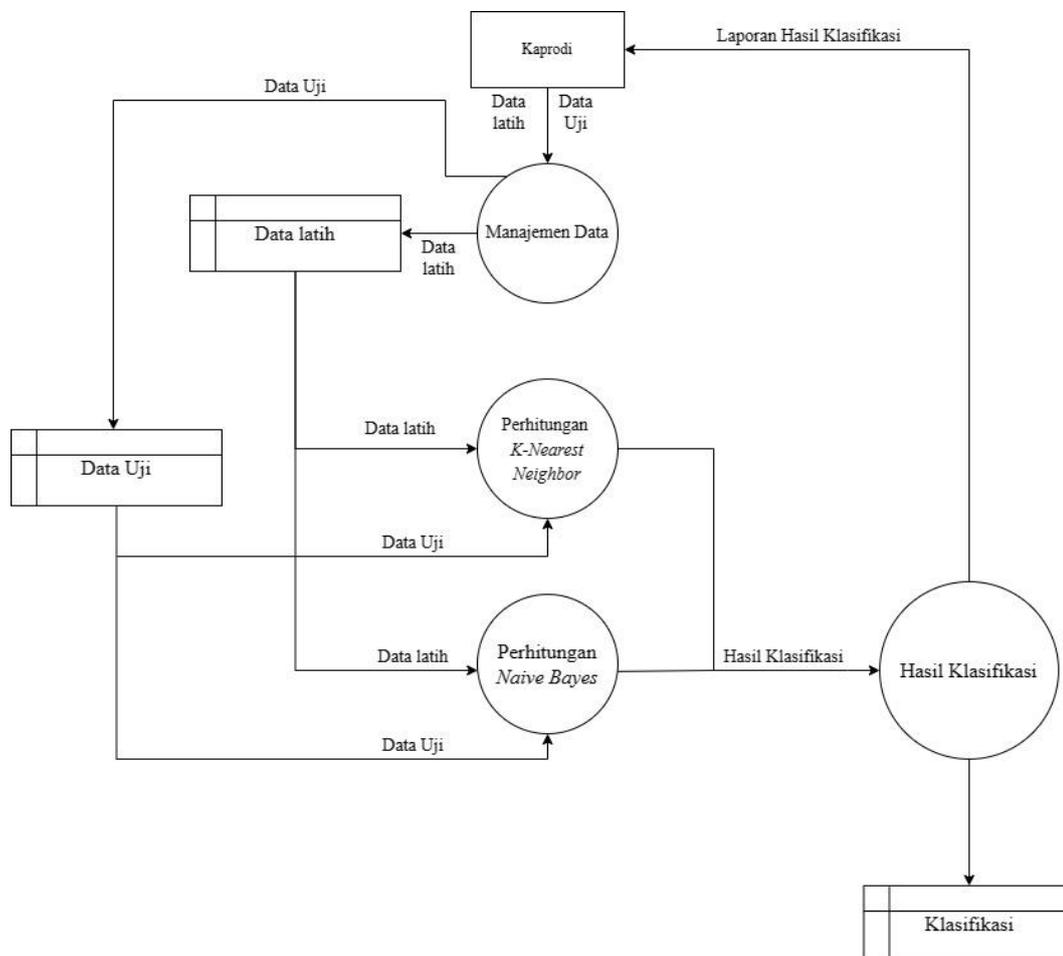
3.1 Perhitungan probabilitas awal kelas dan atribut

