

## BAB III

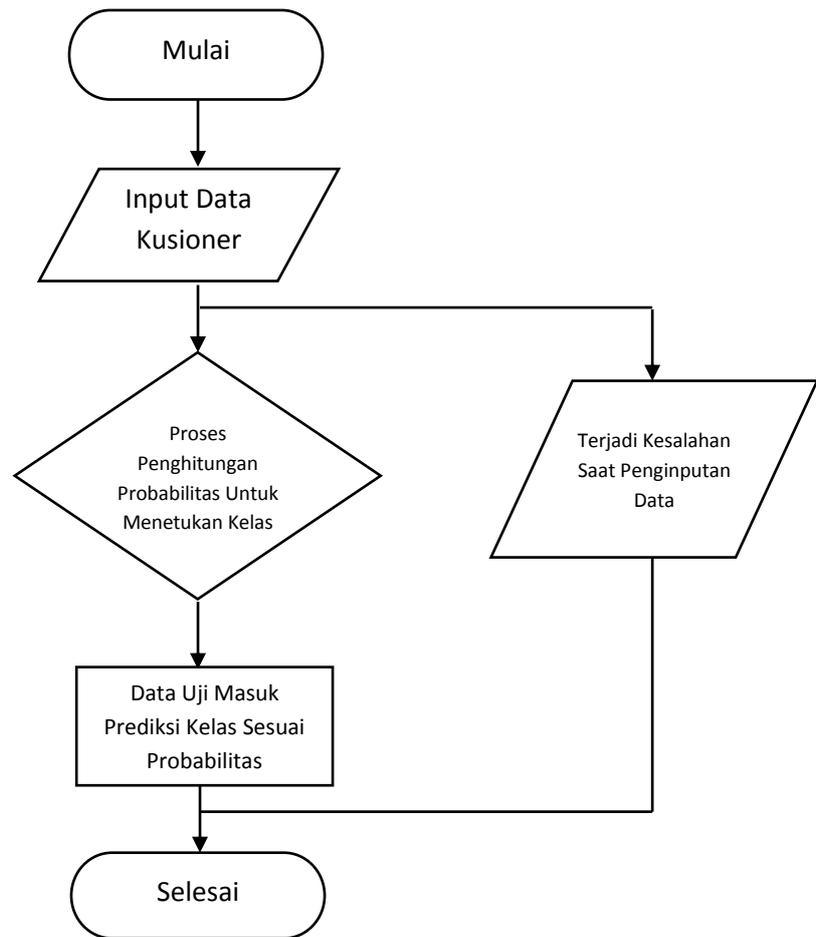
### ANALISIS dan PERANCANGAN

#### 3.1 Analisis Sistem

Sistem yang akan dibangun adalah sebuah aplikasi yang dapat memprediksikan kategori TOEFL mahasiswa berdasarkan data ketika D1 bahasa Inggris pada saat semester 1 dan 2. Pada akhir D1 pada semester 2 selalu diadakan TOEFL untuk mengetahui kemampuan mahasiswa berdasarkan ilmu yang mereka peroleh selama perkuliahan D1 bahasa Inggris. Nilai D1 yang mendasari TOEFL meliputi *reading*, *writing* dan *listening*. Skor untuk mahasiswa non bahasa mempunyai standard minimal skor yang harus dicapai dan belum adanya sebuah prediksi untuk membantu mahasiswa mengetahui gambaran awal kemampuan mereka dengan perbandingan data mahasiswa yang telah melakukan TOEFL.

Berikut **gambar 3.1** penjelasan dari *flowchart* sistem prediksi kategori nilai TOEFL:

1. Mahasiswa akan memasukkan data sesuai dengan pertanyaan yang ada pada kusioner mengenai nilai D1 bahasa Inggris yang mereka peroleh sesuai nilai pada sertifikat D1.
2. Sistem akan menghitung probabilitas kelas dan tiap fitur dari data latih yang tersimpan di dalam database. Untuk fitur bertipe numerik perhitungan akan dilakukan berdasarkan rumus (2.3).
3. Perhitungan dilakukan untuk menghitung probabilitas akhir dari data uji terhadap data latih.
4. Sistem akan mengklasifikasikan kelas dari data uji berdasarkan probabilitas akhir terbesar.



**Gambar 3.1** Flowchart Sitem Prediksi Kategori TOEFL

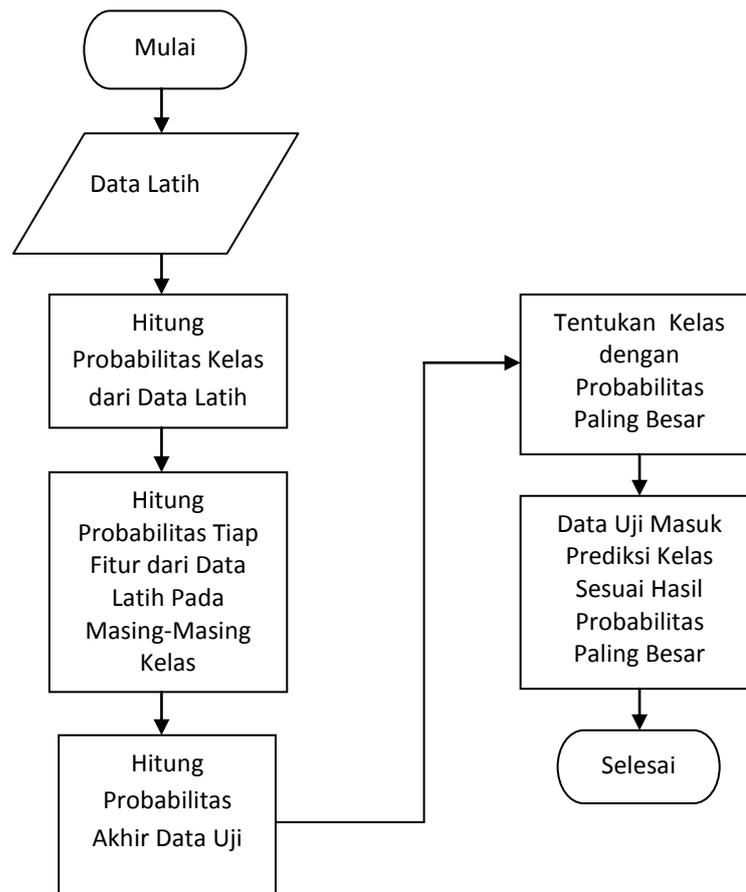
### 3.2 Hasil Analisis

Hasil analisis masalah, didapatkan dari data mahasiswa Teknik Informatika Universitas Muhammadiyah Gresik kelas pagi yang mengikuti D1 bahasa Inggris, diklasifikasikan menjadi 2 kelas, yaitu rendah dan tinggi. Hasilnya berupa informasi yang dapat membantu pihak *Language Center* (LC) dan mahasiswa untuk mengetahui kategori nilai TOEFL secara dini. Terdapat tiga entitas, yaitu :

- a. Mahasiswa : Pihak yang mengisi kusioner (data uji).
- b. Admin LC : Pihak yang mengisi data mahasiswa dan data latih.
- b. Kepala LC : Pihak yang dapat melihat laporan hasil prediksi kategori nilai TOEFL mahasiswa.

Rekomendasi penggunaan metode *Naive Bayes* untuk melakukan prediksi kategori nilai TOEFL. *Naive Bayes* merupakan pengklasifikasian statistik yang

dapat memprediksi probabilitas keanggotaan kelas (Han dan Kamber. 2006). *Naive Bayes* memiliki keunggulan dalam menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian karena hanya membutuhkan jumlah pelatihan data dalam jumlah kecil. Berikut *flowchart* proses prediksi kategori nilai TOEFL menggunakan metode *Naive Bayes* terlihat pada **gambar 3.2**.



