

BAB IV

ANALISA DAN PERANCANGAN

Analisa merupakan suatu proses kajian dalam membahas pokok permasalahan secara lebih mendalam. Tujuan analisa dalam penelitian ini adalah untuk mendapat pemahaman yang tepat dan akurat terhadap masalah, data, proses dan semua hal yang berhubungan dengan penelitian.

4.1 Pengumpulan Data

Data yang digunakan berjumlah 1050 yang diambil dari data sekunder penelitian (Radili, 2016), pada penelitian tersebut data di *download* dalam bentuk pdf yang kemudian di tulis ulang kedalam *notepad* dengan format *.txt*, kemudian data tersebut di tentukan kelompoknya berdasarkan subjek yang tertera pada situs repository ITS dan dipisahkan berdasarkan kelompoknya masing-masing, dan terdapat 8 kelompok. Data kemudian dibagi dengan perbandingan 90:10 untuk data latih dan uji dan data uji, data kemudian dijadikan data latih tidak seimbang dan data latih seimbang untuk mengetahui akurasi mana yang lebih baik.

1. Data latih tidak seimbang.

Data latih tidak seimbang adalah kumpulan dokumen teks yang berekstensi *.txt* yang telah di tentukan kelompoknya terlebih dahulu berdasarkan subject dari situs repository its. Dokumen latih diambil secara manual dengan total jumlah 945 dokumen dengan bidang KCV berjumlah 328 dokumen, KBJ berjumlah 74 dokumen, RPL berjumlah 140 dokumen, IGS berjumlah 73 dokumen, AJK berjumlah 120 dokumen, AP berjumlah 58 dokumen, DTK berjumlah 68 dokumen dan MI berjumlah 84 dokumen. Semua dokumen latih tersebut akan diproses dalam tahap *text mining* yang menghasilkan *output* berupa *fingerprint*, yang menjadi acuan untuk klasifikasi dokumen uji. Dokumen uji yang digunakan berjumlah 105 untuk data uji tidak seimbang dengan bidang KCV berjumlah 28 dokumen, KBJ berjumlah 6 dokumen, RPL berjumlah 10 dokumen, IGS

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berjumlah 10 dokumen, AJK berjumlah 11 dokumen, AP berjumlah 16 dokumen. Dokumen uji dipilih selain dokumen latih yang akan di uji klasifikasi jenis kelompoknya berdasarkan *fingerprint* dokumen latih.

2. Data latih seimbang.

Terdapat 400 dokumen untuk data latih seimbang dengan masing-masing bidang berjumlah 50 dokumen. Semua dokumen latih tersebut akan diproses dalam tahap *text mining* yang menghasilkan *output* berupa *fingerprint*, yang menjadi acuan untuk klasifikasi dokumen uji. Dokumen uji yang digunakan berjumlah 40 untuk data uji seimbang dengan masing-masing bidang berjumlah 5 dokumen telah dipilih secara manual oleh penelitian (Radili, 2016), dokumen uji dipilih selain dokumen latih yang akan di uji klasifikasi jenis kelompoknya berdasarkan *fingerprint* dokumen latih.

4.2 Analisa Proses Text Mining

Yaitu analisa yang dilakukan terhadap dokumen yang dibutuhkan sesuai dengan tahapan *text mining*. Terdapat 4 tahap proses *text mining*, yaitu: *text preprocessing*, *text transformation*, *feature selection*, dan *data mining*.

Pada saat proses pelatihan dan pengujian dokumen abstrak tugas akhir, hal pertama yang dilakukan adalah memasukkan dokumen latih ke sistem dengan cara memasukkan *file* berekstensi *.txt* yang menjadi data koleksi pada sistem. . Kemudian dokumen latih tersebut diproses melalui tahap *text preprocessing*. Yaitu membersihkan teks dengan menghilangkan karakter-karakter yang tidak relevan seperti: tanda baca, spasi, simbol, kata penghubung berdasarkan (*Kamus Besar Bahasa Indonesia*, 2008) dan juga karakter lain serta mengubah semua huruf besar menjadi huruf kecil.

Konfigurasi *Winnowing* merupakan ketentuan aturan nilai yang akan digunakan dalam algoritma *winnowing*. Penelitian yang dilakukan oleh (Ridho, 2013) menyatakan bahwa bilangan basis prima=2 dan jumlah *window*=8 akan menghasilkan ciri dokumen yang lebih baik, dan penelitian yang telah dilakukan (Radili, 2016) menyimpulkan bahwa nilai *k-gram*=8, bilangan prima=2, dan

Hak Cipta Dilindungi Undang-Undang

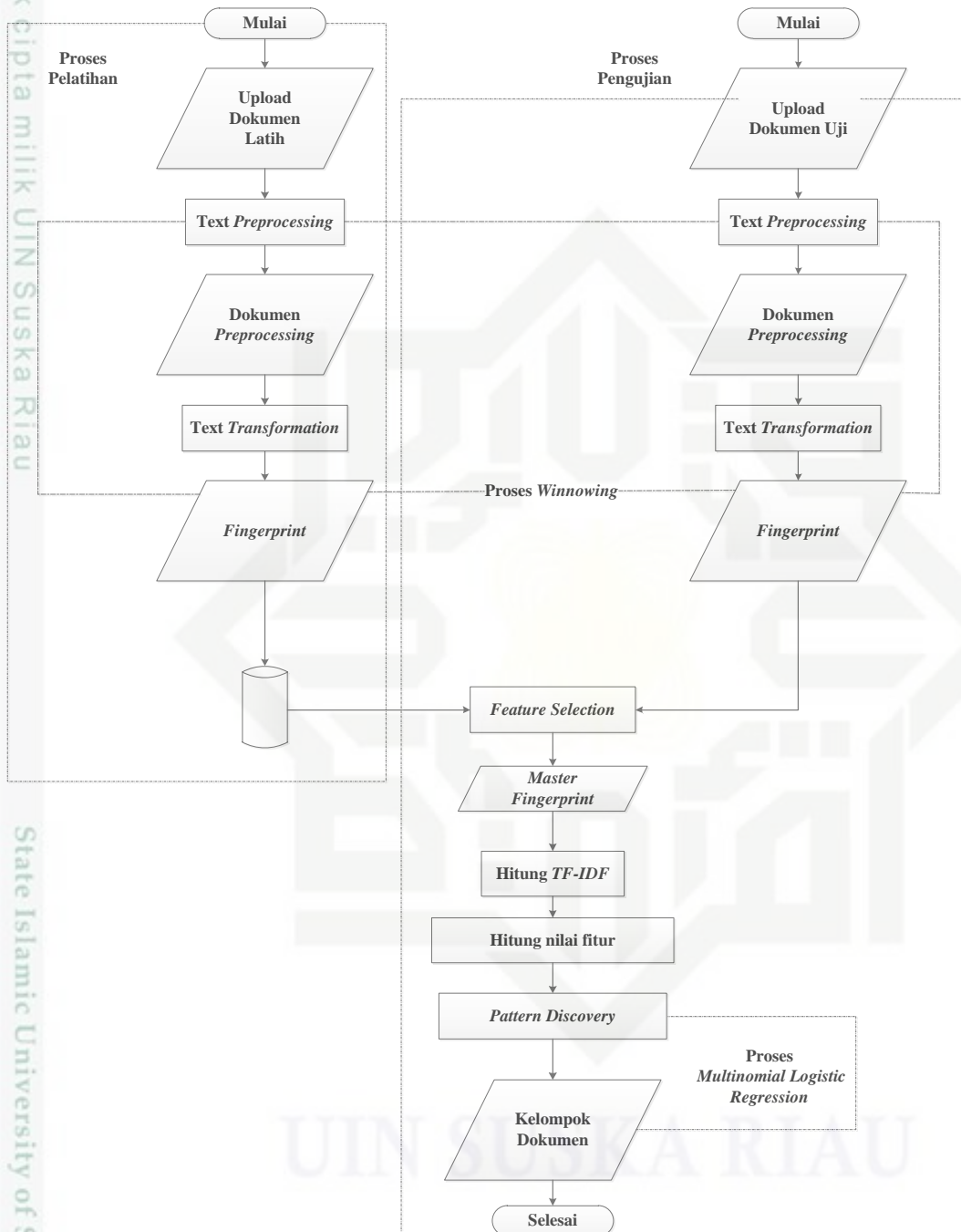
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

jumlah $window=8$ menghasilkan ciri dokumen yang terbaik untuk pengelompokkan dokumen. Oleh karena itu seluruh proses algoritma *winnowing* pada penelitian ini akan menggunakan konfigurasi tersebut.

Kemudian dokumen masuk pada tahap *text transformation*. Pada tahap ini dokumen yang sudah melalui proses *text preprocessing* akan diproses menggunakan metode *k-gram* yaitu membentuk *substring* sepanjang k karakter. Hasil dari proses *k-gram* akan dihitung kedalam nilai *hash*. Perhitungan nilai *hash* dilakukan berdasarkan nilai karakter ASCII. Kemudian nilai *hash* tersebut dikumpulkan menjadi beberapa *window* yang kemudian menjadi hasil *fingerprint*. *Fingerprint* tersebut disimpan kedalam *database*. *Fingerprint* tersebut juga akan melalui tahap *feature selection*. Tahap ini adalah akhir dari proses pelatihan.

Kemudian pada saat proses pengujian dilakukan *feature selection*, yaitu menggabungkan seluruh *fingerprint* dokumen latih dan dokumen uji menjadi *master fingerprint*. Kemudian mencari nilai bobot dengan *TF-IDF* dan menghitung nilai feature dari masing-masing *fingerprint* dalam masing-masing kelompok serta melakukan tahap *pattern discovery* yaitu melakukan proses *multinomial logistic regression* untuk menentukan kelompok dokumen uji.

Berikut adalah *flowchart text mining* pada klasifikasi dokumen :



Gambar 4.1 *Flowchart Text Mining* Klasifikasi Dokumen

4.3.1 Analisa Proses Pelatihan

Merupakan proses memasukkan dokumen latih dan mengambil *fingerprint* dari dokumen latih tersebut dan menyimpan ke dalam *database*. Dokumen latih terbagi dalam 8 kelompok, yaitu : Komputasi Cerdas dan Visualisasi (KCV),

Komputasi Berbasis Jaringan (KBJ), Rekayasa Perangkat Lunak (RPL), Interaksi Grafika dan Seni (IGS), Arsitektur dan Jaringan Komputer (AJK), Algoritma dan Pemrograman (AP), Dasar dan Terapan Komputer (DTK), dan Manajemen Informasi (MI). Beberapa dokumen latihan yang digunakan terdapat dalam tabel 4.1 berikut, untuk lebih lengkapnya dapat dilihat pada **BAB V**:

Tabel 4.1 Dokumen Latihan dan Kelompoknya

No	Nama Dokumen	Kelompok
1	1ITS-Undergraduate-19863-Abstract_id.txt	KCV
2	1ITS-Undergraduate-19868-Abstract_id.txt	KCV
3	2ITS-Undergraduate-20033-5108100004-Abstract_id.txt	KBJ
4	2ITS-Undergraduate-21233-5108100027-Abstract_id.txt	KBJ
5	3ITS-Undergraduate-19857-Abstract_id.txt	RPL
6	3ITS-Undergraduate-19939-5108100196-Abstract_id.txt	RPL
7	4ITS-Undergraduate-19985-5108100101-Abstract_id.txt	IGS
8	4ITS-Undergraduate-21070-5108100603-Abstract_id.txt	IGS
9	5ITS-Undergraduate-21072-Abstract_id.txt	AJK
10	5ITS-Undergraduate-21210-5108100024-Abstract_id.txt	AJK

Dokumen latihan yang digunakan terdapat 945 dokumen, tapi dokumen latihan yang akan dijelaskan pada bab ini adalah dokumen ‘1ITS-Undergraduate-19863-Abstract_id.txt’.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.3.1.1 Text Preprocessing Proses Pelatihan

Adalah tahap pertama dalam proses *text mining*. Berikut proses yang dilakukan pada tahap ini:

1. Memasukkan dokumen latih kedalam sistem yang akan dijadikan rujukan dalam pengelompokan dokumen.
2. Kemudian membersihkan dokumen dengan menghilangkan karakter yang tidak relevan atau tidak berhubungan seperti tanda baca, spasi, kata penghubung dan mengubah huruf besar menjadi huruf kecil. Dapat dilihat pada gambar 4.2 berikut ini adalah alur *text preprocessing* proses pelatihan.