3.2 Menentukan Klasifikasi Kelas

3.3 Menghitung probabilitas akhir kelas dan atribut

3.4.4. Data Flow Diagram

a. Data Flow Diagram Level 0

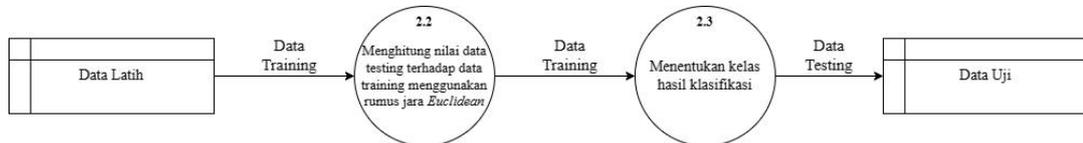


Gambar 3.4. Data Flow Diagram

Dari data flow diagram level 0 yang ada pada **Gambar 3.4.** Data Flow Diagram ada beberapa proses yang terjadi pada flow tersebut, antara lain :

1. Kepala program studi memasukkan data latih dan data uji pada sistem dan data tersebut akan disimpan ke dalam *database* yang sudah dibuat pada sistem.
2. Data latih dan data uji akan dilakukan pengujian menggunakan 2 metode yang dipilih oleh kepala program studi yaitu metode *k-Nearest Neighbor*, dan *Naive Bayes*. Hasil perhitungan dari ke 2 metode tersebut akan disimpan pada *database*, dan hasilnya juga diteruskan kepada kepala program studi.