**Gambar 3.2** *Flowchart* Perhitungan Prediksi Kategori TOEFL dengan *Naive Bayes*

### 3.2.1 Sumber Data

Tahapan awal yang dilakukan dalam penelitian adalah mengumpulkan data, dimana data diperoleh dari *Language Center (LC)* sesuai sertifikat nilai D1 mahasiswa. Data yang digunakan adalah data mahasiswa angkatan tahun 2008-2011 kelas pagi yang telah mengikuti D1 bahasa Inggris. Jumlah data yang digunakan sebanyak 60 record yang akan dibagi menjadi 50 data latih dan 10 data

uji. Data yang didapatkan dibagi menjadi 7 fitur seperti tampak pada **Tabel 3.1**

Fitur Penentuan :

No	Fitur	Keterangan
1.	Nilai Listening D1 bahasa Inggris 1	Nilai Listening saat D1 semester 1 sesuai disertifikat D1 bahasa Inggris
2.	Nilai Listening D1 bahasa Inggris 2	Nilai Listening saat D1 semester 2 sesuai disertifikat D1 bahasa Inggris
3.	Nilai Reading D1 bahasa Inggris 1	Nilai Reading saat D1 semester 1 sesuai disertifikat D1 bahasa Inggris
4.	Nilai Reading D1 bahasa Inggris 2	Nilai Reading saat D1 semester 2 sesuai disertifikat D1 bahasa Inggris
5.	Nilai Writting D1 bahasa Inggris 1	Nilai Writting saat D1 semester 1 sesuai disertifikat D1 bahasa Inggris
6.	Nilai Writting D1 bahasa Inggris 2	Nilai Writting saat D1 semester 2 sesuai disertifikat D1 bahasa Inggris
7.	Nilai TOEFL D1	Nilai Hasil TOEFL saat selesai program D1

Berikut contoh sebagian isi tabel dari data latihan terlihat pada **Tabel 3.2:**

**Tabel 3.2** Contoh Tabel Data Latihan

NO	NAMA	D1 SEMESTER 1			D1 SEMESTER 2			TOEFL D1	TOEFL
		L1	R1	W1	L2	R2	W2		
1	FH	B	C	B	C	BC	C	370	373
2	MZA	B	B	AB	C	BC	C	360	350
3	MDTS	AB	C	D	C	D	D	330	360

### 3.2.2 Analisa Kebutuhan Fungsional

Kebutuhan fungsional untuk prediksi nilai TOEFL mahasiswa adalah :

1. Sistem dapat melakukan validasi *login* berdasarkan hak akses user.
2. Sistem dapat melakukan proses *input* data kusioner mahasiswa.

3. Sistem dapat melakukan prediksi klasifikasi kategori TOEFL data uji baru yang diinputkan mahasiswa berdasarkan data latih yang telah tersimpan dalam database menggunakan metode *Naive Bayes*.

### 3.2.3 Persiapan Data dan Perhitungan

Sebelum data digunakan dilakukan *preprocessing* data untuk meningkatkan efisiensi dari sebuah sistem prediksi, dimana langkah-langkah yang dilakukan antara lain menghilangkan kerangkapan data, menggabungkan data (agregasi) dan penentuan kelas pada data latih berdasarkan skor TOEFL. Data latih setelah *preprocessing* mengalami perubahan fitur skor TOEFL akan diganti menjadi fitur kelas Rendah dan Tinggi berdasarkan kategori nilai TOEFL. Data diklasifikasikan menjadi 2 kelas yaitu :

- Nilai TOEFL Rendah ( $\geq 200 - 379$ ).
- Nilai TOEFL Tinggi ( $380 - \leq 500$ ).

Data kusioner yang sudah melalui tahap *preprocessing* maka akan dijadikan data latih untuk mengklasifikasikan data uji menggunakan metode *Naive Bayes*. Berikut **tabel 3.3** berisi sebagian data latih setelah *preprocessing* untuk prediksi kategori nilai TOEFL, keseluruhan tabel data latih yang berjumlah 50 data dapat dilihat pada lampiran.

**Tabel 3.3** Data Latih setelah *preprocessing*

NO	NAMA	D1 SEMESTER 1			D1 SEMESTER 2			TOEFL D1	KELAS
		L1	R1	W1	L2	R2	W2		
1	FH	B	C	B	C	BC	C	370	RENDAH
2	MZA	B	B	AB	C	BC	C	360	RENDAH
3	MDTS	AB	C	D	C	D	D	330	RENDAH
4	EI	A	B	B	B	AB	B	397	TINGGI

5	UA	A	AB	B	B	C	E	357	RENDAH
---	----	---	----	---	---	---	---	-----	--------

Data uji keseluruhan dapat dilihat pada lampiran. Berikut sebagian tampilan data uji terlihat pada **Tabel 3.4**:

**Tabel 3.4** Tabel Data Uji

NO	NAMA	D1 SEMESTER 1			D1 SEMESTER 2			TOEFL D1
		L1	R1	W1	L2	R2	W2	
1	IH	AB	B	B	B	B	AB	460
2	SCF	A	AB	AB	AB	BC	B	427
3	MKN	A	B	AB	A	BC	AB	387
4	MA	A	AB	BC	AB	C	AB	393
5	SA	AB	A	A	A	A	A	407

1. Menghitung nilai probabilitas kelas, kelas rendah dan tinggi. Perhitungan probabilitas kelas menggunakan persamaan rumus (2.4)

$P(Y)$  = Probabilitas Kelas

$\sum Y$  = Jumlah seluruh data pada tiap kelas (rendah dan tinggi)

$N$  = Jumlah total data latih

$$P(\text{Tinggi}) = 25/50 = 0,5$$

$$P(\text{Rendah}) = 25/50 = 0,5$$

2. Menghitung Nilai Probabilitas Tiap Fitur sesuai persamaan rumus (2.5)

$\sum X$  = Jumlah fitur setiap kelas

$\sum Y$  = Jumlah seluruh data pada tiap kelas (rendah dan tinggi)

- Nilai D1 Bhs. Inggris

**Tabel 3.5** Nilai Probabilitas Fitur D1 Listening 1

	Tinggi	Rendah
A	$\sum A / \sum \text{Tinggi} = 8/25 = 0,32$	$\sum A / \sum \text{Rendah} = 1/25 = 0,04$

AB	$\sum AB / \sum \text{Tinggi} = 14/25 = 0.56$	$\sum AB / \sum \text{Rendah} = 3/25 = 0.12$
B	$\sum B / \sum \text{Tinggi} = 1/25 = 0.04$	$\sum B / \sum \text{Rendah} = 7/25 = 0.28$
BC	$\sum BC / \sum \text{Tinggi} = 1/25 = 0.04$	$\sum BC / \sum \text{Rendah} = 5/25 = 0,2$
C	$\sum C / \sum \text{Tinggi} = 1/25 = 0.04$	$\sum C / \sum \text{Rendah} = 7/25 = 0,28$
D	$\sum D / \sum \text{Tinggi} = 0$	$\sum D / \sum \text{Rendah} = 1/25 = 0,04$
E	$\sum E / \sum \text{Tinggi} = 0$	$\sum E / \sum \text{Rendah} = 1/25 = 0,04$

**Tabel 3.6** Nilai Probabilitas Fitur D1 Reading 1

	Tinggi	Rendah
A	$\sum A / \sum \text{Tinggi} = 7/25 = 0,28$	$\sum A / \sum \text{Rendah} = 0$
AB	$\sum AB / \sum \text{Tinggi} = 5/25 = 0,2$	$\sum AB / \sum \text{Rendah} = 1/25 = 0,04$
B	$\sum B / \sum \text{Tinggi} = 11/25 = 0,44$	$\sum B / \sum \text{Rendah} = 6/25 = 0,24$
BC	$\sum BC / \sum \text{Tinggi} = 1/25 = 0,04$	$\sum BC / \sum \text{Rendah} = 8/25 = 0,32$
C	$\sum C / \sum \text{Tinggi} = 1/25 = 0,04$	$\sum C / \sum \text{Rendah} = 7/25 = 0,28$
D	$\sum D / \sum \text{Tinggi} = 0$	$\sum D / \sum \text{Rendah} = 2/25 = 0,08$
E	$\sum E / \sum \text{Tinggi} = 0$	$\sum E / \sum \text{Rendah} = 1/25 = 0,04$