Gambar 4.2 Flowchart Text Preprocessing Pelatihan

Pada Gambar 4.2, proses *preprocessing* dimulai dengan memasukkan dokumen latih ke dalam sistem. Kemudian dilakukan pembersihan pada dokumen dengan menghilangkan tanda baca, spasi, kata penghubung dan mengubah huruf besar menjadi huruf kecil. Tabel 4.2 dan 4.3 berikut ini merupakan proses *preprocessing* pada dokumen.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4.2 Isi Dokumen Sebelum *Preprocessing*

Nama Dokumen	Isi Dokumen
IITS-Undergraduate-19863-Abstract_id.txt	<p data-bbox="627 409 1337 555">OPTIMASI NILAI AMBANG WAVELET BERBASIS LOGIKA FUZZY PADA DENOISING CITRA BERWARNA</p> <p data-bbox="627 689 1337 1989">Proses pengurangan noise dengan metode transformasi wavelet diskrit dapat meyebabkan pergeseran varians (shift variance) dan selektifitas arah (directional selectivity) yang buruk pada citra. Oleh karena itu untuk mengatasi hal tersebut, domain citra diubah menjadi domain wavelet melalui proses Transformasi Wavelet Diskrit Dual-Tree , kemudian dilakukan dekomposisi untuk mendapatkan koefisien aproksimasi (subband Low-Low) dan koefisien detail (Low-High, High-Low, HighHigh). Proses denoising koefisien detail dilakukan pada setiap subbandnya, yang meliputi perhitungan fuzzy, fuzzy membership function dan perhitungan citra bebas noise di setiap subbandnya. Dilakukan proses rekonstruksi untuk menggabungkan semua subband yang telah bebas noise. Pada proses fuzzy membership function untuk mendapatkan nilai keanggotaan didefinisikan dengan proses thresholding. Nilai threshold dipengaruhi oleh nilai konstanta dan nilai estimasi varians noise. Nilai konstanta ditentukan dengan melakukan beberapa percobaan untuk memperoleh nilai terbaik sedangkan nilai estimasi varians noise dilakukan dengan menggunakan median absolute deviation. Hasil percobaan membuktikan bahwa setiap data uji</p>

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	<p>memiliki nilai konstanta dan nilai estimasi varians noise yang berbeda-beda sesuai dengan nilai variansnya. Tingkat keberhasilan diukur dengan menggunakan metode signal to noise ratio vii</p> <p>metode wavelet, logika fuzzy, dekomposisi, rekonstruksi, thresholding , median absolute deviation</p>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabel 4.3 Isi Dokumen Setelah *Preprocessing*

Nama Dokumen	Isi Dokumen
IITS-Undergraduate-19863-Abstract_id.txt	<p>optimasinilaiambangwaveletberbasislogikafuzzypadadenoisingsitraberwarnaprosespengurangannoisemetoderansformasiwaveletdiskritdapatmeyebabkanpergeseran variansshiftvarianceselektifitasarahdirectionalselectivityburukpadacitraolehitudmengatasihaltersebutdomaincitra diubahmenjadi domainwaveletmeiproses transformasi waveletdiskritdualtreedilakukandekomposisimendapat kankoefisienaproksimasisubbandlowlowkoefisiendetail lowhighhighlowhighhighprosesdenoisingskoefisiendet aildilakukanpadasetiapsubbandnyameliputiperhitungan fuzzyfuzzymembershipfunctionperhitungancitrabebasnoisedisetiapsubbandnyadilakukanprosesrekonstruksim enggabungkansemuasubbandtelahbebasnoisepadaprose sfuzzymembershipfunctionmendapatkanilaikeanggota andidefinisikanprosesthresholdingnilaitresholddipeng aruhiolehnilaikonstantanilaiestimasi variansnoisenilaik onstantaditentukanmelakukanbeberapapercobaanmem perolehnilaiterbaikseggannilaiestimasi variansnoisedila kukanmedianabsolutedeviationhasilpercobaanmembuk tikansetiapdjimemilikinilaikonstantanilaiestimasi varia nsnoiseberbedabedasesuainilai variansnyatingkatkeberh</p>

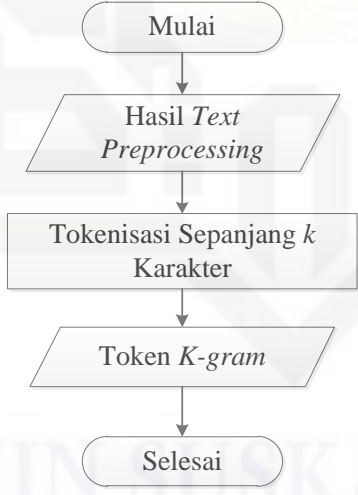
	asilandiukurmetodesignaltonoiserasatioviimetodewavelet logikafuzzydekomposisirekonstruksithresholdingmedi anabsolutedeivation
--	-------------------------------------------------------------------------------------------------------------------------------------

Dapat dilihat perbedaan dokumen sebelum dan sesudah *preprocessing* pada tabel 4.2 dan 4.3.

4.3.1.2 Text Transformation Proses Pelatihan

Berikut adalah penjelasan mengenai tahap *text transformation* pada dokumen latih.

1. Hasil dari *text preprocessing* akan dilakukan pemotongan *string* atau yang disebut tokenisasi. Pada proses ini akan menggunakan metode *k-gram*, yaitu kata akan dipotong menjadi beberapa bagian sepanjang *k* karakter yang telah ditentukan. Gambar 4.3 berikut merupakan alur *text transformation* proses pelatihan.



Gambar 4.3 Flowchart Text Transformation Pelatihan

Hasil dari tokenisasi menggunakan metode *k-gram* dapat dilihat pada Tabel 4.4

Tabel 4.4 Isi Dokumen Setelah *K-gram*

Nama Dokumen	Isi Dokumen
IITS-Undergraduate-19863-Abstract_id.txt	optimasi ptimasin timasini imasinil masinila asinilai sinilaia inilaiam nilaiamb ilaiamba laiamban aiambang iambangw ambangwa mbangwav bangwave angwavel ngwavele gwavelet waveletb aveletbe veletber eletberb letberba etberbas tberbasi berbasis erbasisl rbasislo basislog asislogi sislogik islogika slogikaf logikafu ogikafuz gikafuzz ikafuzzy kafuzzyp afuzzypa

2. Setelah membentuk token, kemudian akan dilakukan perhitungan nilai *hash* dengan rumus *hash* 2.1, dan rumus *Rolling Hash* 2.2 untuk nilai *hash* selanjutnya. Untuk lebih jelasnya dapat dilihat pada *Flowchart* berikut:



Gambar 4.4 Flowchart Proses Perhitungan Nilai Hash Pelatihan

Dapat dilihat pada Tabel 4.4 pemotongan *string* menggunakan nilai $k=8$, dan diketahui bilangan basis prima $b=2$ kemudian dapat dihitung nilai *hash* sebagai berikut:

$$\begin{aligned}
 H_{(\text{optimasi})} &= \text{ascii}(o) * 2^{(7)} + \text{ascii}(p) * 2^{(6)} + \text{ascii}(t) * 2^{(5)} + \text{ascii}(i) * 2^{(4)} + \text{ascii}(m) \\
 &\quad * 2^{(3)} + \text{ascii}(a) * 2^{(2)} + \text{ascii}(s) * 2^{(1)} + \text{ascii}(i) * 2^{(0)} \\
 &= 111 * 128 + 112 * 64 + 116 * 32 + 105 * 16 + 109 * 8 + 97 * 4 + 115 * \\
 &\quad 2 + 105 * 1
 \end{aligned}$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$= 14208 + 7168 + 3712 + 1680 + 872 + 388 + 230 + 105$$

$$= 28363$$

Untuk menentukan nilai *hash* selanjutnya digunakan persamaan *Rolling hash*, sebagai berikut :

$$H_{(ptimasin)} = (28363 - \text{ascii}(o) * 2^{(7)}) * 2 + \text{ascii}(n) * 2^{(0)}$$

$$= (28363 - 111 * 128) * 2 + 110 * 1$$

$$= (28363 - 14208) * 2 + 110 * 1$$

$$= (14155 * 2) + 110$$

$$= 28310 + 110$$

$$= 28420$$

$$H_{(timasini)} = (28420 - \text{ascii}(p) * 2^{(7)}) * 2 + \text{ascii}(i) * 3^{(0)}$$

$$= (28420 - 112 * 128) * 2 + 105 * 1$$

$$= (28420 - 14336) * 2 + 105 * 1$$

$$= (14084 * 2) + 105$$

$$= 28168 + 105$$

$$= 28273$$

Tabel 4.5 Nilai Hash Proses Pelatihan

Nama Dokumen	Isi Dokumen
11TS-Undergraduate-19863-Abstract_id.txt	28363 28420 28273 26958 27133 26467 28199 27067 27352 26641 26512 25479 26245 25707 26700 25597 26214 27697 27350 28430 26497 28276 26442 27125 26717 27683 25785 26590 27435 25789 26595 28465 27587 28396 27469 27412 26530 26813 26858 26421

3. Kemudian nilai *hash* tersebut dibagi ke dalam beberapa *window* dengan ukuran $w=8$ sesuai dengan aturan konfigurasi *winnowing* yang sudah ditetapkan. Proses *window* merupakan proses menggeser *string* nilai *hash* kekanan sebanyak 8 kali nilai *hash* sehingga ditemukannya nilai *hash* yang terakhir. Untuk lebih jelasnya dapat dilihat pada *Flowchart* berikut:



Gambar 4.5 Flowchart Proses Pembentukan Window

Tabel 4.6 Pembentukan Window Proses Pelatihan

Nama Dokumen	Window
IITS-	[28363 28420 28273 26958 27133 26467 28199 27067]
Undergraduate-	[28420 28273 26958 27133 26467 28199 27067 27352]
19863-	[28273 26958 27133 26467 28199 27067 27352 26641]
Abstract_id.txt	[26958 27133 26467 28199 27067 27352 26641 26512]
	[27133 26467 28199 27067 27352 26641 26512 25479]
	[26467 28199 27067 27352 26641 26512 25479 26245]
	[28199 27067 27352 26641 26512 25479 26245 25707]
	[27067 27352 26641 26512 25479 26245 25707 26700]
	[27352 26641 26512 25479 26245 25707 26700 25597]
	[26641 26512 25479 26245 25707 26700 25597 26214]
	[26512 25479 26245 25707 26700 25597 26214 27697]
	[25479 26245 25707 26700 25597 26214 27697 27350]
	[26245 25707 26700 25597 26214 27697 27350 28430]
	[25707 26700 25597 26214 27697 27350 28430 26497]
	[26700 25597 26214 27697 27350 28430 26497 28276]
	[25597 26214 27697 27350 28430 26497 28276 26442]
	[26214 27697 27350 28430 26497 28276 26442 27125]
	[27697 27350 28430 26497 28276 26442 27125 26717]
	[27350 28430 26497 28276 26442 27125 26717 27683]
	[28430 26497 28276 26442 27125 26717 27683 25785]
	[26497 28276 26442 27125 26717 27683 25785 26590]
	[28276 26442 27125 26717 27683 25785 26590 27435]
	[26442 27125 26717 27683 25785 26590 27435 25789]
	[27125 26717 27683 25785 26590 27435 25789 26595]
	[26717 27683 25785 26590 27435 25789 26595 28465]
	[27683 25785 26590 27435 25789 26595 28465 27587]
	[25785 26590 27435 25789 26595 28465 27587 28396]

