



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.

## BAB IV ANALISA DAN PERANCANGAN

Pada bab ini berisikan analisa proses pendeteksian plagiarisme proposal tugas akhir menggunakan algoritma *Latent Semantic Analysis* (LSA) terhadap masukan dari *file* proposal yang akan diperiksa dan melakukan perancangan sistem serta perancangan antarmuka dari sistem.

### 4.1 Analisa Sistem

Analisa sistem dilakukan untuk membantu dan memudahkan dalam melakukan proses pencarian dari informasi yang berguna untuk kepentingan sistem yang akan dibangun. Pada penelitian ini informasi yang dibutuhkan adalah *soft file* dari proposal Jurusan Teknik Informatika yang akan dijadikan *corpus* sebagai data latih untuk melakukan pengujian pendeteksian plagiarisme.

#### 4.1.1 Analisa Masukan

Analisa masukan dilakukan untuk mengetahui data-data yang akan digunakan dalam pembangunan sistem. Data masukan yang akan digunakan dalam penelitian ini adalah sebagai berikut:

1. Proposal yang berisi bab I
2. Kumpulan daftar *stopwords*
3. Kumpulan matriks
4. Kumpulan kamus bahasa Indonesia


**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.

#### 4.1.2 Analisa Proses

Terdapat subproses yang akan dilaksanakan dalam penelitian ini sebagai berikut:

1. Subproses *preprocessing*.
  - a. Mengumpulkan proposal tugas akhir mahasiswa dalam bentuk *soft file* berformat \*.doc/docx yang akan diuji.
  - b. Melakukan pembuangan *stopword*.
  - c. Melakukan proses mengubah kata yang berimbuhan menjadi kata dasar (*Stemming*).
  - d. Melakukan proses pembobotan kata menggunakan TF-IDF.
2. Subproses pemodelan.
  - a. Menerapkan Algoritma *Latent Semantic Analysis* (LSA).
  - b. Melakukan perhitungan kemiripan (*similarity*) atau relevansi antara proposal tugas akhir dengan proposal yang akan diuji.
3. Subproses antarmuka.

Proses antarmuka sistem berisikan tampilan, fungsi-fungsi dari sistem seperti menu-menu yang berada disistem, masukan dari sistem, dll.

Dari tahapan proses analisa sistem, data *corpus* yang digunakan dalam bentuk *soft file* berformat \*.doc/docx yang akan dikonversi ke *file* berformat text (\*.txt). Sistem dapat melakukan pekerjaan seperti melakukan tahap *preprocessing* terhadap dokumen proposal tugas akhir dan menerapkan model *Latent Semantic Analysis* (LSA) sebagai perhitungan kemiripan dokumen yang akan diproses. Penggunaan model LSA juga digunakan dalam menentukan hasil akhir dari proses pendeteksian plagiarisme proposal tugas akhir yang akan diperiksa.

Pada tahapan subproses *preprocessing* yang dilakukan adalah melakukan penghapusan tanda baca. Akan melakukan proses pembuangan kata yang dianggap tidak penting (*stopword*) berdasarkan kamus kata yang dianggap tidak penting. Sistem akan melakukan proses pemisahan rangkaian kata ini meliputi perubahan



- 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.

kata berimbuhan menjadi kata dasar (*stemming*). Pemisahan kata ini seperti kata bermain menjadi kata dasar. Kemudian melakukan proses pembobotan kata menggunakan perhitungan TF-IDF sebelum melakukan proses algoritma *Latent Semantic Analysis* (LSA).

Selanjutnya tahapan subproses pemodelan yang berisi penerapan mengenai model yang akan digunakan pada sistem, yaitu *Latent Semantic Analysis* (LSA) yang akan melakukan perhitungan matrik pada dokumen proposal tugas akhir. Matrik tersebut juga akan menghitung nilai kemiripan pada dokumen proposal tugas akhir yang ada pada data latih dengan dokumen proposal tugas akhir yang akan diuji nilai kemiripannya. Penggunaan model *Latent Semantic Analysis* (LSA) diharapkan mampu menghasilkan performa sistem yang baik dan relevan.

Tahapan selanjutnya adalah subproses antarmuka yang berisi tampilan dari sistem yang akan dibangun, dimana terdapat menu dari sistem, submenu, media dan masukan. Pada tahapan ini juga berisi mengenai perancangan antarmuka (*interface*) dari sistem yang akan dibangun agar sistem nantinya akan mudah digunakan. Antarmuka sistem nantinya akan dirancang menggunakan bahasa pemrograman PHP.

#### 4.1.3 Analisa Keluaran

Analisa keluaran merupakan hasil keluaran dari sistem pendeteksian plagiarisme proposal tugas akhir menggunakan algoritma *Latent Semantic Analysis* (LSA). Selanjutnya sistem akan melakukan perhitungan menggunakan algoritma *Latent Semantic Analysis* (LSA) dan memberikan hasil berupa nilai kemiripan dokumen proposal yang diuji dengan yang ada pada data latih.