**Tabel 3.7** Nilai Probabilitas Fitur D1 Writting 1

	Tinggi	Rendah
A	$\sum A / \sum \text{Tinggi} = 4/25 = 0,16$	$\sum A / \sum \text{Rendah} = 0$
AB	$\sum AB / \sum \text{Tinggi} = 12/25 = 0,48$	$\sum AB / \sum \text{Rendah} = 1/25 = 0,04$
B	$\sum B / \sum \text{Tinggi} = 6/25 = 0,24$	$\sum B / \sum \text{Rendah} = 9/25 = 0,36$
BC	$\sum BC / \sum \text{Tinggi} = 2/25 = 0,08$	$\sum BC / \sum \text{Rendah} = 2/25 = 0,08$
C	$\sum C / \sum \text{Tinggi} = 1/25 = 0,04$	$\sum C / \sum \text{Rendah} = 9/25 = 0,36$

D	$\sum D / \sum \text{Tinggi} = 0$	$\sum D / \sum \text{Rendah} = 3/25 = 0,12$
E	$\sum E / \sum \text{Tinggi} = 0$	$\sum E / \sum \text{Rendah} = 1/25 = 0,04$

**Tabel 3.8** Nilai Probabilitas Fitur D1 Listening 2

	Tinggi	Rendah
A	$\sum A / \sum \text{Tinggi} = 5/25 = 0,2$	$\sum A / \sum \text{Rendah} = 0$
AB	$\sum AB / \sum \text{Tinggi} = 10/25 = 0,4$	$\sum AB / \sum \text{Rendah} = 1/25 = 0,04$
B	$\sum B / \sum \text{Tinggi} = 4/25 = 0,16$	$\sum B / \sum \text{Rendah} = 4/25 = 0,16$
BC	$\sum BC / \sum \text{Tinggi} = 3/25 = 0,12$	$\sum BC / \sum \text{Rendah} = 6/25 = 0,24$
C	$\sum C / \sum \text{Tinggi} = 3/25 = 0,12$	$\sum C / \sum \text{Rendah} = 10/25 = 0,4$
D	$\sum D / \sum \text{Tinggi} = 0$	$\sum D / \sum \text{Rendah} = 3/25 = 0,12$
E	$\sum E / \sum \text{Tinggi} = 0$	$\sum E / \sum \text{Rendah} = 1/25 = 0,04$

**Tabel 3.9** Nilai Probabilitas Fitur D1 Reading 2

	Tinggi	Rendah
A	$\sum A / \sum \text{Tinggi} = 3/25 = 0,12$	$\sum A / \sum \text{Rendah} = 1/25 = 0,04$
AB	$\sum AB / \sum \text{Tinggi} = 8/25 = 0,32$	$\sum AB / \sum \text{Rendah} = 2/25 = 0,08$
B	$\sum B / \sum \text{Tinggi} = 11/25 = 0,44$	$\sum B / \sum \text{Rendah} = 2/25 = 0,08$
BC	$\sum BC / \sum \text{Tinggi} = 1/25 = 0,04$	$\sum BC / \sum \text{Rendah} = 6/25 = 0,24$
C	$\sum C / \sum \text{Tinggi} = 2/25 = 0,08$	$\sum C / \sum \text{Rendah} = 12/25 = 0,48$
D	$\sum D / \sum \text{Tinggi} = 0$	$\sum D / \sum \text{Rendah} = 1/25 = 0,04$
E	$\sum E / \sum \text{Tinggi} = 0$	$\sum E / \sum \text{Rendah} = 1/25 = 0,04$

**Tabel 3.10** Nilai Probabilitas Fitur D1 Writting 2

	Tinggi	Rendah
A	$\sum A / \sum \text{Tinggi} = 3/25 = 0,12$	$\sum A / \sum \text{Rendah} = 0$
AB	$\sum AB / \sum \text{Tinggi} = 13/25 = 0,52$	$\sum AB / \sum \text{Rendah} = 5/25 = 0,2$
B	$\sum B / \sum \text{Tinggi} = 5/25 = 0,2$	$\sum B / \sum \text{Rendah} = 5/25 = 0,2$
BC	$\sum BC / \sum \text{Tinggi} = 3/25 = 0,12$	$\sum BC / \sum \text{Rendah} = 0$
C	$\sum C / \sum \text{Tinggi} = 1/25 = 0,04$	$\sum C / \sum \text{Rendah} = 12/25 = 0,48$
D	$\sum D / \sum \text{Tinggi} = 0$	$\sum D / \sum \text{Rendah} = 2/25 = 0,08$
E	$\sum E / \sum \text{Tinggi} = 0$	$\sum E / \sum \text{Rendah} = 1/25 = 0,04$

3. Menghitung Probabilitas Numerik pada Fitur “TOEFL D1” tiap data uji, sesuai persamaan rumus (2.3).

$\mu$  = Rata-rata semua data pada masing-masing kelas (rendah dan tinggi)

$ij$  = Kelas pada data latih (rendah dan tinggi)

$\sigma^2$  = Ragam/varian

$\sigma$  = Akar dari ragam/varian

$x_i$  = Nilai tiap fitur pada data uji (Nilai D1 bahasa Inggris)

$\pi$  = 3,14

exp = 2,7182

Rata-rata diperoleh dari semua data di kelas tinggi dibagi banyaknya data pada kelas tinggi, dan sebaliknya juga kelas rendah.  $\sigma^2$  diperoleh dari masing-masing fitur pada tiap baris TOEFL D1 dikurangi rata-rata pangkat dua lalu dijumlah, setelah itu dibagi banyaknya data pada kelas tinggi yang sudah dikurangi 1.  $\sigma$  diperoleh dari hasil akar dari varian.

$$\mu_{\text{Tinggi}} = \frac{10333}{25} = 413,32$$

$$\sigma^2_{\text{Tinggi}} = 460,6433$$

$$\sigma_{\text{Tinggi}} = \sqrt{460,6433} = 21,4626$$

$$\mu_{\text{Rendah}} = \frac{8956}{25} = 358,24$$

$$\sigma^2_{\text{Rendah}} = 428,60667$$

$$\sigma_{\text{Rendah}} = \sqrt{428,60667} = 20,70282$$

- Data Uji Numerik Pertama

Perhitungan data uji pertama sesuai persamaan rumus (2.3). Untuk perhitungan data uji 2 s/d 10 dapat dilihat pada lampiran.

$$P(X_i = x_i | Y = ij) = \frac{1}{\sqrt{2\pi} \sigma_{ij}} \exp \frac{-(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \dots\dots\dots(2.3)$$

Keterangan :

$\mu_{ij}$  = rata-rata  $X_i$  ( $\bar{x}$ ) dari semua data latih

$\sigma_{ij}^2$  = varian dari data latih

$\sigma_{ij}$  = Akar varian dari data latih

$$P(\text{TOEFL D1} = 393 | \text{Tinggi}) = \frac{1}{\sqrt{2\pi} 21,4626} \exp \frac{-(393-413,32)^2}{2 \times 460,6433} = 0.0118$$

$$P(\text{TOEFL D1} = 393 | \text{Rendah}) = \frac{1}{\sqrt{2\pi} 20,70282} \exp \frac{-(393-358,24)^2}{2 \times 428,60667} = 0.00471$$

4. Menghitung Nilai Probabilitas Akhir sesuai rumus (2.6), mengalikan hasil perhitungan nilai probabilitas kelas dan nilai probabilitas tiap fitur sesuai data pada data uji. Untuk perhitungan probabilitas akhir data uji 2 s/d 10 dapat dilihat pada lampiran.