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[26590 27435 25789 26595 28465 27587 28396 27469] [27435 25789 26595 28465 27587 28396 27469 27412] [25789 26595 28465 27587 28396 27469 27412 26530] [26595 28465 27587 28396 27469 27412 26530 26813] [28465 27587 28396 27469 27412 26530 26813 26858] [27587 28396 27469 27412 26530 26813 26858 26421]

4.3.1.3 Feature Selection

Kemudian berikutnya adalah tahap *feature selection*, yang dilakukan pada tahap ini adalah :

1. Memilih nilai *hash* minimum pada posisi paling kanan dari setiap *window* yang telah terbentuk untuk dijadikan *fingerprint*. Untuk lebih jelasnya dapat dilihat pada *Flowchart* pembentukan *fingerprint* berikut ini:

Tabel 4.7 Pembentukan *Fingerprint* Proses Pelatihan

Nama Dokumen	Window
IITS- Undergraduate- 19863- Abstract_id.txt	[28363 28420 28273 26958 27133 26467 28199 27067]
	[28420 28273 26958 27133 26467 28199 27067 27352]
	[28273 26958 27133 26467 28199 27067 27352 26641]
	[26958 27133 26467 28199 27067 27352 26641 26512]
	[27133 26467 28199 27067 27352 26641 26512 25479]
	[26467 28199 27067 27352 26641 26512 25479 26245]
	[28199 27067 27352 26641 26512 25479 26245 25707]
	[27067 27352 26641 26512 25479 26245 25707 26700]
	[27352 26641 26512 25479 26245 25707 26700 25597]
	[26641 26512 25479 26245 25707 26700 25597 26214]
	[26512 25479 26245 25707 26700 25597 26214 27697]
	[25479 26245 25707 26700 25597 26214 27697 27350]
	[26245 25707 26700 25597 26214 27697 27350 28430]
	[25707 26700 25597 26214 27697 27350 28430 26497]
	[26700 25597 26214 27697 27350 28430 26497 28276]
	[25597 26214 27697 27350 28430 26497 28276 26442]
	[26214 27697 27350 28430 26497 28276 26442 27125]
	[27697 27350 28430 26497 28276 26442 27125 26717]
	[27350 28430 26497 28276 26442 27125 26717 27683]
	[28430 26497 28276 26442 27125 26717 27683 25785]
[26497 28276 26442 27125 26717 27683 25785 26590]	
[28276 26442 27125 26717 27683 25785 26590 27435]	
[26442 27125 26717 27683 25785 26590 27435 25789]	
[27125 26717 27683 25785 26590 27435 25789 26595]	
[26717 27683 25785 26590 27435 25789 26595 28465]	

[27683 <u>25785</u> 26590 27435 25789 26595 28465 27587]
[<u>25785</u> 26590 27435 25789 26595 28465 27587 28396]
[26590 27435 <u>25789</u> 26595 28465 27587 28396 27469]
[27435 <u>25789</u> 26595 28465 27587 28396 27469 27412]
[<u>25789</u> 26595 28465 27587 28396 27469 27412 26530]
[26595 28465 27587 28396 27469 27412 <u>26530</u> 26813]
[28465 27587 28396 27469 27412 <u>26530</u> 26813 26858]
[27587 28396 27469 27412 26530 26813 26858 <u>26421</u>]
.....

Terlihat pada tabel 4.7 pemilihan *fingerprint* dari kumpulan nilai *hash* terkecil pada tiap *window* (yang digaris bawahi). Kemudian akan dipilih nilai *hash* terkecil pada tiap *window* dan didapat hasil sebagai berikut:

Tabel 4.8 Fingerprint IITS-Undergraduate-19863-Abstract_id.txt

Nama Dokumen	Window
IITS-Undergraduate-19863-Abstract_id.txt	26467 25479 25597 26214 26442 25785 25789 26530 26421 25095 25451 26180 26333 25422 26109 26473 26674 26759 25951 26311 25861 25967 26233 26348 26427 26598 26505 26485 26576 25831 25448 25141

2. Hasil *fingerprint* akan disimpan ke dalam *database* beserta informasi dokumen, yaitu nama dokumen serta kelompok dokumen yang akan digunakan sebagai dokumen latih untuk menguji dokumen uji.
- Jika semua *fingerprint* dokumen latih telah disimpan ke dalam *database*, selanjutnya dilakukan klasifikasi terhadap data baru yang ingin di uji. Dokumen uji tersebut tetap harus melewati *text preprocessing* dan *text transformation* sampai menghasilkan *fingerprint* dari dokumen uji tersebut.

4.3.2 Analisa Proses Pengujian

Pada proses ini yang di lakukan adalah memasukkan dokumen uji dan mengambil *fingerprint* dari dokumen uji tersebut dan mencocokkannya dengan *fingerprint* yang ada di dalam *database*. Dokumen uji yang digunakan berjumlah 105 dokumen, dokumen uji yang akan dijelaskan pada bab ini adalah dokumen ‘IITS-Undergraduate-25671-5108100089-Abstract_id’. Untuk melihat lebih

Hak Cipta Dilindungi Undang-Undang

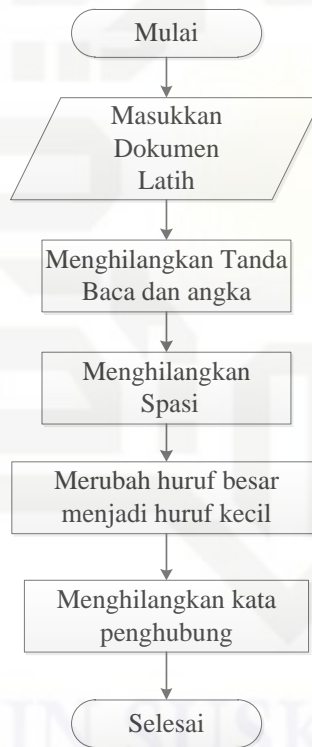
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

lengkap dokumen uji yang digunakan pada penelitian ini dapat dilihat pada **LAMPIRAN B**.

4.3.2.1 Text Preprocessing Proses Pengujian

Adalah tahap pertama dalam proses *text mining*. Berikut proses yang dilakukan pada tahap ini:

1. Memasukkan dokumen uji kedalam sistem.
2. Kemudian membersihkan dokumen dengan menghilangkan karakter yang tidak relevan atau tidak berhubungan seperti tanda baca, spasi, kata penghubung dan mengubah huruf besar menjadi huruf kecil. Dapat dilihat pada gambar 4.6 berikut ini adalah alur *text preprocessing* proses pengujian.



Gambar 4.6 Flowchart Text Preprocessing Pengujian

Pada Gambar 4.6, proses *preprocessing* dimulai dengan memasukkan dokumen uji ke dalam sistem. Kemudian dilakukan pembersihan pada dokumen dengan menghilangkan tanda baca, spasi, kata penghubung dan mengubah huruf besar menjadi huruf kecil. Tabel 4.9 dan 4.10 berikut ini merupakan proses *preprocessing* pada dokumen.

Tabel 4.9 Isi Dokumen Uji Sebelum *Preprocessing*

Nama Dokumen	Isi Dokumen
IITS-Undergraduate-25671-5108100089-Abstract_id	<p data-bbox="627 409 1337 555">IMPLEMENTASI METODE KERNEL PARTIAL LEAST SQUARE UNTUK KLASIFIKASI KANKER OVARIUM</p> <p data-bbox="627 633 1337 1989">Kanker ovarium merupakan kanker yang berkembang dengan sangat cepat. Perkembangan kanker dari stadium awal hingga stadium lanjut dapat terjadi dalam kurun waktu 1 tahun. Kanker ovarium merupakan penyebab kematian terbesar kelima di Indonesia. Pertumbuhan jumlah penderita kanker ini terus bertambah setiap tahunnya. Untuk itulah, sebuah perangkat lunak yang mampu melakukan diagnosa dini kanker dibutuhkan untuk mengurangi resiko kematian. Sebuah studi yang dilakukan oleh Tang et. al (Tang, Li, Xiong, & Chen, 2010) mengusulkan sebuah model yang memanfaatkan Kernel Partial Least Square (KPLS) untuk memprediksi adanya kanker. Tugas Akhir ini bertujuan untuk mengembangkan perangkat lunak dengan memanfaatkan metode KPLS untuk memprediksi penyakit kanker ovarium menggunakan data spektometri massa ovarium SELDI-TOF High Resolution seperti yang telah diusulkan oleh (Tang, Li, Xiong, & Chen, 2010). KPLS sendiri merupakan pengembangan dari Partial Least Square dengan melibatkan fungsi kernel di dalam prosesnya. Fungsi kernel yang digunakan dalam implementasi metode ini adalah fungsi kernel polinomial. KPLS terbagi ke dalam dua tahap. Tahap pertama adalah tahap latihan</p>

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	<p>untuk mencari koefisien regresi. Tahap kedua adalah tahap uji untuk memprediksi label data uji berdasarkan koefisien regresi dari tahap latih. Uji coba perangkat lunak dilakukan dengan memprediksi kanker pada empat dataset yang berbeda. Berdasarkan uji coba, KPLS menghasilkan rata-rata akurasi terbaik sebesar 93,53%, sensitivitas sebesar 89,81% dan spesifisitas sebesar 98,89% pada data Set A. Berdasarkan performa uji coba klasifikasi data, dapat ditarik kesimpulan bahwa perangkat lunak yang memanfaatkan KPLS mampu melakukan klasifikasi kanker ovarium pada data ekspresi protein dalam format SELDI-TOF.</p> <p>kanker ovarium, SELDI-TOF kernel, klasifikasi, KPLS,</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabel 4.10 Isi Dokumen Uji Setelah *Preprocessing*

Nama Dokumen	Isi Dokumen
<p>11TS-Undergraduate-25671-5108100089-Abstract_id</p>	<p>implementasi metode kernel partial least square klasifikasi kanker ovarium kanker ovarium merupakan kanker berke mbang sangat cepat perkembangan kanker dari stadium awal stadium lanjut dapat terjadi dalam kurun waktu tahunan kanker ovarium merupakan penyebab kematian terbesar kelima di Indonesia pertumbuhan jumlah penderita kanker ini terus bertambah setiap tahunnya yaitu lah sebuah perangkat lunak mampu melakukan diagnosis dini kanker dibutuhkan men gurangi resiko kematian sebuah studi dilakukan oleh tang et al. Tanglixiongchen mengusulkan sebuah model memanfaatkan kernel partial least square kpls memprediksi kanker tugas akhir ini bertujuan mengembangkan perangkat lunak</p>

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	<p>kmemanfaatkanmetodekplsmemprediksipenyakitkanke rovariumdataspektometrimassaovariumselditofhighres olutiontelahdiusulkanohtanglixiongchenkplssendirim erupakanpengembangandaripartialleastsquaremelibatk anfungsikerneldidalamprosesnyafungsikerneldigunaka ndalamimplementasimetodeinifungsikernelpolinomial kplsterbagikedalamduatahaptahappertamatahaplatihme ncarikoeffisienregresitahapkeduatahapujimemprediksila beldjiberdasarkankoeffisienregresidaritahaplatihujicoba perangkatlunakdilakukanmemprediksikanterpadaempa tdatasetberbedaberdasarkanujicobakplsmenghasilkanra tarataakurasiterbaiksebesarsensitivitassebesarspesifisit assebesarpadadatasetaberdasarkanperformaujicobaklas ifikasadapatditarikkesimpulanperangkatlunakmem anfaatkankplsmampumelakukanklasifikasikanrovari umpadadataekspresiproteindalamformatselditofkanker ovariumselditofkernelklasifikasikpls</p>
--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

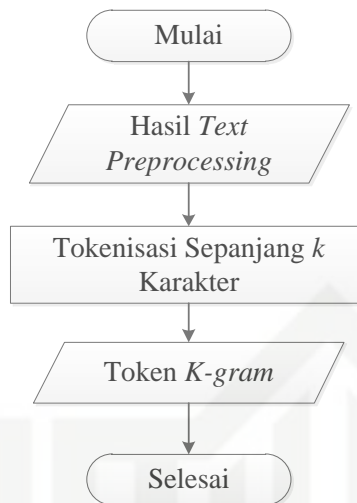
Dapat dilihat perbedaan dokumen sebelum dan sesudah *preprocessing* pada tabel 4.9 dan 4.10.