b. Data Flow Diagram Level 1 pada metode *k-Nearest Neighbor*

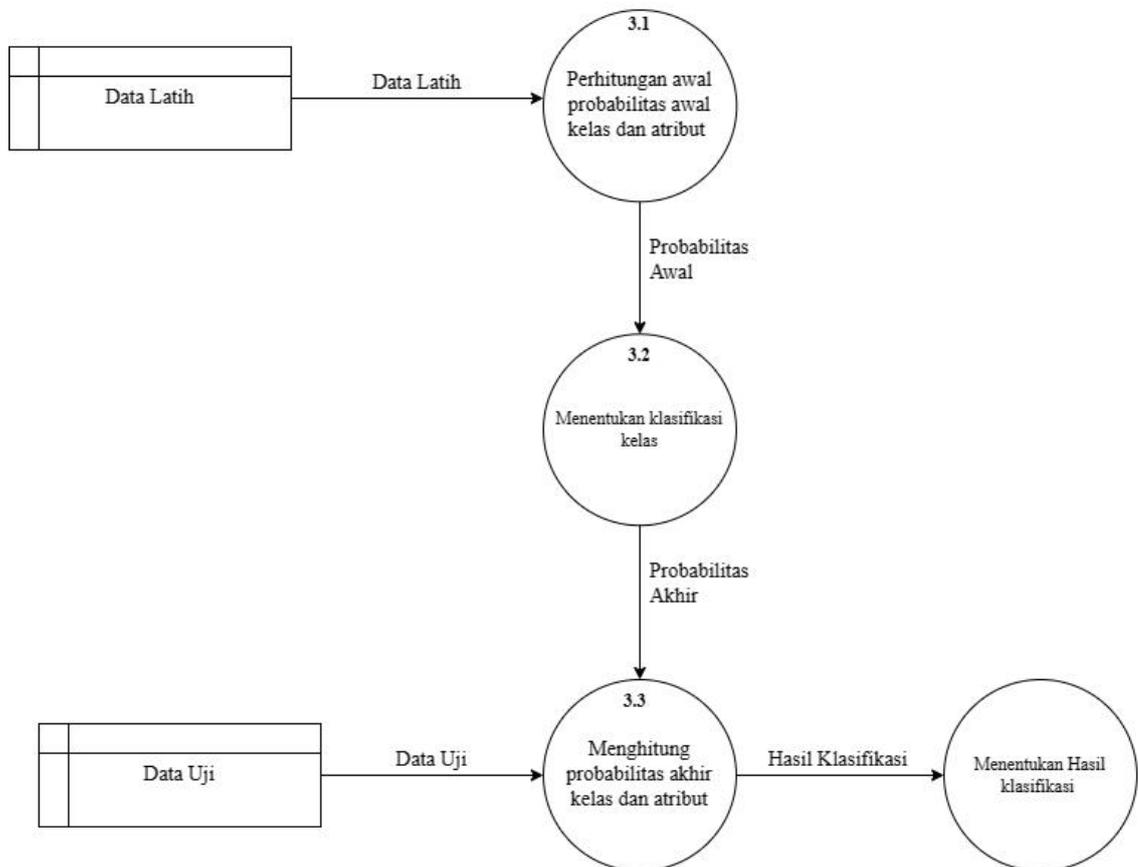


Gambar 3.5. Data Flow *k-Nearest Neighbor*

Berdasarkan data *flow diagram* level 1 yang ada pada **Gambar 3.5.** Data Flow *k-Nearest Neighbor* terdapat beberapa proses yang terjadi, antara lain :

1. Menghitung nilai data testing terhadap data *training* menggunakan rumus jarak *Euclidean*.
2. Menentukan kelas label hasil klasifikasi berdasarkan kelas mayoritas yang paling banyak muncul.

c. *Data Flow Diagram Level 1* pada metode *Naive Bayes*



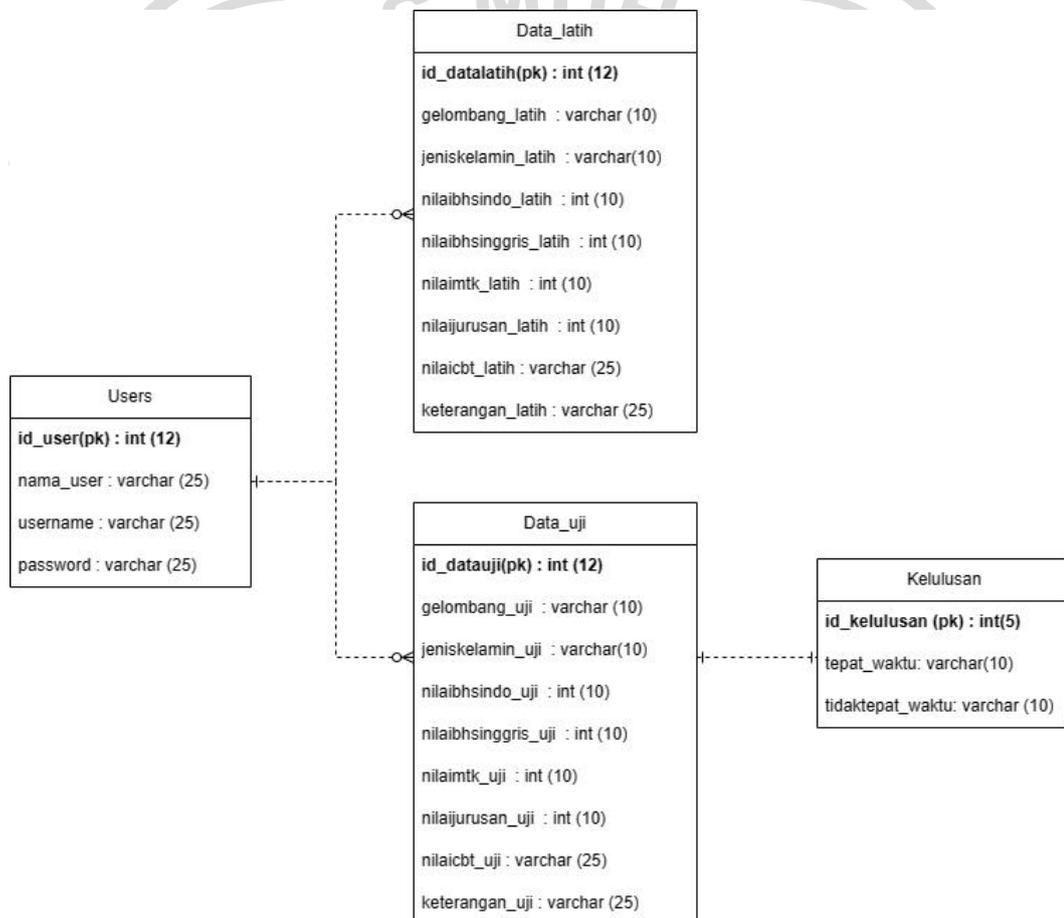
Gambar 3.6. Data Flow Diagram *Naive Bayes*