$P(X)$  = Nilai probabilitas masing-masing kelas data uji (rendah dan tinggi)

$i$  = Nilai tiap fitur pada data uji

$P(X_i|Y)$  = Nilai probabilitas fitur pada data latih

a. Data Uji Pertama

- Kelas Tinggi

$$\begin{aligned}
 P(X|\text{Tinggi}) &= P(\text{Tinggi}) * P(L1=A|\text{Tinggi}) * P(R1=AB|\text{Tinggi}) * P(W1=BC|\text{Tinggi}) \\
 & * P(L2=AB|\text{Tinggi}) * P(R2=C|\text{Tinggi}) * P(W2=AB|\text{Tinggi}) * P(\text{TOEFLD1}=393|\text{Tinggi}) \\
 & = 0,5 * 0,32 * 0,2 * 0,08 * 0,4 * 0,08 * 0,52 * 0.0118 \\
 & = 5.0593E-7
 \end{aligned}$$

- Kelas Rendah

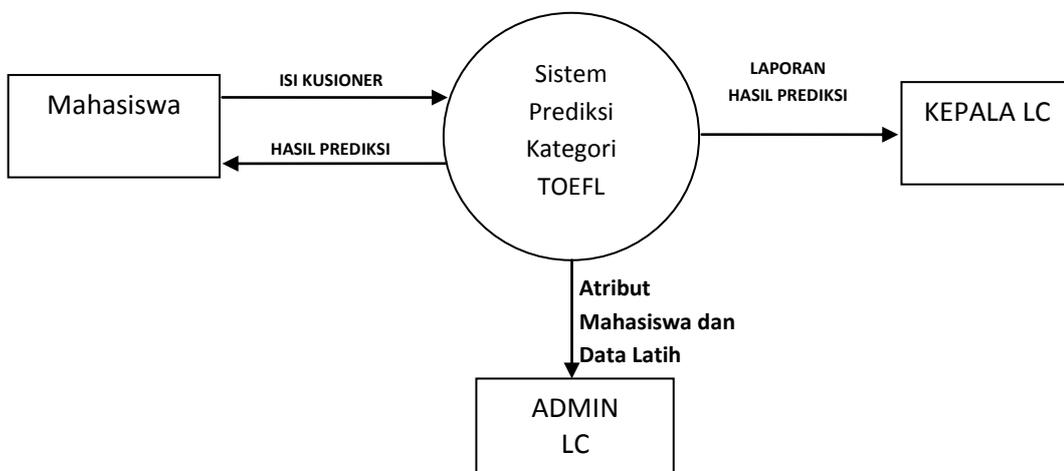
$$\begin{aligned}
 P(X|\text{Rendah}) &= P(\text{Rendah}) * P(L1=A|\text{Rendah}) * P(R1=AB|\text{Rendah}) * P(W1=BC|\text{Rendah}) \\
 & * P(L2=AB|\text{Rendah}) * P(R2=C|\text{Rendah}) * P(W2=AB|\text{Rendah}) * P(\text{TOEFLD1}=393|\text{Tinggi}) \\
 & = 0,5 * 0,04 * 0,04 * 0,08 * 0,04 * 0,48 * 0,2 * 0.00471 \\
 & = 1.16E-09
 \end{aligned}$$

Data Uji masuk kelas Tinggi karena jika dibandingkan hasil probabilitas akhir kelas Rendah dan Tinggi maka nilai maksimal terdapat pada kelas **Tinggi**.

### 3.3 Perancangan Sistem

Tahapan ini akan membahas mengenai *context diagram*, *data flow diagram*, perancangan *database* dan *interface* aplikasi.

#### 3.3.1 Context Diagram Sistem



**Gambar 3.3** Context Diagram Prediksi Kategori TOEFL

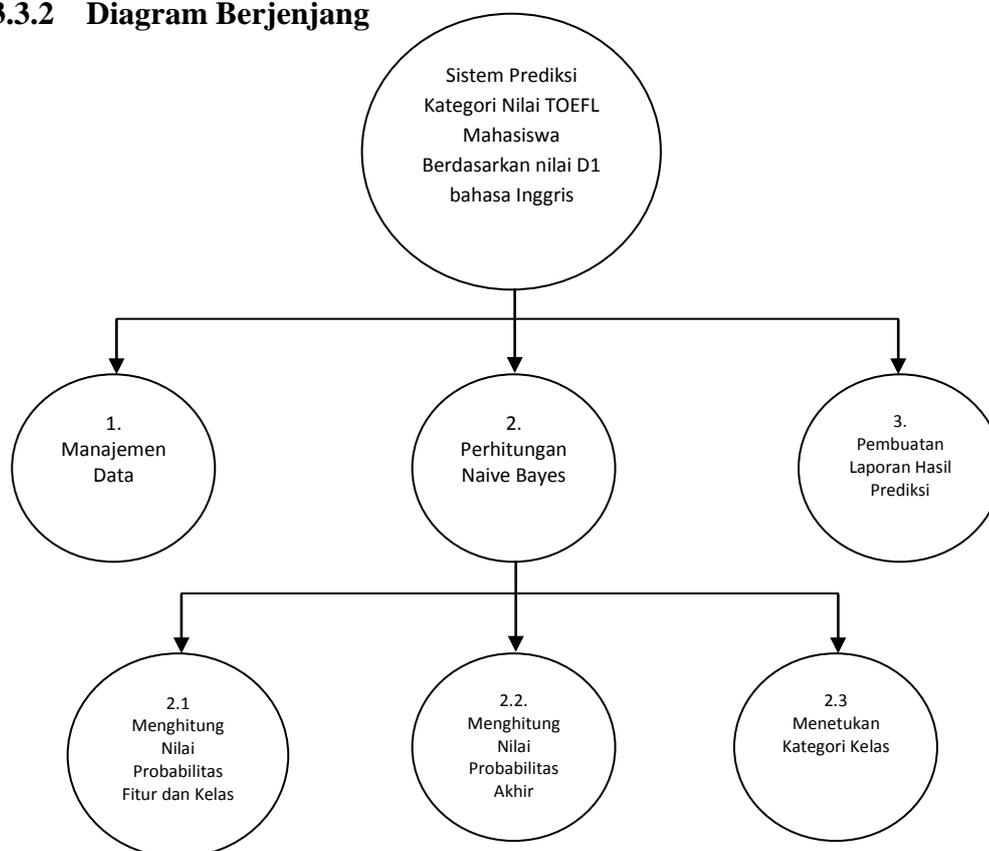
Pada *context diagram 3.3* adalah gambaran sistem secara garis besar, dimana terdapat tiga entitas luar yang berhubungan dengan sistem, yaitu :

1. Mahasiswa merupakan pihak yang akan menginputkan kusioner (data uji) dan melakukan proses klasifikasi.
2. Admin LC merupakan pihak yang dapat memanipulasi data latih mahasiswa.
3. Kepala LC merupakan pihak yang dapat melihat laporan hasil prediksi.

Penjelasan *diagram context* :

Admin memasukkan data atribut mahasiswa dan juga data latih. Mahasiswa memasukkan data kusioner sebagai data uji untuk diprediksi nilai TOEFL-nya. Data uji tersebut diklasifikasikan dalam sistem dengan menggunakan metode *Naive Bayes* yang dihitung berdasarkan nilai D1 bahasa Inggris. Mahasiswa akan menerima hasil prediksi dari data uji yang telah diprediksi tadi, sedangkan Kepala LC akan menerima laporan atau daftar hasil prediksi nilai TOEFL dari semua mahasiswa yang telah melakukan proses klasifikasi.