4.3.2.2 Text Transformation Proses Pengujian

Berikut adalah penjelasan mengenai tahap *text transformation* pada dokumen uji.

1. Hasil dari *text preprocessing* akan dilakukan pemotongan *string* atau yang disebut tokenisasi. Pada proses ini akan menggunakan metode *k-gram*, yaitu kata akan dipotong menjadi beberapa bagian sepanjang *k* karakter yang telah ditentukan. Gambar 4.7 berikut merupakan alur *text transformation* proses pengujian.

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.7 Flowchart Text Transformation Pengujian

Hasil dari tokenisasi menggunakan metode *k-gram* dapat dilihat pada Tabel 4.11

Tabel 4.11 Isi Dokumen Uji Setelah *K-gram*

Nama Dokumen	Isi Dokumen
1ITS-Undergraduate-25671-5108100089-Abstract_id	implemen mplement plementa lementas ementasi mentasim entasime ntasimet tasimeto asimetod simetode imetodek metodeke etodeker todekern odekerne dekernel ekernelp kernelpa ernelpar rnelpart

2. Setelah membentuk token, kemudian akan dilakukan perhitungan nilai *hash* dengan rumus *hash 2.1*, dan rumus *Rolling Hash 2.2* untuk nilai *hash* selanjutnya. Untuk lebih jelasnya dapat dilihat pada *Flowchart* berikut:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.8 Flowchart Proses Perhitungan Nilai Hash Pengujian

Dapat dilihat pada Tabel 4.11 pemotongan *string* menggunakan nilai $k=8$, dan diketahui bilangan basis prima $b=2$ sehingga dapat kita hitung nilai *hash* sebagai berikut:

$$\begin{aligned}
 H_{(\text{implemen})} &= \text{ascii}(i) * 2^{(7)} + \text{ascii}(m) * 2^{(6)} + \text{ascii}(p) * 2^{(5)} + \text{ascii}(l) * 2^{(4)} + \text{ascii}(e) * 2^{(3)} + \text{ascii}(m) * 2^{(2)} + \text{ascii}(e) * 2^{(1)} + \text{ascii}(n) * 2^{(0)} \\
 &= 105 * 128 + 109 * 64 + 112 * 32 + 108 * 16 + 101 * 8 + 109 * 4 + 101 * 2 + 110 * 1 \\
 &= 13440 + 6976 + 3584 + 1728 + 808 + 436 + 202 + 110 \\
 &= 27284
 \end{aligned}$$

Untuk mencari nilai *hash* selanjutnya digunakan persamaan *Rolling hash*, sebagai berikut :

$$\begin{aligned}
 H_{(\text{plement})} &= (27284 - \text{ascii}(i) * 2^{(7)}) * 2 + \text{ascii}(t) * 2^{(0)} \\
 &= (27284 - 105 * 128) * 2 + 116 * 1 \\
 &= (27284 - 13440) * 2 + 116 * 1 \\
 &= (13844 * 2) + 116 \\
 &= 27688 + 116 \\
 &= 27804
 \end{aligned}$$

$$\begin{aligned}
 H_{(\text{plementa})} &= (27804 - \text{ascii}(m) * 2^{(7)}) * 2 + \text{ascii}(a) * 3^{(0)} \\
 &= (27804 - 109 * 128) * 2 + 97 * 1 \\
 &= (27804 - 13952) * 2 + 97 * 1
 \end{aligned}$$

$$\begin{aligned}
 &= (13852 * 2) + 97 \\
 &= 27704 + 97 \\
 &= 27801
 \end{aligned}$$

Tabel 4.12 Nilai Hash Proses Pengujian

Nama Dokumen	Isi Dokumen
1HTS-Undergraduate-25671-5108100089-Abstract_id	27284 27804 27801 27045 26547 27347 26891 28042 28035 26474 28217 27101 27423 27056 28366 27137 25966 26444 27129 26980 28220 27361 26659 27570 27600 26629 28523 27977 26374 25983 27247 26963

3. Kemudian nilai *hash* tersebut dibagi ke dalam beberapa *window* dengan ukuran $w=8$ sesuai dengan aturan konfigurasi *winning* yang sudah ditetapkan. Proses *window* merupakan proses menggeser *string* nilai *hash* kekanan sebanyak 8 kali nilai *hash* sehingga ditemukannya nilai *hash* yang terakhir. Untuk lebih jelasnya dapat dilihat pada *Flowchart* berikut:



Gambar 4.9 Flowchart Proses Pembentukan Window Pengujian

Tabel 4.13 Pembentukan Window Proses Pengujian

Nama Dokumen	Window
1HTS-Undergraduate-25671-5108100089-	[27284 27804 27801 27045 26547 27347 26891 28042] [27804 27801 27045 26547 27347 26891 28042 28035] [27801 27045 26547 27347 26891 28042 28035 26474] [27045 26547 27347 26891 28042 28035 26474 28271] [26547 27347 26891 28042 28035 26474 28271 27101]

Abstract_id	[27347 26891 28042 28035 26474 28271 27101 27423] [26891 28042 28035 26474 28271 27101 27423 27056] [28042 28035 26474 28271 27101 27423 27056 28366] [28035 26474 28271 27101 27423 27056 28366 27137] [26474 28271 27101 27423 27056 28366 27137 25966] [28271 27101 27423 27056 28366 27137 25966 26444] [27101 27423 27056 28366 27137 25966 26444 27129] [27423 27056 28366 27137 25966 26444 27129 26980] [27056 28366 27137 25966 26444 27129 26980 28220]
-------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

4.3.2.3 Feature Selection

Kemudian berikutnya adalah tahap *feature selection*, yang dilakukan pada tahap ini adalah :

1. Memilih nilai *hash* minimum pada posisi paling kanan dari setiap *window* yang telah terbentuk untuk dijadikan *fingerprint*. Untuk lebih jelasnya dapat dilihat pada *Flowchart* pembentukan *fingerprint* berikut ini:

Tabel 4.14 Pembentukan *Fingerprint* Proses Pengujian

Nama Dokumen	Window
IITS- Undergraduate- 25671-5108100089- Abstract_id	[27284 27804 27801 27045 <u>26547</u> 27347 26891 28042]
	[27804 27801 27045 <u>26547</u> 27347 26891 28042 28035]
	[27801 27045 26547 27347 26891 28042 28035 <u>26474</u>]
	[27045 26547 27347 26891 28042 28035 <u>26474</u> 28271]
	[26547 27347 26891 28042 28035 <u>26474</u> 28271 27101]
	[27347 26891 28042 28035 <u>26474</u> 28271 27101 27423]
	[26891 28042 28035 <u>26474</u> 28271 27101 27423 27056]
	[28042 28035 <u>26474</u> 28271 27101 27423 27056 28366]
	[28035 <u>26474</u> 28271 27101 27423 27056 28366 27137]
	[26474 28271 27101 27423 27056 28366 27137 <u>25966</u>]
	[28271 27101 27423 27056 28366 27137 <u>25966</u> 26444]
	[27101 27423 27056 28366 27137 <u>25966</u> 26444 27129]
	[27423 27056 28366 27137 <u>25966</u> 26444 27129 26980]
	[27056 28366 27137 <u>25966</u> 26444 27129 26980 28220]

.....	

Terlihat pada tabel 4.14 pemilihan *fingerprint* dari kumpulan nilai *hash* terkecil pada tiap *window* (yang digaris bawahi). Kemudian akan dipilih nilai *hash* terkecil pada tiap *window* dan didapat hasil sebagai berikut:

Tabel 4.15 Fingerprint IITS-Undergraduate-25671-5108100089

Nama Dokumen	Window
IITS-Undergraduate-25671-5108100089-Abstract_id	26547 26474 25966 26444 26629 26374 25983 26375 26501 27117 26939 26345 26379 26185 26568 27268 27132 25699 26056 26196 25899 25905 25447 25903

4.3.2.4 Pembobotan TF-IDF

Pada tahap ini akan diimplementasikan pembobotan *TF-IDF* untuk mencari nilai bobot suatu *fingerprint*, sesuai dengan persamaan *TF-IDF* (2.4). Beberapa hal yang dilakukan pada tahap ini adalah :

1. Menghitung banyaknya masing-masing *fingerprint* dalam dokumen uji terkait. Tabel 4.16 berikut beberapa frekuensi dari masing-masing *fingerprint* di dokumen uji IITS-Undergraduate-25671-5108100089.

Tabel 4.16 Frekuensi dari masing-masing Fingerprint di dokumen uji

No	Fingerprint	Frekuensi
1	26547	4
2	26474	4
3	25966	2
4	26444	2
5	26629	4
6	26374	6

Terdapat 274 total jumlah *fingerprint* dengan total jumlah frekuensi sebanyak 696 di dokumen uji IITS-Undergraduate-25671-5108100089

2. Menghitung nilai *Term Frequency*, yaitu membagi jumlah muncul suatu *fingerprint* dalam dokumen uji dengan total jumlah *fingerprint* di dokumen uji terkait, diketahui total jumlah frekuensi *fingerprint* di dokumen IITS-Undergraduate-25671-5108100089 ada sebanyak 696 frekuensi *fingerprint*.

$$\text{Nilai TF fingerprint } 26547 = 4/696$$

$$= 0.0057471264367816$$

Begitu juga untuk *fingerprint* selanjutnya dapat dilihat pada tabel 4.17.

Tabel 4.17 Nilai Term Frequency dari masing-masing Fingerprint di dokumen uji

No	Fingerprint	Term frequency
1	26547	0.0057471264367816
2	26474	0.0057471264367816
3	25966	0.0028735632183908
4	26444	0.0028735632183908
5	26629	0.0057471264367816
6	26374	0.0086206896551724

Perhitungan tersebut berlaku juga untuk keseluruhan 274 fingerprint.

- Menghitung banyaknya dokumen di dokumen latih.
Total data latih ada sebanyak 945.
- Menghitung *inverse document frequency* dari masing-masing fingerprint.

Dengan persamaan (2.4), yaitu:

$$idf =$$

$$\text{Log}\left(\frac{\text{Total banyaknya dokumen di dokumen latih}+1}{\text{Banyaknya dokumen yang memiliki kemunculan Fingerprint terkait di dokumen latih}+1}\right)$$

$$Idf = \log (945+1 / 60+1) = 1.19055859278$$

- Menghitung nilai *TF-IDF* dari masing-masing fingerprint dan menyimpannya kedalam array, yaitu mengkalikan $tf \times idf$.

$$TF-IDF \text{ Fingerprint } 26547 = 0.0057471264367816 \times 1.19055859278$$

$$TF-IDF \text{ Fingerprint } 26547 = 0.00688092274$$

Perhitungan tersebut juga berlaku pada keseluruhan 274 fingerprint.

Beberapa nilai *TF-IDF* selanjutnya bisa dilihat pada tabel 4.18 berikut.