Berdasarkan pada data *flow* yang ada pada **Gambar 3.6.** Data Flow Diagram *Naive Bayes* diagram level 1 yang ada pada metode *naive bayes* ini, terdapat beberapa proses yang terjadi. Proses tersebut antara lain :

1. Menghitung probabilitas awal kemunculan nilai atribut pada tiap kelas dan probabilitas kelas dan probabilitas awal kelas pada data latih.
2. Menghitung probabilitas akhir kemunculan nilai atribut pada tiap kelas dan probabilitas awal kelas pada data latih.
3. Menentukan klasifikasi kelas pada data uji dengan menggunakan nilai probabilitas akhir yang sudah didapatkan pada data latih.

3.5. PERANCANGAN BASIS DATA

Untuk tempat penyimpanan data pada sistem klasifikasi kelulusan calon mahasiswa baru ini, di butuh kan sebuah *database* yang terdiri dari 4 tabel. Tabel tersebut terdiri dari tabel *user*, tabel data latih, tabel data uji, dan tabel kelulusan. Model konseptual dalam menggambarkan hubungan antara tabel dan memodelkan struktur data dari tabel termuat pada *Entity Relationship Diagram*. Berikut *Entity Relationship Diagram* pada sistem klasifikasi kelulusan calon mahasiswa baru ini seperti pada **Gambar 3.7**. Perancangan Basis Data Sistem :



Gambar 3.7. Perancangan Basis Data Sistem

3.6. PERANCANGAN ANTARMUKA SISTEM

a. Halama *Login*

Halaman *login* adalah halaman awal dari sistem klasifikasi calon mahasiswa baru ini. *Users* pada sistem ini harus memasukkan *username* dan *password* yang sesuai untuk dapat masuk ke dalam sistem. Tampilan pada halaman *login* sistem ini seperti pada gambar **Gambar 3.8.** Halaman Login.



SISTEM KLASIFIKASI CALON MAHASISWA BARU
DI UNIVERSITAS MUHAMMADIYAH GRESIK

Nama

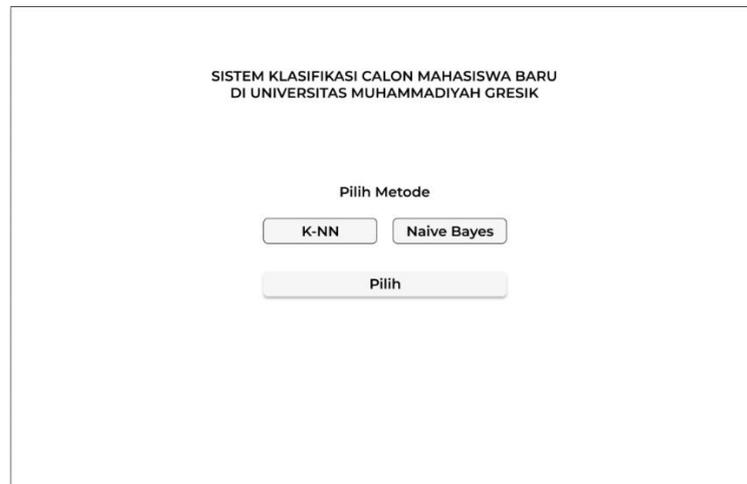
Kata Sandi

Masuk

Gambar 3.8. Halaman Login

b. Halaman Pilih Metode

Halaman pilih metode ini nantinya di gunakan untuk memilih metode yang digunakan, dikarenakan sistem ini adalah untuk membandingkan antara 2 metode yaitu *k-Nearest Neighbor* Dan *Naive Bayes* seperti pada gambar **Gambar 3.9.** Halaman Pilih Metode.