### 3.3.2 Diagram Berjenjang



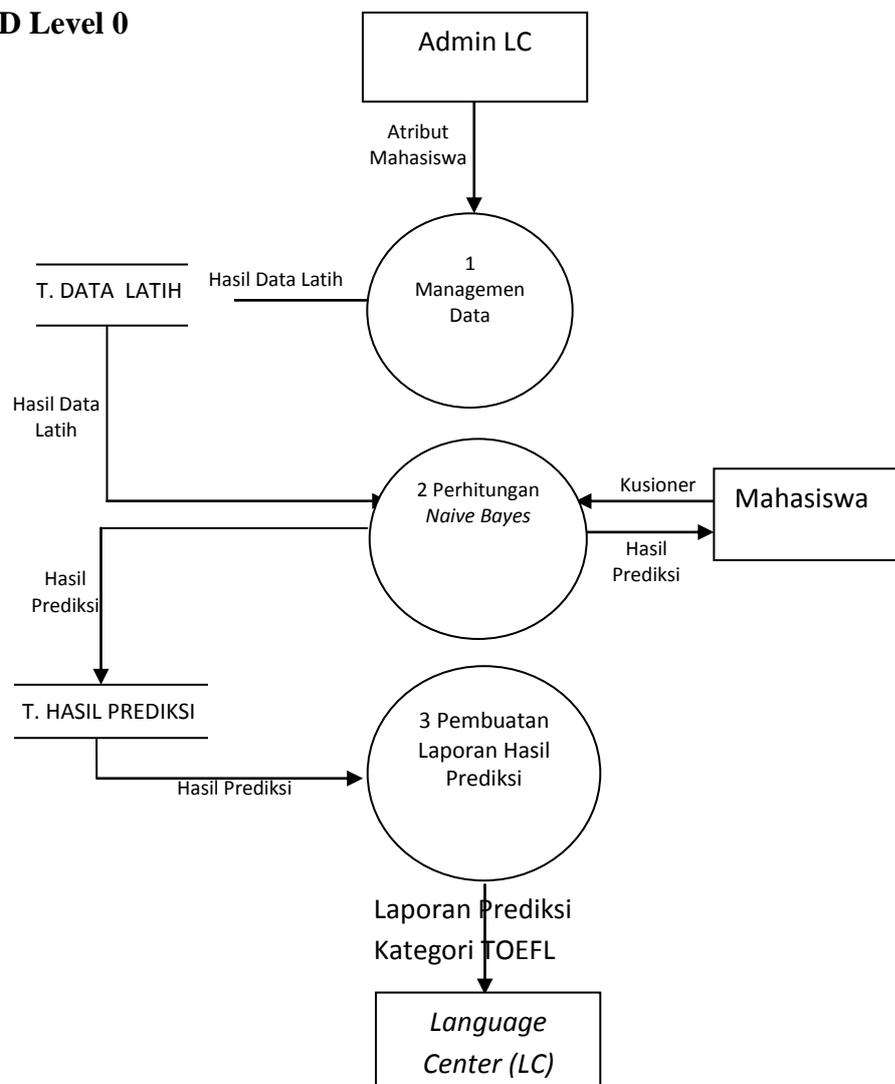
**Gambar 3.4** Diagram Berjenjang Prediksi Kategori TOEFL

Pada **gambar 3.4** dapat dijelaskan sebagai berikut :

- Top level : Sistem prediksi nilai TOEFL berdasarkan nilai nilai D1 bahasa Inggris
- Level 0 : 1. Manajemen Data  
2. Perhitungan Naive Bayes  
3. Pembuatan Laporan Hasil Prediksi
- Level 1 : 1.1 Menghitung Nilai Probabilitas Fitur dan Kelas  
1.2 Menghitung Nilai Probabilitas Akhir  
1.3 Menentukan Kategori Kelas

### 3.3.3 Data Flow Diagram (DFD)

#### a. DFD Level 0

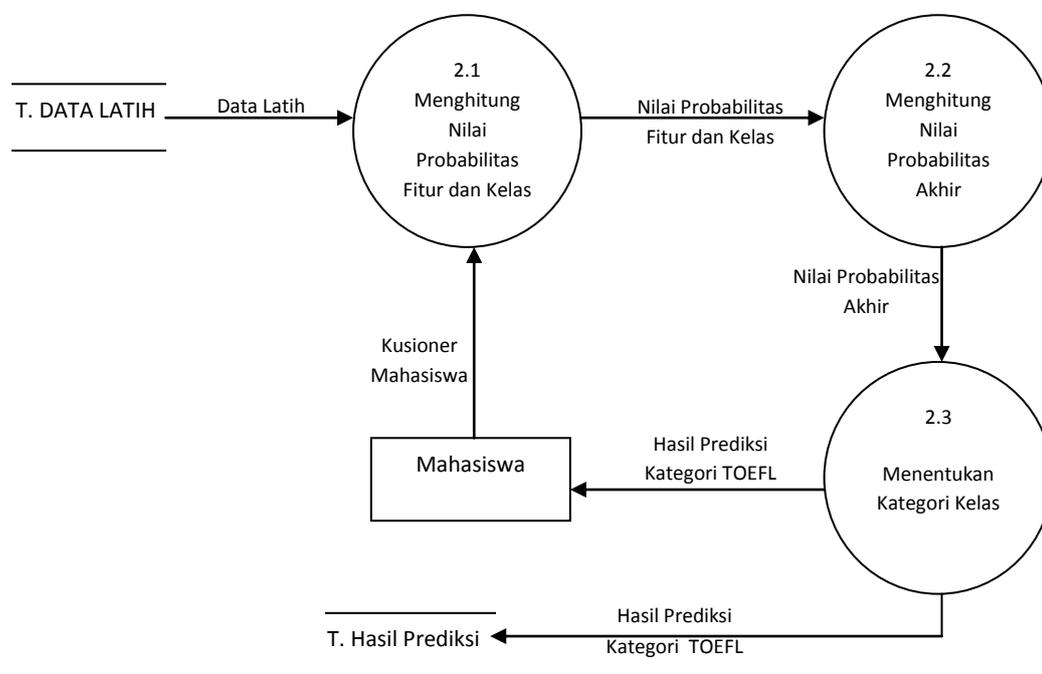


**Gambar 3.5** DFD Level 0 Prediksi Kategori TOEFL

Pada **gambar 3.5** dapat dijelaskan sebagai berikut:

- Proses 1 adalah proses manajemen data yang diinputkan oleh admin LC. Data kusioner yang diinputkan oleh panitia admin digunakan sebagai data latih.
- Proses 2 adalah perhitungan *Naive Bayes*, yaitu proses perhitungan klasifikasi data uji terhadap data latih menggunakan metode *Naive Bayes*.
- Proses 3 adalah pembuatan laporan dari daftar hasil prediksi yang telah dilakukan mahasiswa.

#### b. DFD Level 1



**Gambar 3.6** DFD Level 1 Perhitungan Prediksi Kategori TOEFL

Adapun keterangan dari **gambar 3.6** diatas ini adalah sebagai berikut :

- Proses 2.1 adalah proses menghitung nilai probabilitas fitur dan kelas, yang digunakan dalam memprediksi kategori TOEFL. Fitur yang digunakan adalah nilai D1 bahasa Inggris dihitung dari latih. Kategori kelas yang diklasifikasikan adalah kelas tinggi dan rendah.
- Proses 2.2 adalah proses menghitung nilai probabilitas akhir dari nilai probabilitas pada proses pertama tiap kelas

- Proses 2.3 adalah proses menentukan kategori TOEFL data uji. Kelas prediksi TOEFL ditentukan berdasarkan nilai probabilitas akhir terbesar.