Tabel 4.18 TF-IDF dari masing-masing Fingerprint di dokumen uji

No	Fingerprint	TF-IDF
1	26547	0.00684229076
2	26474	0.00684229076
3	25966	0.00342114638
4	26444	0.00342114638
5	26629	0.00684229076
6	26374	0.01026343614

6. Nilai *TF-IDF* dikategorikan sebagai bobot dari masing-masing *fingerprint*

4.3.2.5 Perhitungan Nilai Fitur

Pada tahap ini akan menghitung nilai fitur dari masing-masing *fingerprint* dalam masing-masing kategori

1. Tahap pertama untuk mencari nilai fitur adalah menghitung jumlah kemunculan *fingerprint* dari masing-masing kategori atau kelompok. Tabel 4.19 berikut bebarapa jumlah kemunculan *fingerprint* dari masing-masing kategori.

Tabel 4.19 Jumlah kemunculan *fingerprint* dari masing-masing kelompok

No	<i>Fingerprint</i>	Frekuensi	Kelompok
1	25647	550	1
2	26474	157	1
3	25966	145	1
4	26444	123	1
5	26629	59	1
6	26374	109	1

Terdapat 274 total *fingerprint* dengan masing-masing *fingerprint* dan kelompok

2. Menghitung total jumlah kemunculan *fingerprint* dalam masing-masing kelompok.

Tabel 4.20 Total jumlah kemunculan *fingerprint* dari masing-masing kelompok

No	Kelompok	Total Frekuensi
1	Komputasi Cerdas Visualisasi (KCV)	189666
2	Komputasi Berbasis Jaringan (KBJ)	48617
3	Rekayasa Perangkat Lunak (RPL)	86752
4	Interaksi Grafika dan Seni (IGS)	39533
5	Arsitektur dan Jaringan Komputer (AJK)	72330
6	Algoritma dan Pemrograman (AP)	35796
7	Dasar dan Terapan Komputasi (DTK)	34751
8	Manajemen Informasi (MI)	51184

3. Menghitung nilai fitur dengan cara membagi jumlah kemunculan *fingerprint* terkait di data latih dalam masing-masing kelompok dengan total jumlah kemunculan *fingerprint* dari masing-masing kelompok.

Contoh :

$$\begin{aligned} \text{Nilai fitur fingerprint 26547} &= 550/189666 \\ &= 0.0028154756255734 \end{aligned}$$

Begitu juga untuk *fingerprint* selanjutnya dapat dilihat pada tabel 4.21.

Perhitungan tersebut juga berlaku pada keseluruhan 274 *fingerprint*.

Tabel 4.21 Beberapa nilai fitur masing-masing *fingerprint* dokumen uji

Kelompok	No.	Fingerprint	Nilai Fitur
Komputasi Cerdas Visualisasi (KCV)	1	26547	0.0028778333577513
	2	26474	0.0008267230373176
	3	25966	0.0007639339458758
	4	26444	0.0006488206115657
	5	26629	0.0003139454572092
	6	26374	0.0005755666715502
Komputasi Berbasis Jaringan (KBJ)	1	26547	0.0021803072999157
	2	26474	0.0006946856547412
	3	25966	0.0004495024824796
	4	26444	0.0008172772408720
	5	26629	0.0004495024824796
	6	26374	0.0003269108963488
Rekayasa Perangkat Lunak (RPL)	1	26547	0.002330336139996
	2	26474	0.0004177017609426
	3	25966	0.0006375447930177
	4	26444	0.0004177017609426
	5	26629	0.0003517488513201
	6	26374	0.0001538901224525
Interaksi Grafika dan Seni (IGS)	1	26547	0.0038196124366581
	2	26474	0.0006620661556874

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	3	25966	0.0003055689949326
	4	26444	0.0008148506531537
	5	26629	0.0008148506532537
	6	26374	0.0003055689949326
Arsitektur dan Jaringan Komputer (AJK)	1	26547	0.0034815115981887
	2	26474	0.0002037126632884
	3	25966	0.0003889212989971
	4	26444	0.0003611412062116
	5	26629	0.0002222407422840
	6	26374	0.0003611412062116
Algoritma dan Pemrograman (AP)	1	26547	0.0018874029335634
	2	26474	0.0005392579810181
	3	25966	0.0004853321829163
	4	26444	0.0010245901633944
	5	26629	0.0003235547886108
	6	26374	0.0002696289905090
Dasar dan Terapan Komputasi (DTK)	1	26547	0.0028164176541302
	2	26474	0.0007556242486690
	3	25966	0.0008930104756998
	4	26444	0.0008930104756998
	5	26629	0.0003434655675768
	6	26374	0.0005495449081229
Manajemen Informasi (MI)	1	26547	0.0027707317073171
	2	26474	0.0002731707317073
	3	25966	0.0002341463414634
	4	26444	0.0007414634146341
	5	26629	0.0005853658536585
	6	26374	0.0003512195121951

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.3.2.4 Pattern Discovery

Pada tahap ini akan diimplementasikan metode *multinomial logistic regression* untuk mencari *probabilitas* nilai kedekatan atau kesamaan dokumen untuk mendapatkan kelompok, sesuai dengan persamaan (2.5).

Berikut beberapa hal yang dilakukan pada tahap ini:

1. Mengkalikan *weight (TF-IDF)* dan *feature* dari masing-masing *fingerprint* pada masing-masing kelompok

Contoh :

Fingerprint 26547 mempunyai:

bobot *TF-IDF* = 0.00684229076

nilai fitur = 0.0028778333577513

kelompok = komputasi cerdas dan visualisasi

Maka $0.00684229076 \times 0.0028778333577513 = 0.00001969097$

Perhitungan tersebut juga berlaku pada keseluruhan 274 *fingerprint*.

Begitu juga untuk *fingerprint* selanjutnya, dapat dilihat beberapa pada tabel 4.22.

Tabel 4.22 Beberapa nilai hasil perkalian bobot x fitur pada dokumen uji

Kelompok	No.	<i>Fingerprint</i>	Nilai bobot x fitur
Komputasi Cerdas Visualisasi (KCV)	1	26547	0.00001969097
	2	26474	0.00000565667
	3	25966	0.00000261352
	4	26444	0.00000221971
	5	26629	0.0000021481
	6	26374	0.00000590729
Komputasi Berbasis Jaringan (KBJ)	1	26547	0.00001491829
	2	26474	0.00000475324
	3	25966	0.00000153781
	4	26444	0.00000279602
	5	26629	0.00000307562
	6	26374	0.00000335522

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Rekayasa Perangkat Lunak (RPL)	1	26547	0.00001594483
	2	26474	0.00000285803
	3	25966	0.00000218113
	4	26444	0.00000142901
	5	26629	0.00000240676
	6	26374	0.00000157944
Interaksi Grafika dan Seni (IGS)	1	26547	0.00002613489
	2	26474	0.00000453004
	3	25966	0.00000209079
	4	26444	0.00000557544
	5	26629	0.00000278772
	6	26374	0.00000313618
Arsitektur dan Jaringan Komputer (AJK)	1	26547	0.00002382151
	2	26474	0.00000139386
	3	25966	0.00000266111
	4	26444	0.00000247103
	5	26629	0.00000076031
	6	26374	0.00000370654
Algoritma dan Pemrograman (AP)	1	26547	0.00001291415
	2	26474	0.00000368975
	3	25966	0.00000332078
	4	26444	0.00000701054
	5	26629	0.00000110692
	6	26374	0.00000276731
Dasar dan Terapan Komputasi (DTK)	1	26547	0.00001927074
	2	26474	0.0000051702
	3	25966	0.00000611023
	4	26444	0.00000611023
	5	26629	0.00000117504
	6	26374	0.00000564021

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Manajemen Informasi (MI)	1	26547	0.00001895815
	2	26474	0.00000186911
	3	25966	0.00000160209
	4	26444	0.0000050733
	5	26629	0.00000200626
	6	26374	0.00000360471

2. Menjumlahkan keseluruhan *weight x feature* dari masing-masing *fingerprint* di masing-masing kelompok

Contoh:

Kelompok Komputasi Cerdas Visualisasi (KCV)

Nilai *Fingerprint* awal + nilai *fingerprint* kedua + + nilai *fingerprint* terakhir

$$0.00001969097 + 0.00000565667 + + 0.00000064443$$

$$= 0.00055046379137724$$

Begitu juga untuk perhitungan kelompok berikutnya, dapat dilihat pada tabel 4.23.

Tabel 4.23 Hasil perhitungan *multinomial logistic regression (MLR)* dengan kelas latih

No	Kelompok	Hasil Perhitungan <i>MLR</i>
1	Komputasi Cerdas Visualisasi (KCV)	0.00055046379137724
2	Komputasi Berbasis Jaringan (KBJ)	0.00048782067843249
3	Rekayasa Perangkat Lunak (RPL)	0.0005348111917426
4	Interaksi Grafika dan Seni (IGS)	0.00051719678662911
5	Arsitektur dan Jaringan Komputer (AJK)	0.00049860871557478
6	Algoritma dan Pemrograman (AP)	0.00050720050319699
7	Dasar dan Terapan Komputasi (DTK)	0.00053444904667248
8	Manajemen Informasi (MI)	0.00052467027824481

3. Mencari nilai tertinggi dari masing-masing kalkulasi kelompok tersebut.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Mendapatkan nama kelompok yang memiliki nilai *probabilitas* tertinggi, sehingga diketahui dokumen uji ‘1ITS-Undergraduate-25671-5108100089-Abstract_id’ termasuk kelompok Komputasi Cerdas dan Visualisasi (KCV) dengan nilai *probabilitas* tertinggi.

4.3 Analisa Sistem

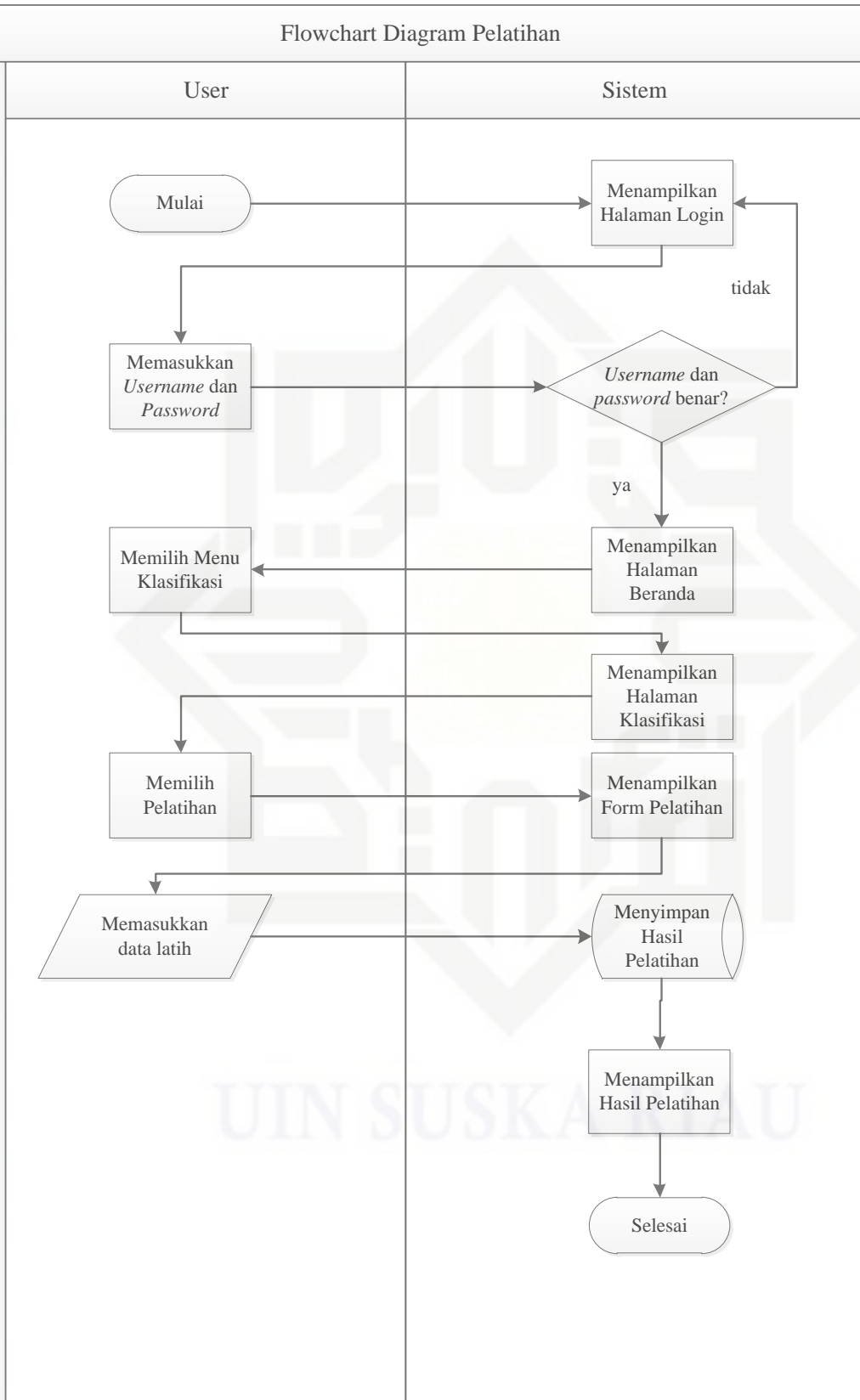
Pada analisa sistem akan menjelaskan mengenai perancangan sistem yang akan dibuat menggunakan *Flow Diagram* (FD) meliputi *Flowchart*, *Context Diagram*, *Data Flow Diagram*(DFD) dan *Entity Relational Diagram* (ERD)