Gambar 3.9. Halaman Pilih Metode

c. Halaman Data Latih

Halaman data latih ini digunakan *users* untuk menambahkan data latih ke dalam sistem yang nantinya akan di proses oleh sistem ke dalam *database* yang sudah dibuat seperti pada gambar **Gambar 3.10**. Halaman Data Latih.



Gambar 3.10. Halaman Data Latih

d. Halaman Data Uji

Halaman data uji ini digunakan untuk *users* dalam menambahkan data uji ke dalam sistem yang nantinya akan disimpan ke dalam *database*, dan melihat detail dari data uji seperti pada gambar **Gambar 3.11**. Halaman Data Uji.

**SISTEM KLASIFIKASI CALON MAHASISWA BARU
DI UNIVERSITAS MUHAMMADIYAH GRESIK**

Data Uji Tambah Data

No	Gelombang	Jenis Kelamin	Kategorial Rata2 Nilai Ijazah	Kategorial Rata2 Nilai CBT	Relevansi Jurusan	Kesimpulan
1	Gelombang 1	L	SANGAT BAIK	BAIK	KURANG RELEVAN	TEPAT JURUSAN
2	Gelombang 3	L	SANGAT BAIK	CUKUP	KURANG RELEVAN	TEPAT JURUSAN
3	Gelombang 2	L	SANGAT BAIK	BAIK	RELEVAN	TEPAT JURUSAN
4	Gelombang 2	P	BAIK	BAIK	RELEVAN	TIDAK TEPAT JURUSAN
5	Gelombang 2	P	SANGAT BAIK	BAIK	RELEVAN	TEPAT JURUSAN
6	Gelombang 3	L	CUKUP	SANGAT BAIK	TIDAK RELEVAN	TIDAK TEPAT JURUSAN
7	Gelombang 3	L	BAIK	SANGAT BAIK	TIDAK RELEVAN	TIDAK TEPAT JURUSAN

Gambar 3.11. Halaman Data Uji

e. Halaman Klasifikasi

Halaman klasifikasi ini menjadi halaman untuk menampilkan semua hasil perhitungan klasifikasi baik dari metode *k-Nearest Neighbor*, dan *Naive Bayes* seperti yang sudah di jelaskan pada **Gambar 3.12**. Halaman Klasifikasi.

**SISTEM KLASIFIKASI CALON MAHASISWA BARU
DI UNIVERSITAS MUHAMMADIYAH GRESIK**

Naive Bayes Tambah Data

No	Gelombang	Jenis Kelamin	Kategorial Rata2 Nilai Ijazah	Kategorial Rata2 Nilai CBT	Relevansi Jurusan	Kesimpulan
1	Gelombang 1	L	SANGAT BAIK	BAIK	KURANG RELEVAN	TEPAT JURUSAN
2	Gelombang 3	L	SANGAT BAIK	CUKUP	KURANG RELEVAN	TEPAT JURUSAN
3	Gelombang 2	L	SANGAT BAIK	BAIK	RELEVAN	TEPAT JURUSAN
4	Gelombang 2	P	BAIK	BAIK	RELEVAN	TIDAK TEPAT JURUSAN
5	Gelombang 2	P	SANGAT BAIK	BAIK	RELEVAN	TEPAT JURUSAN
6	Gelombang 3	L	CUKUP	SANGAT BAIK	TIDAK RELEVAN	TIDAK TEPAT JURUSAN
7	Gelombang 3	L	BAIK	SANGAT BAIK	TIDAK RELEVAN	TIDAK TEPAT JURUSAN

Hasil Akurasi : 80.5%

Gambar 3.12. Halaman Klasifikasi