### 3.3.4 Perancangan Database

Basis data diperlukan untuk menyimpan data yang berhubungan dengan user *login*, data latih dan hasil klasifikasi yang akan digunakan dalam proses prediksi kategori TOEFL. Berikut struktur tabel dalam basis data sistem prediksi kategori TOEFL mahasiswa berdasarkan data nilai D1 bahasa Inggris.

#### a. Struktur Tabel Mahasiswa

Tabel Mahasiswa berfungsi untuk memberikan informasi detail data mahasiswa yang akan diprediksi sesuai pada **Tabel 3.11**.

**Tabel 3.11** Struktur Tabel Mahasiswa

No	Nama Field	Tipe Data	Ukuran	Relasi	Keterangan
1.	Nim	int	8	Primary	Nomor induk mahasiswa
2.	nama	varchar	50		Nama mahasiswa
3.	jekel	char	1		Jenis kelamin
4.	angkatan	year	4		Tahun masuk mahasiswa

#### b. Struktur Tabel User

Tabel user berfungsi untuk menyimpan data user yang digunakan untuk *login* kedalam sistem dan memberikan hak akses bagi user dalam mengakses sistem sesuai pada **Tabel 3.12**.

**Tabel 3.12** Struktur Tabel User

No	Nama Field	Tipe Data	Ukuran	Relasi	Keterangan
1.	id_user	int		Primary	id pengguna sistem
2.	Username	varchar	20		Username saat <i>login</i>
3.	Password	text			Password saat <i>login</i>
4.	Level	char	1		Hak akses user

c. Struktur Tabel Data Latih

Tabel data latih berfungsi untuk menyimpan data mahasiswa yang diinputkan oleh *Language Center* (LC) yang digunakan sebagai data latih untuk memprediksi kategori TOEFL sesuai pada **Tabel 3.13**.

**Tabel 3.13** Struktur Tabel Data Latih

No	Nama Field	Tipe Data	Ukuran	Relasi	Keterangan
1.	id_latih	int		Primary	id data latih
2.	Nim	int	8	Foreign	Relasi dengan id_mahasiswa
3.	L1	varchar	2		Nilai Listening Semester 1 D1 bahasa Inggris
4.	R1	varchar	2		Nilai Reading Semester 1 D1 bahasa Inggris
5.	W1	varchar	2		Nilai Writting Semester 1 D1 bahasa Inggris
6.	L2	varchar	2		Nilai Listening Semester 2 D1 bahasa Inggris
7.	R2	varchar	2		Nilai Reading Semester 2 D1 bahasa Inggris
8.	W2	varchar	2		Nilai Writting Semester 2 D1 bahasa Inggris
9.	TOEFL D1	double			Skor TOEFL D1
10.	TOEFL	double			Skor TOEFL
11.	Kelas	char	6		Kelas prediksi nilai TOEFL

d. Struktur Tabel Data Uji

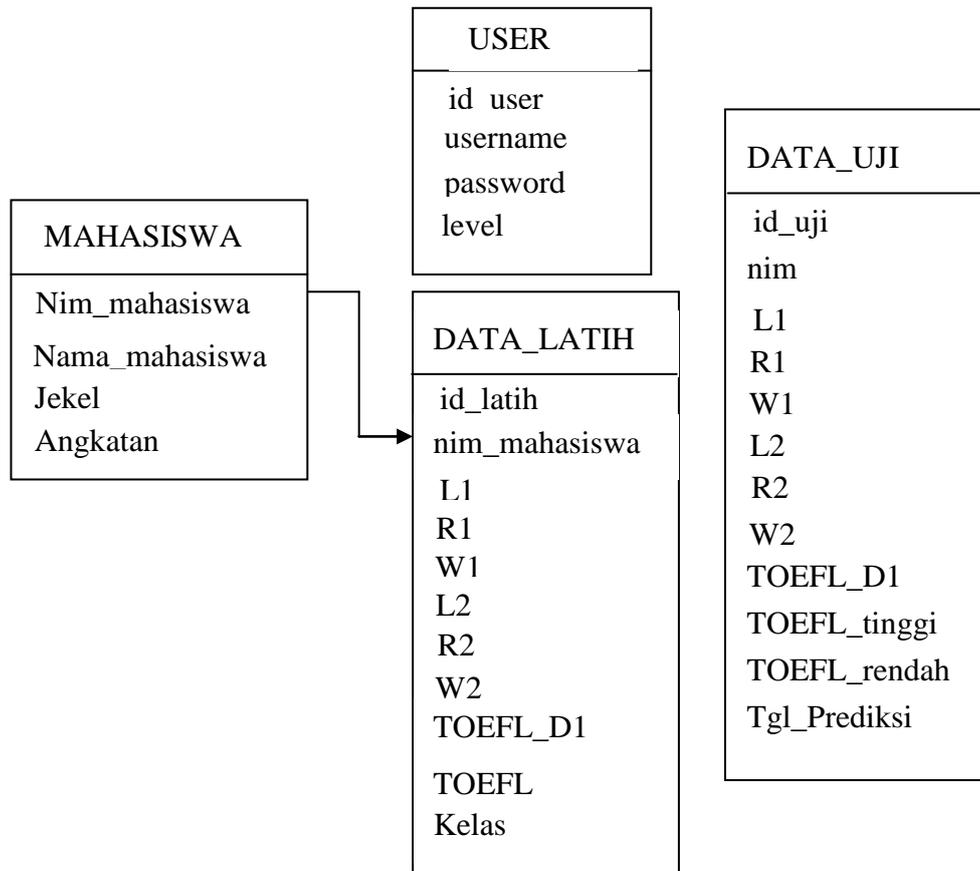
Tabel hasil prediksi berfungsi untuk menyimpan hasil dari prediksi kategori TOEFL dari data uji yang telah diujikan. Data uji diperoleh dari kusioner yang diinputkan mahasiswa sesuai pada **Tabel 3.14**.

**Tabel 3.14** Struktur Tabel Data Uji

No	Nama Field	Tipe Data	Ukuran	Relasi	Keterangan
1.	id_uji	int		Primary	id data uji
2.	Nim	int	8		Nim mahasiswa
3.	L1	varchar	2		Nilai Listening Semester 1 D1 bahasa Inggris
4.	R1	varchar	2		Nilai Reading Semester 1 D1 bahasa Inggris
5.	W1	varchar	2		Nilai Writting Semester 1 D1 bahasa Inggris
6.	L2	varchar	2		Nilai Listening Semester 2 D1 bahasa Inggris
7.	R2	varchar	2		Nilai Reading Semester 2 D1 bahasa Inggris
8.	W2	varchar	2		Nilai Writting Semester 2 D1 bahasa Inggris
9.	toefl_d1	int	3		Skor TOEFL D1
10.	kelas	varchar	6		Kelas prediksi nilai TOEFL
11.	toefl_tinggi	double			Nilai probabilitas akhir kelas
12.	toefl_rendah	double			Nilai probabilitas akhir kelas
13.	tgl_prediksi	date			Tanggal dilakukannya prediksi

### 3.3.5 Relasi Antar Tabel

Relasi antar pada **Gambar 3.7** merupakan memperlihatkan keseluruhan struktur tabel termasuk nama tabel (*entitas*), nama atribut, atribut *primary key* dan atribut *foreign key* yang menunjukkan hubungan antar tabel yang saling berelasi.