4.4.1 Flowchart

Flowchart dapat menggambarkan alur proses yang terjadi pada sistem, Gambar 4.10 berikut ini menjelaskan proses yang terjadi pada saat pelatihan dan pengujian.

Hak Cipta Dilindungi Undang-Undang

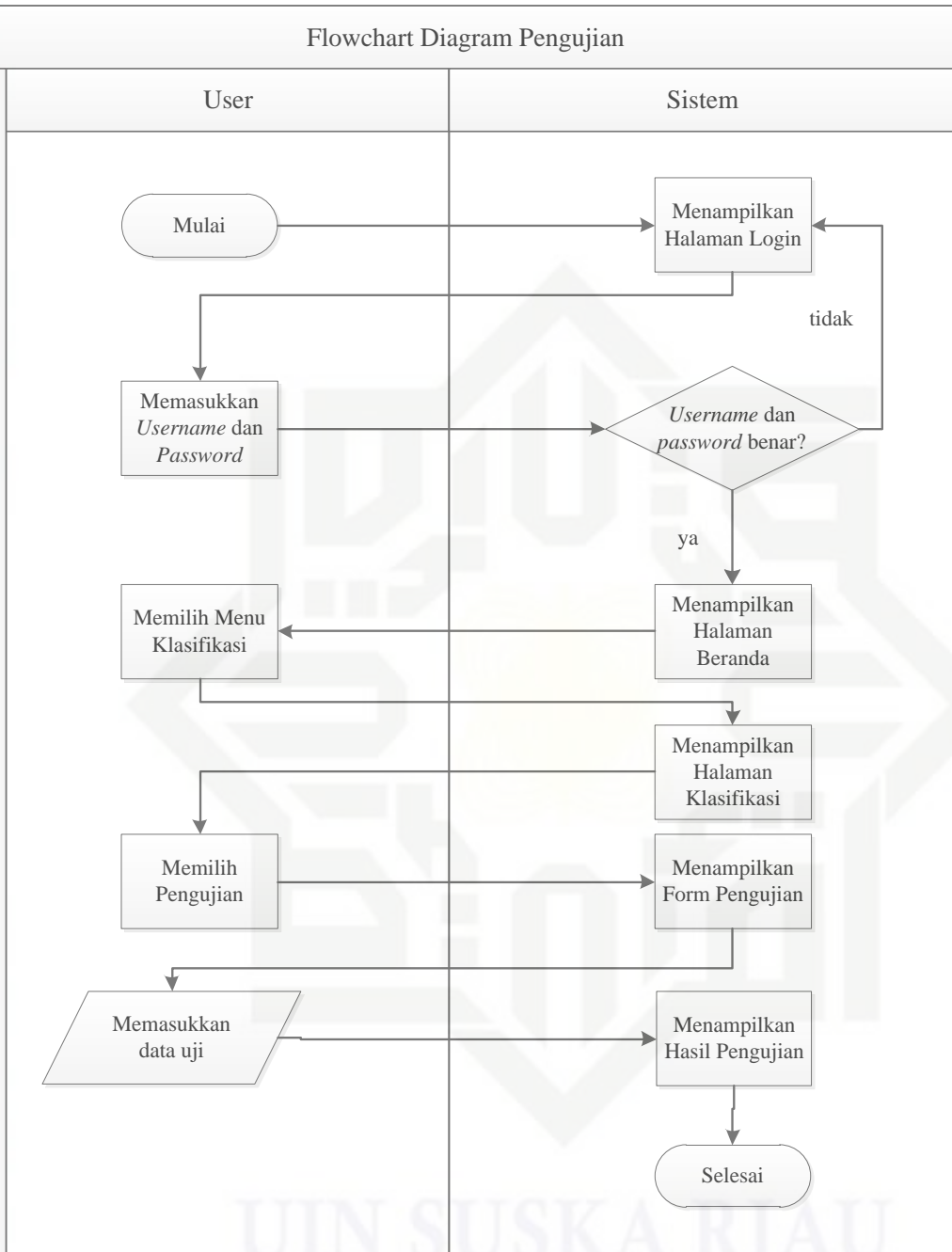
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.10 Flowchart Pelatihan

Hak Cipta Dilindungi Undang-Undang

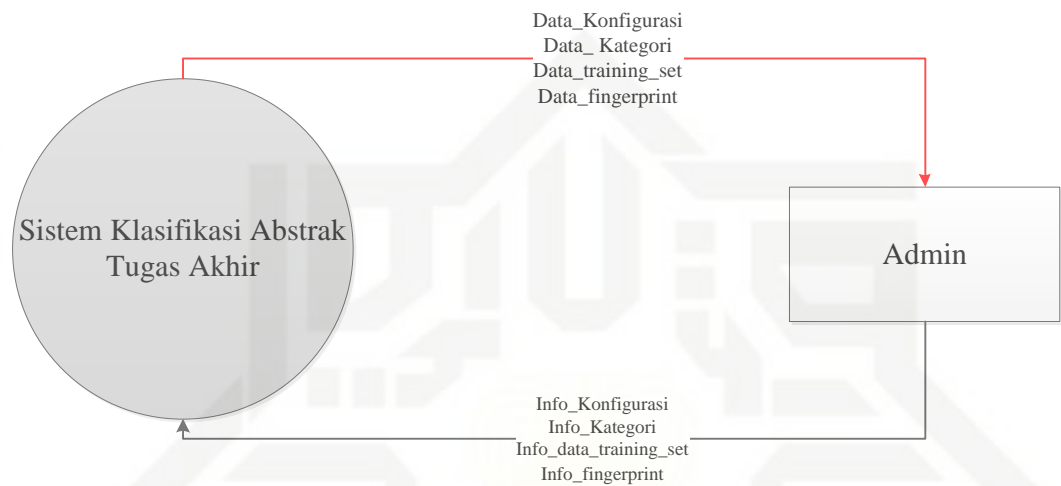
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.11 Flowchart Pengujian

4.4.2 Context Diagram

Context Diagram menggambarkan aliran fungsional suatu proses pada aplikasi. Gambar 4.12 berikut ini dijelaskan semua proses yang terjadi pada sistem.



Gambar 4.12 Context Diagram

Entitas yang berinteraksi pada sistem ini adalah *Administrator* atau *Admin*

Dalam sistem ini, *administrator* dapat melakukan penghapusan, penambahan dan perbaikan data. Aliran data tersebut dapat menghasilkan aliran data baru setelah proses sistem seperti aliran data training set dari hasil proses pelatihan.

4.4.3 Data Flow Diagram (DFD) Level 1

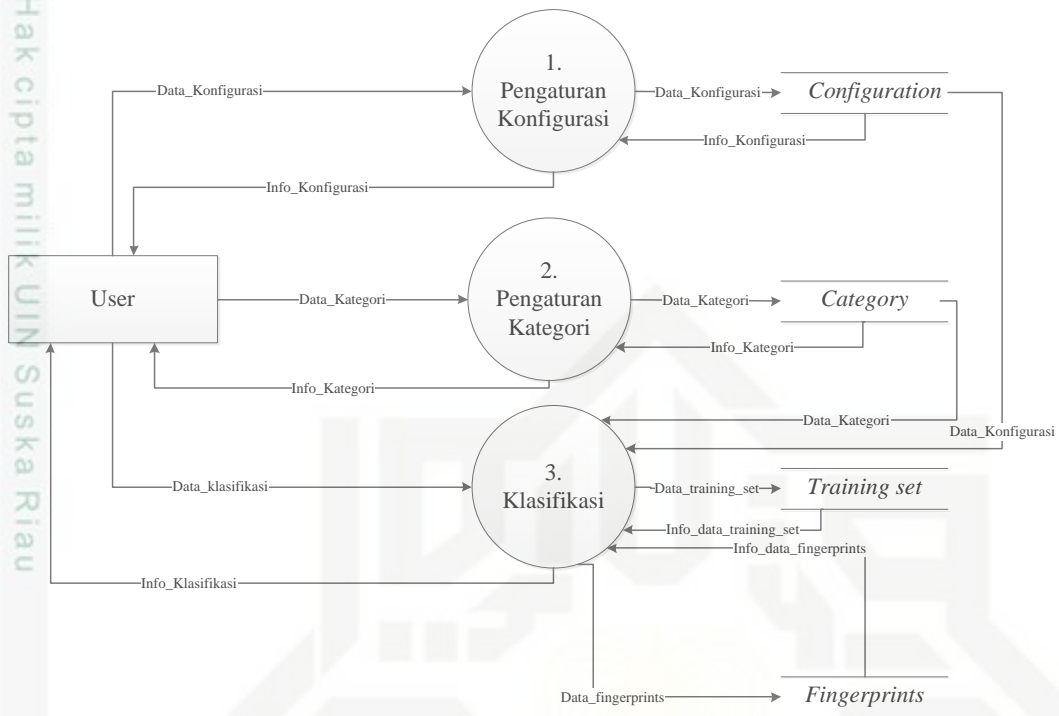
Gambar 4.13 berikut ini adalah gambaran DFD *Level 1* dari sistem klasifikasi abstrak tugas akhir.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.13 DFD Level 1

Pada tabel 4.24 dibawah ini adalah proses DFD *Level 1*. Terdapat 3 proses, yaitu pengaturan konfigurasi, pengaturan kategori, dan Klasifikasi.

Tabel 4.24 Proses DFD Level 1

No	Nama Proses	Deskripsi
1.	Pengaturan konfigurasi	Proses mengisi konfigurasi aturan <i>winnowing</i>
2.	Pengaturan kategori	Proses mengatur kelompok
3.	Klasifikasi	Proses klasifikasi pelatihan dan pengujian sistem.

Pada tabel 4.25 dibawah ini adalah aliran data DFD *Level 1*. Terdapat 6 aliran data yaitu data konfigurasi, info konfigurasi, data kategori, info kategori, data klasifikasi dan info klasifikasi.

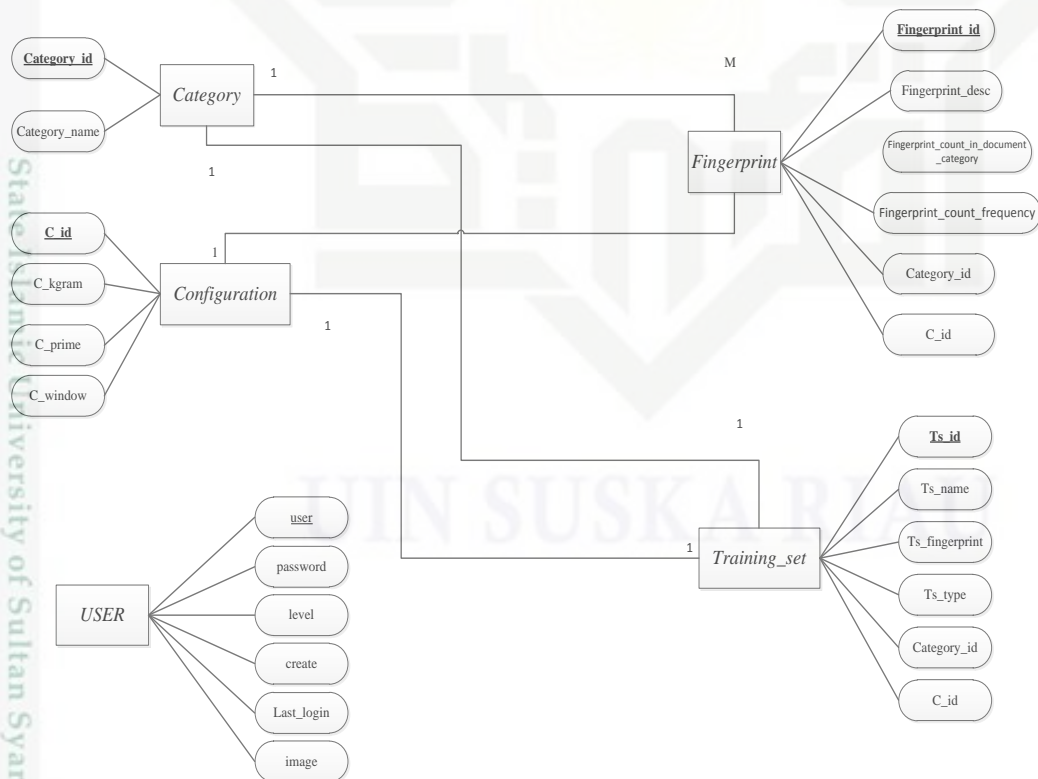
Tabel 4.25 Aliran Data DFD Level 1

No	Nama Proses	Deskripsi
1.	Data Konfigurasi	Data konfigurasi yang dimasukkan oleh <i>user</i>
2.	Info Konfigurasi	<i>User</i> mendapat info konfigurasi
3.	Data Kategori	Data kategori yang dimasukkan oleh <i>user</i>
4.	Info Kategori	<i>User</i> mendapat info kategori
5.	Data Klasifikasi	Data klasifikasi diinputkan oleh <i>user</i>
6.	Info Klasifikasi	<i>User</i> mendapatkan info klasifikasi

Penjelasan DFD selanjutnya dapat dilihat pada **LAMPIRAN A**.

4.4.4 Entity Relationship Diagram (ERD)

Gambar 4.14 adalah ERD pada sistem klasifikasi abstrak tugas akhir. Menggambarkan hubungan antar tabel dari sistem klasifikasi abstrak tugas akhir.



Gambar 4.14 Entity Relational Diagram (ERD)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4.26 berikut ini penjelasan dari ERD pada gambar 4.14.

Tabel 4.26 Keterangan Entity Relational Diagram (ERD)

No	Nama	Deskripsi	Atribut	Primary Key	Foreign Key
1	<i>Fingerprints</i>	Menyimpan data fingerprint	<ul style="list-style-type: none"> - <i>Fingerprint_id</i> - <i>Fingerprint_desc</i> - <i>Fingerprint_count_in_document_category</i> - <i>Fingerprint_count_frequency</i> - <i>Category_id</i> - <i>C_id</i> 	<i>Fingerprint_id</i>	- <i>Category_id</i> <i>C_id</i>
2	<i>Configuration</i>	Menyimpan data konfigurasi	<ul style="list-style-type: none"> - <i>C_id</i> - <i>C_kgram</i> - <i>C_prime</i> - <i>C_window</i> 	<i>C_id</i>	-
3	<i>Category</i>	Menyimpan data kategori	<ul style="list-style-type: none"> - <i>Category_id</i> - <i>Category_name</i> 	<i>Category_id</i>	-
4	<i>Traininset</i>	Menyimpan data latihan	<ul style="list-style-type: none"> - <i>Ts_id</i> - <i>Ts_name</i> - <i>Ts_fingerprint</i> - <i>Ts_type</i> - <i>Category_id</i> - <i>C_id</i> 	<i>Ts_id</i>	- <i>Category_id</i> <i>C_id</i>
5	<i>User</i>	Menyimpan data user	<ul style="list-style-type: none"> - <i>user</i> - <i>password</i> - <i>level</i> - <i>create</i> - <i>last_login</i> - <i>image</i> 	<i>user</i>	-

4.4 Perancangan Sistem

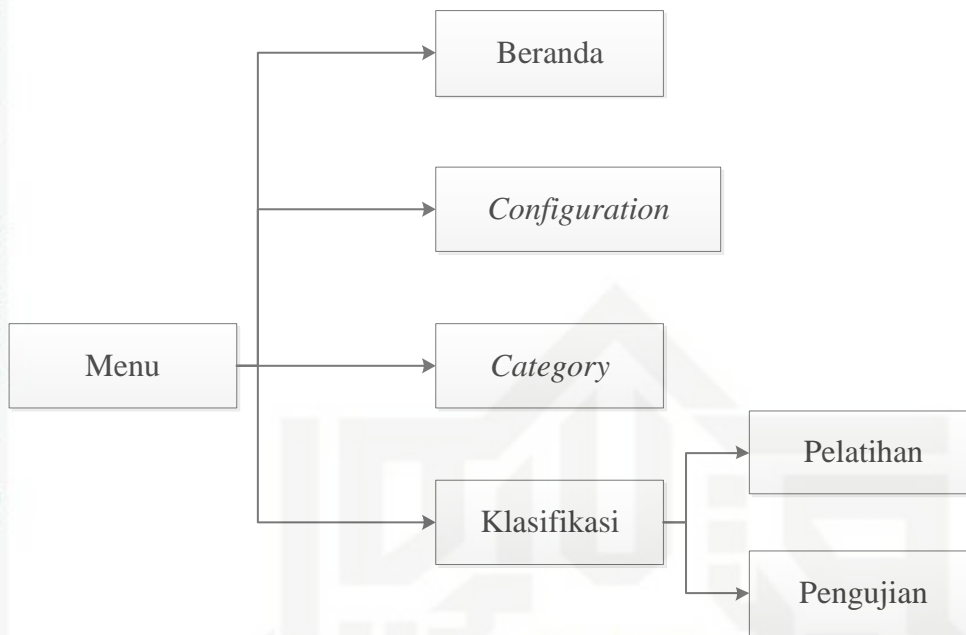
Tahap selanjutnya yang akan dilakukan adalah menentukan rancangan terhadap proses-proses yang telah dijelaskan sebelumnya. Perancangan sistem meliputi perancangan *database*, perancangan struktur menu dan perancangan *interface*.

4.5.1 Perancangan Struktur Menu

Gambar 4.15 berikut ini menjelaskan perancangan struktur menu dari sistem klasifikasi abstrak tugas akhir yang dikembangkan, supaya mempermudah dalam melakukan hubungan antar modul.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.15 Perancangan Struktur Menu

4.5.2 Perancangan *Interface*

Perancangan *interface* atau antarmuka sistem adalah desain awal untuk sebagai acuan pada saat implementasi nantinya. Sistem atau aplikasi ini akan diimplementasikan menggunakan *framework* Yii, sehingga gambar desain berikut ini digunakan sebagai acuan.

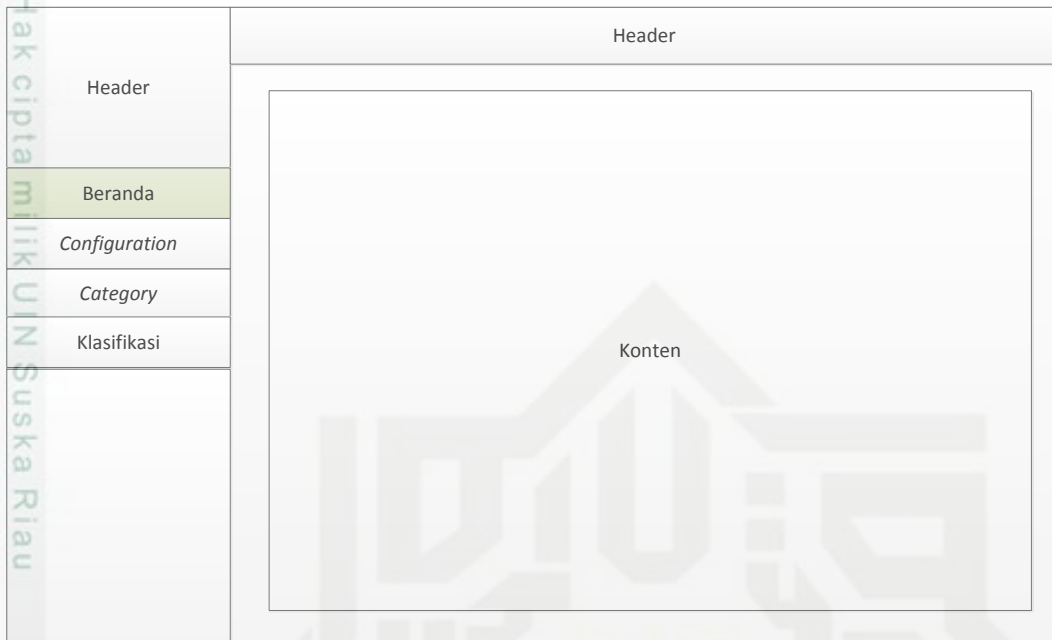
4.5.2.1 Perancangan Menu Beranda

Gambar 4.16 berikut ini merupakan rancangan tampilan awal pada sistem aplikasi klasifikasi dokumen yang akan dikembangkan.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

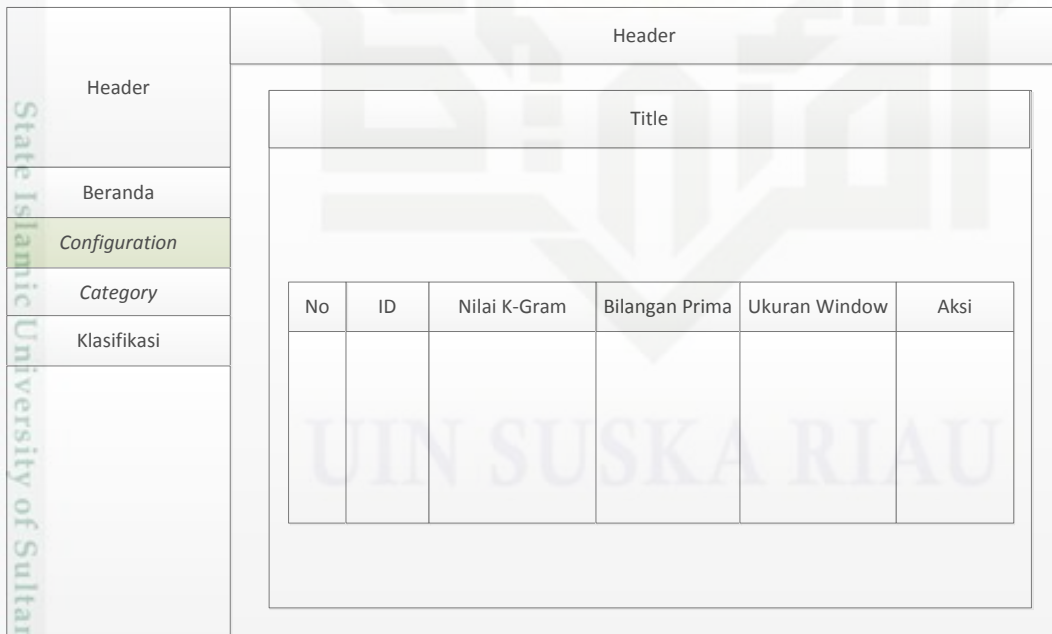
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.16 Perancangan Menu Beranda

4.5.2.2 Perancangan Menu Konfigurasi

Gambar 4.17 berikut ini merupakan perancangan *interface* menu konfigurasi pada sistem aplikasi klasifikasi dokumen yang dikembangkan.