**Gambar 3.7** Relasi Antar Tabel

### 3.3.6 Analisa Kebutuhan Pembuatan Sistem

#### 1. Kebutuhan Perangkat Keras

Perangkat keras adalah alat yang digunakan untuk menunjang dalam pembuatan sistem. Dalam pembuatan sistem ini perangkat keras yang digunakan yaitu laptop dengan spesifikasi :

- a. *Processor Intel Dual Core*
- b. RAM 2 GB
- c. HDD 250 GB
- d. *Laptop 14"*

#### 2. Kebutuhan Perangkat Lunak

Perangkat lunak adalah program atau aplikasi yang digunakan untuk membangun sistem. Perangkat lunak yang dibutuhkan dalam pembuatan sistem ini adalah :

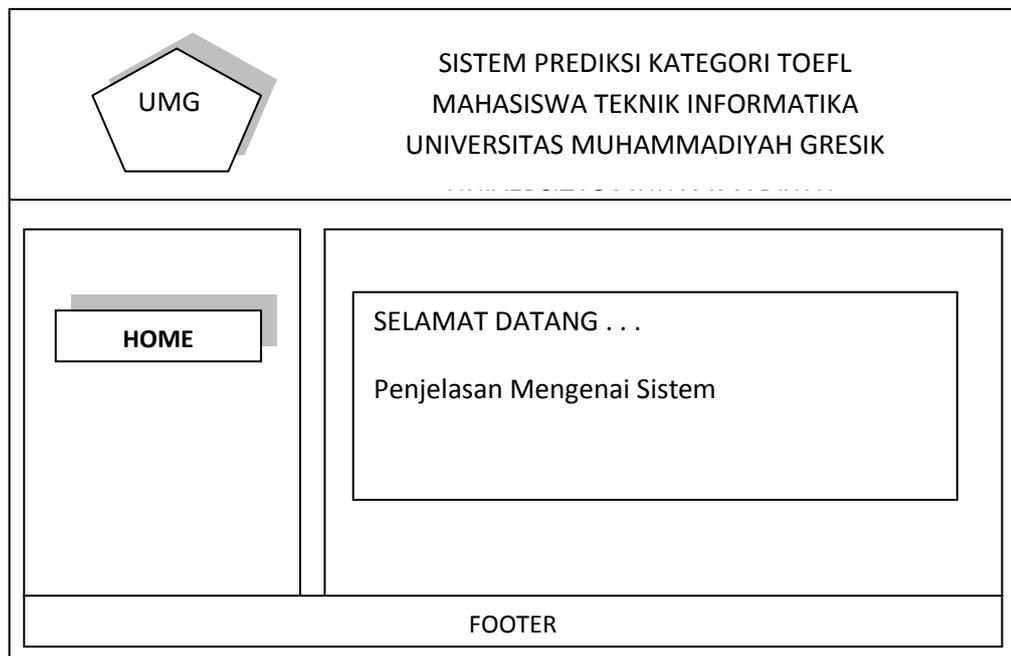
- a. *Windows 8*
- b. *Web Server* : Apache
- c. *Database Server* : MySQL
- d. Bahasa Pemrograman : PHP
- e. SQLyog Enterprise
- f. *Browser Internet (HTML 5)*

### 3.3.7 Perancangan Interface

*Interface* atau antarmuka adalah bentuk tampilan grafis yang menghubungkan antara pengguna dengan sistem. Sistem ini akan dibangun dengan menggunakan bahasa pemrograman PHP.

- a. Halaman Awal (Home)

Halaman awal (Home) pada **gambar 3.8** merupakan halaman yang berisi tentang deskripsi sistem dan merupakan halaman utama jika kita pertama membuka sistem.



**Gambar 3.8** Tampilan Awal Sistem

- b. Halaman Login

Halaman login seperti **gambar 3.9** berguna untuk memberi hak akses user untuk membedakan peran masing-masing user. Untuk mahasiswa menu yang disediakan adalah Home, Daftar, Input Kusioner dan Logout. Untuk *Language*

*Center* (LC), menu yang disediakan adalah Home, Master user, Master data latih dan uji, Laporan dan Logout.

**Gambar 3.9** Tampilan Antar Muka Halaman Login

c. Halaman Daftar (*Sign Up*)

Halaman *sign up* pada **gambar 3.10** memudahkan *Language Center* (LC) untuk mendaftarkan mahasiswa yang belum punya akun untuk *login*.

**Gambar 3.10** Tampilan Daftar Untuk *Login*

d. Halaman Untuk Mengubah Akun *Login*

Halaman edit akun *login* pada **gambar 3.11** memudahkan pihak *Language Center* (LC) untuk mengubah akun mahasiswa semisal mahasiswa tersebut ingin mengubah *username* atau *password*.

SISTEM PREDIKSI KATEGORI TOEFL  
MAHASISWA TEKNIK INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH GRESIK

HOME  
USER  
DATA  
LAPORAN  
LOGOUT

USERNAME: Mr. X  
PASSWORD: \*\*\*\*\*  
LEVEL:

UBAH CANCEL

FOOTER

**Gambar 3.11** Tampilan Untuk Ubah Data Login

e. Halaman Data Mahasiswa

**Gambar 3.12** menampilkan data mahasiswa yang melakukan prediksi.

SISTEM PREDIKSI KATEGORI TOEFL  
MAHASISWA TEKNIK INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH GRESIK

HOME  
USER  
DATA  
MAHASISWA  
DATA LATIH  
LAPORAN

ADD

No	NIM	NAMA	ANGKATAN	ACTION
1.	Mr. X	x	x	EDIT DELETE

FOOTER

**Gambar 3.12** Tampilan Master Data Mahasiswa

## f. Halaman Data Latih

Halaman pada **gambar 3.13** menampilkan isi semua data latih mahasiswa Teknik Informatika tahun 2008-2011.

SISTEM PREDIKSI KATEGORI TOEFL MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS MUHAMMADIYAH GRESIK													
													
TAMBAH DATA													
<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">HOME</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">USER</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">DATA</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">MAHASISWA</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">DATA LATIH</div> <div style="border: 1px solid black; padding: 5px;">LAPORAN</div>	No	NAMA	D1 Bhs. Inggris								TOEFL L D1	KELAS	ACTION
	Speaking		Listening		writing		reading						
	1	2	1	2	1	2	1	2					
	1	Mr. X	x	x	x	x	x	x	x	x	390	Tinggi	EDIT   DELETE
FOOTER													

**Gambar 3.13** Tampilan Master Data Latih

## g. Halaman Kusioner

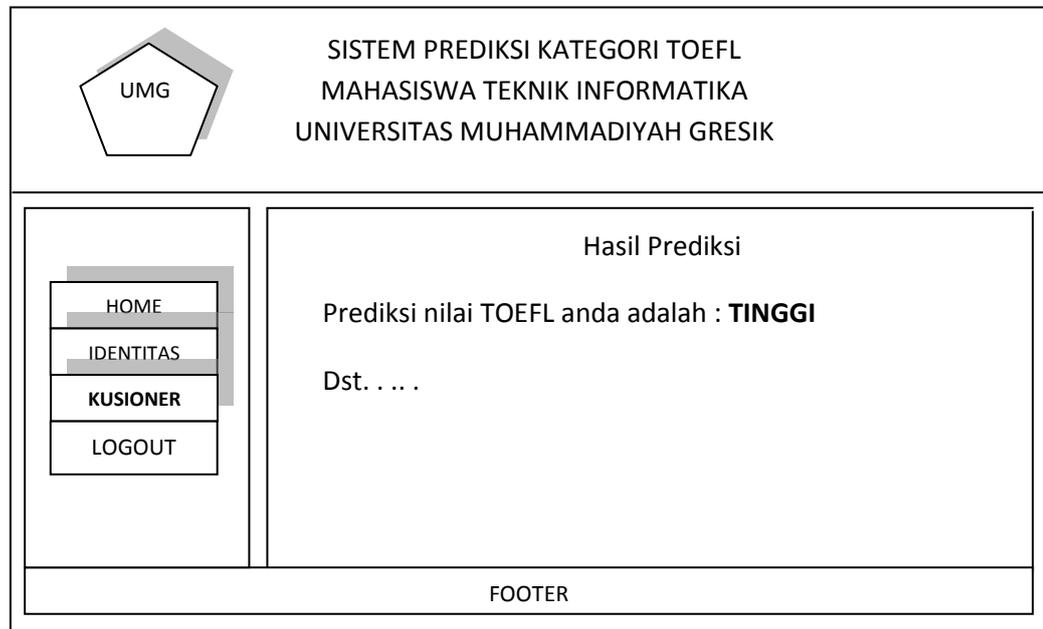
Halaman kusioner pada **gambar 3.14** hanya akan muncul jika kita login sebagai mahasiswa. Mahasiswa mempunyai hak penuh untuk mengisi kusioner. Isi kusioner tersebut yang akan menjadi sumber acuan untuk prediksi nilai TOEFL.