Gambar 4.17 Perancangan Menu Konfigurasi

4.5.2.3 Perancangan *Form* Konfigurasi

Gambar 4.18 berikut ini merupakan perancangan *interface form* konfigurasi pada sistem aplikasi klasifikasi dokumen yang dikembangkan.

Header	Header	
Beranda	Title	
Configuration	<input type="text" value="Nilai K-gram"/>	<input type="text" value="textField"/>
Category	<input type="text" value="Bilangan Prima"/>	<input type="text" value="textField"/>
Klasifikasi	<input type="text" value="Ukuran Window"/>	<input type="text" value="textField"/>
		<input type="button" value="Simpan"/>

Gambar 4.18 Perancangan *Form* Konfigurasi

4.5.2.4 Perancangan Menu Kategori

Gambar 4.19 berikut ini merupakan perancangan *interface* menu kategori pada sistem aplikasi klasifikasi dokumen yang dikembangkan.

Header	Header			
Beranda	Title			
Category	No	Kategori	Nama	Aksi
Klasifikasi				

Gambar 4.19 Perancangan Menu Kategori

4.5.2.5 Perancangan *Form* Kategori

Gambar 4.20 berikut ini merupakan perancangan *interface* menu *form* kategori pada sistem aplikasi klasifikasi dokumen yang dikembangkan.

Header	Header	
Beranda	Title	
Configuration		
Category	Nama Kelompok	textField
Klasifikasi	Simpan	

Gambar 4.20 Perancangan *Form* Kategori

4.5.2.6 Perancangan Menu Klasifikasi

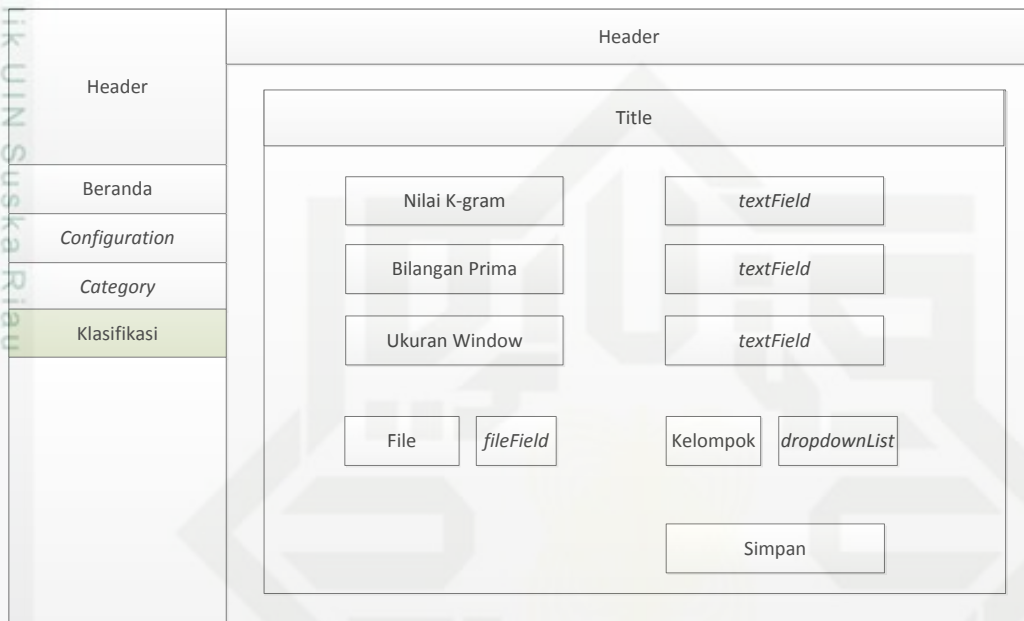
Gambar 4.21 berikut ini merupakan perancangan *interface* menu klasifikasi pada sistem aplikasi klasifikasi dokumen yang dikembangkan.

Header	Header		
Beranda	Title		
Configuration			
Category			
Klasifikasi	Aturan	Pelatihan	Pengujian

Gambar 4.21 Perancangan Menu Klasifikasi

4.5.2.7 Perancangan *Form* Pelatihan

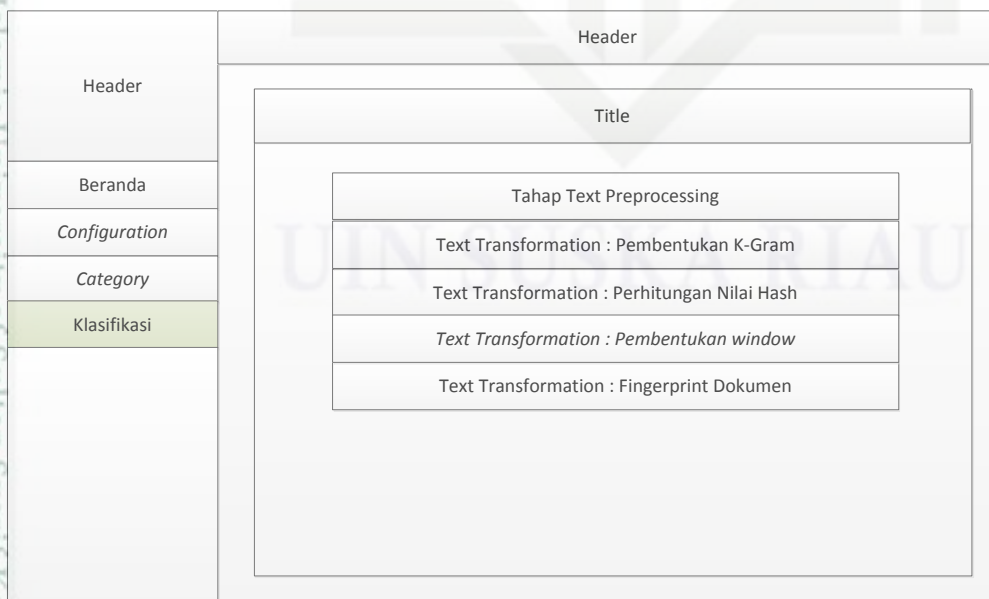
Gambar 4.22 berikut ini merupakan perancangan *interface form* pelatihan pada sistem aplikasi klasifikasi dokumen yang dikembangkan.



Gambar 4.22 Perncangan *Form* Pelatihan

4.5.2.8 Perancangan Halaman Hasil Pelatihan

Gambar 4.23 berikut ini merupakan perancangan *interface* halaman hasil pelatihan pada sistem aplikasi klasifikasi dokumen yang dikembangkan.



Gambar 4.23 Perancangan Halaman Hasil Pelatihan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.5.2.9 Perancangan *Form* Pengujian

Gambar 4.24 berikut ini merupakan perancangan *interface form* pengujian pada sistem aplikasi klasifikasi dokumen yang dikembangkan.

Header	Header	
Beranda	Title	
Configuration	Bilangan Prima	textField
Category	Nilai K-Gram	textField
Klasifikasi	Ukuran Window	textField
	File	fileField
		Simpan

Gambar 4.24 Perancangan *Form* Pengujian

4.5.2.10 Perancangan Halaman Hasil Pengujian

Gambar 4.25 berikut ini merupakan perancangan *interface* halaman hasil pengujian pada sistem aplikasi klasifikasi dokumen yang dikembangkan.

Header	Header	
Beranda	Title	
Configuration	Tahap Text Preprocessing	
Category	Text Transformation : Fingerprint dokumen	
Klasifikasi	Jumlah masing-masing fingerprint di testing set terkait	
	Nilai TF-IDF masing-masing fingerprint	
	Banyaknya jumlah masing-masing fingerprint di testing set terkait	
	Hasil perhitungan Multinomial Logistic Regression	
	Pemilihan kelas klasifikasi berdasar nilai tertinggi	
	Kesimpulan	

Gambar 4.25 Perancangan Halaman Hasil Pengujian

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.5.3 Perancangan Database

Perancangan *database* berikut menjelaskan deskripsi tabel yang telah dirancang pada *database* berdasarkan *class* diagram yang telah dibuat.

Nama : *Configuration*

Deskripsi isi : Berisikan data konfigurasi *winnowing* untuk pro.ses pelatihan dan pengujian.

Primary key : *c_id*

Penjelasan lebih lengkapnya dapat dilihat pada tabel 4.27 berikut.

Tabel 4.27 Conceptual Data Model Tabel Configuration

Nama Field	Type	Null
c_id	varchar (5)	No
c_kgram	int (11)	No
c_prima	int (11)	No
c_window	int (11)	No

Nama : *Category*

Deskripsi isi : Berisi data kelompok untuk proses pelatihan

Primary key : **category_id**

Penjelasan lebih lengkapnya dapat dilihat pada Tabel 4.28 berikut.

Tabel 4.28 Conceptual Data Model Tabel Category

Nama Field	Type	Null
category_id	int (11)	No
category_name	varchar (100)	No

Nama : *training_set*

Deskripsi isi : Berisi dokumen latih yang telah melalui proses pelatihan

Primary key : **ts_id**

Foreign key : **category_id, c_id**

Penjelasan lebih lengkapnya dapat dilihat pada Tabel 4.29 berikut.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4.29 Conceptual Data Model Tabel *training set*

Nama Field	Type	Null
ts_id	int (11)	No
ts_name	varchar (50)	No
ts_fingerprint	text	No
ts_type	int (11)	No
ts_kgram	date	No
ts_prima	int (11)	No
ts_window	int (11)	No
category_id	int (11)	No
c_id	int(11)	No

Nama : *fingerprints*

Deskripsi isi : Berisi deskripsi fingerprint dan perjumlahannya yang menjadi acuan untuk proses pengujian

Primary key : *fingerprint_id*

Foreign key : *kategori_id, c_id*

Penjelasan tabel *fingerprints* dapat dilihat selengkapnya pada Tabel 4.30 berikut.

Tabel 4.30 Conceptual Data Model Tabel *fingerprints*

Nama Field	Type	Null
fingerprint_id	int (11)	No
fingerprint_desc	varchar (50)	No
fingerprint_count_in_document_category	int (11)	No
fingerprint_count_frequency	int (11)	No
fingerprint_kgram	int (11)	No
fingerprint_prima	int (11)	No
fingerprint_window	int (11)	No
category_id	int(11)	No
c_id	int(11)	No

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Nama : user

Deskripsi isi : Berisi data pengguna

Primary key : **user**

Penjelasan tabel user dapat dilihat selengkapnya pada Tabel 4.31 berikut.

Tabel 4.31 Conceptual Data Model Tabel User

Nama Field	Type	Null
user	varchar (50)	No
password	varchar (50)	No
level	varchar (10)	No
create	timestamp	No
last_login	timestamp	No
image	varchar (255)	No