SISTEM PREDIKSI KATEGORI TOEFL MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS MUHAMMADIYAH GRESIK													
													
TOEFL D1 <input type="text"/>													
<div style="border: 1px solid black; padding: 2px 10px;">PROSES</div>													
<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">HOME</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">PROFIL</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">KUSIONER</div> <div style="border: 1px solid black; padding: 5px;">LOGOUT</div>	D1 Bhs Inggris :												
	- Listening 1		<input type="radio"/> A	<input type="radio"/> AB	<input type="radio"/> B	<input type="radio"/> BC	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E				
	- Reading 1		<input type="radio"/> A	<input type="radio"/> AB	<input type="radio"/> B	<input type="radio"/> BC	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E				
	- Writing 1		<input type="radio"/> A	<input type="radio"/> AB	<input type="radio"/> B	<input type="radio"/> BC	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E				
	- Listening 2		<input type="radio"/> A	<input type="radio"/> AB	<input type="radio"/> B	<input type="radio"/> BC	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E				
	- Reading 2		<input type="radio"/> A	<input type="radio"/> AB	<input type="radio"/> B	<input type="radio"/> BC	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E				
- Writing 2		<input type="radio"/> A	<input type="radio"/> AB	<input type="radio"/> B	<input type="radio"/> BC	<input type="radio"/> C	<input type="radio"/> D	<input type="radio"/> E					
FOOTER													

**Gambar 3.14** Tampilan Input Kusioner

#### h. Halaman Hasil Prediksi

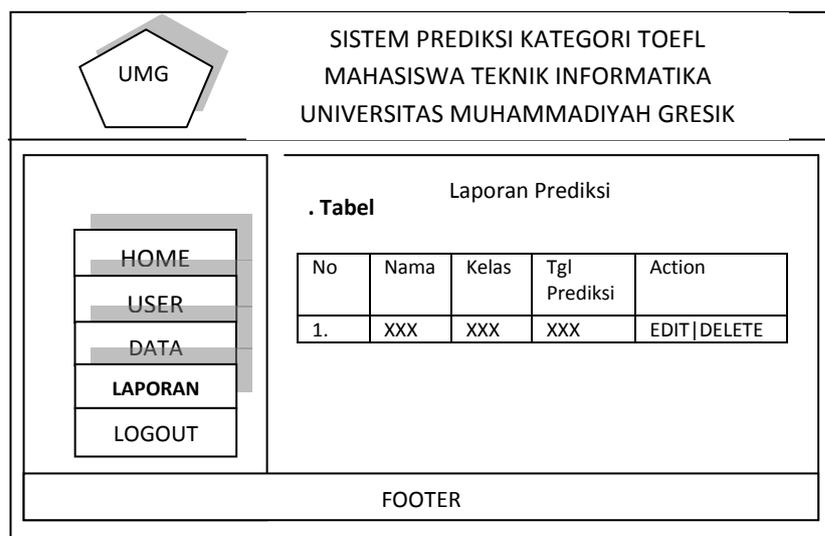
Halaman pada **gambar 3.15** berfungsi untuk menampilkan hasil prediksi setelah mahasiswa menginputkan kusioner yang digunakan sebagai data uji.



**Gambar 3.15** Tampilan Hasil Prediksi

#### i. Halaman Laporan Prediksi

Halaman laporan prediksi pada **gambar 3.16** berfungsi untuk menampilkan semua hasil prediksi yang telah dilakukan oleh mahasiswa.



**Gambar 3.16** Tampilan Laporan Prediksi

### 3.4 Skenario Pengujian

Evaluasi kinerja sistem ini akan dilakukan dengan membandingkan antar hasil prediksi kelas asli dengan hasil prediksi menggunakan aplikasi prediksi kategori nilai TOEFL yang dibagi menjadi dua kelas tinggi dan rendah untuk mahasiswa Teknik Informatika menggunakan *Naive Bayes*. Pengujian menggunakan 7 (tujuh) macam fitur yaitu: nilai *reading1*, *writing1*, *listening1*, *reading2*, *writing2*, *listening2* dan nilai TOEFL setelah D1. Data yang digunakan untuk pengujian prediksi kategori nilai TOEFL adalah data mahasiswa Teknik Informatika Universitas Muhammadiyah Gresik tahun 2008-2011 kelas pagi dengan jumlah 60 data, terbagi menjadi 50 data uji dan 10 data latih.

Setelah dilakukan pemodelan data untuk klasifikasi maka selanjutnya adalah menentukan seberapa akurat data hasil prediksi. Evaluasi dilakukan dengan menguji dataset yang diprediksi secara benar kategori kelas TOEFL-nya dengan menggunakan *Confusion Matrix*. *Confusion Matrix* digunakan untuk menganalisis seberapa baik pengklasifikasian tersebut dapat mengenali tupel dalam kelas-kelas yang berbeda. Berikut tabel *Confusion Matrix* dalam memprediksikan kelas kategori nilai TOEFL seperti pada **tabel 3.15**

**Tabel 3.15** *Confusion Matrix*

		Kelas Hasil Prediksi	
		Tinggi	Rendah
Kelas Asli	Tinggi	True Positive (TP)	False Negative (FN)
	Rendah	False Positive (FP)	True Negative (TN)

Keterangan :

TP : Mahasiswa dengan nilai TOEFL tinggi yang diprediksi secara benar mempunyai nilai TOEFL tinggi

FP : Mahasiswa dengan nilai TOEFL rendah yang diprediksi secara salah mempunyai nilai TOEFL tinggi

TN : Mahasiswa dengan nilai TOEFL rendah yang diprediksi secara benar mempunyai nilai TOEFL rendah

FN : Mahasiswa dengan nilai TOEFL tinggi yang diprediksi secara salah mempunyai nilai TOEFL rendah

Dari tabel *confusion matrix* tersebut, dapat dihitung tingkat akurasi, laju error, sensitivitas dan spesifisitas seperti dibawah ini :

a. Akurasi Pengelompokan

Akurasi digunakan untuk mengukur prosentase pengenalan secara keseluruhan dan dihitung sebagai jumlah data uji yang dikenali dengan benar, dibagi dengan jumlah seluruh data uji. Berikut rumus akurasi dan laju *error* berdasarkan tabel *confusion matrix*.

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \\ &= \frac{TP+TN}{TP+FN+FP+TN} \end{aligned}$$

$$\begin{aligned} \text{Laju error} &= \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{Jumlah prediksi yang dilakukan}} \\ &= \frac{FN+FP}{TP+FN+FP+TN} \end{aligned}$$

b. Sensitivitas dan Spesifisitas

Sensitivitas digunakan untuk mengukur proporsi positif asli yang dikenali (diprediksi) secara benar sebagai positif asli. Sedangkan spesifisitas digunakan untuk mengukur proporsi negatif asli yang dikenali (diprediksi) secara benar sebagai negatif asli. Berikut rumus sensitivitas dan spesifisitas berdasarkan tabel *confusion matrix*.

$$\text{Sensitivitas} = \frac{TP}{TP+FN}$$

$$\text{Spesifisitas} = \frac{TN}{FP+TN}$$