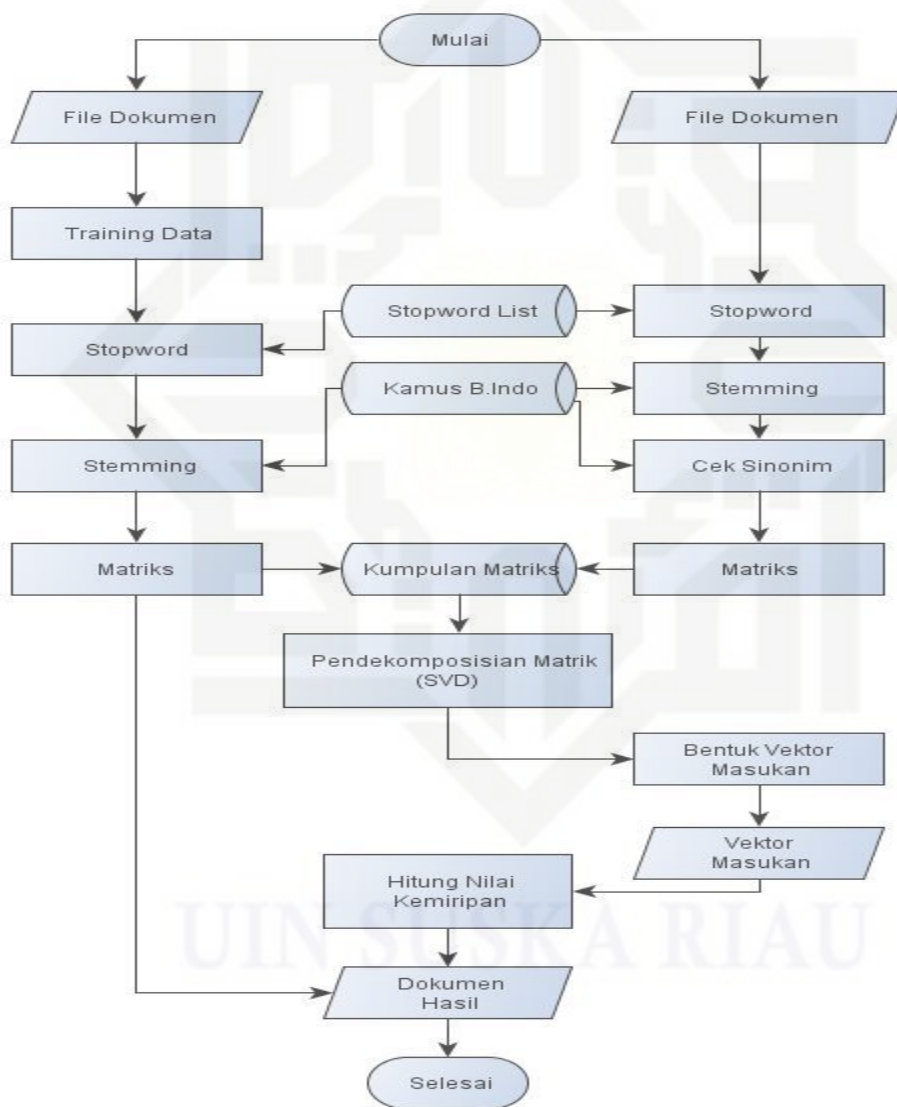
## 4.2 Analisa Tahapan Sistem

Pada tahapan ini terdapat beberapa proses yang akan dilakukan oleh sistem, yaitu melakukan tahap *preprocessing* dan tahap penerapan algoritma *Latent Semantic Analysis* (LSA) untuk melakukan perhitungan nilai kemiripan dokumen. Terdapat beberapa analisa yang dilakukan dalam tahapan sistem, yaitu merancang model

menggunakan *Unified Modeling Language* (UML). Perancangan *interface* sistem dan struktur yang mudah untuk digunakan oleh pengguna (*user friendly*) serta merancang tampilan yang menarik sehingga pengguna nyaman menggunakan sistem tersebut.

#### 4.2.1 Flowchart Diagram

Di bawah ini merupakan *Flowchart Diagram* yang merupakan gambaran umum alur dari sistem yang akan dibangun.



Gambar 4.1 *Flowchart Diagram*

Dari diagram di atas dapat dilihat dua proses penginputan *corpus* dan deteksi. Proses yang pertama melakukan penginputan *corpus* dengan melakukan upload *file* dokumen proposal tugas akhir yang berekstensi \*.doc/docx. Setelah selesai melakukan upload *file* dokumen proposal, pengguna harus memilih tab menu *training* data agar *corpus* yang di-upload sukses disimpan. Dalam proses *training* data, sistem akan melakukan proses *stopword* dan *stemming* serta melakukan pembuatan matriks.

Proses yang kedua adalah dalam melakukan deteksi. Pengguna melakukan *input file* dokumen proposal yang dicurigai plagiat. Kemudian sistem akan melakukan tahapan *stopword* dengan membuang kata yang dianggap tidak penting. Proses *stopword* ini dilakukan sesuai dengan *stopword list* yang ada. Tahap selanjutnya adalah melakukan proses *stemming*, yaitu melakukan penghapusan kata yang memiliki imbuhan. Setelah mendapatkan bobot dari kata, maka proses selanjutnya akan dihitung menggunakan algoritma *Latent Semantic Analysis* (LSA) untuk melakukan perhitungan matrik demi mendapatkan nilai kemiripan dari dokumen proposal tugas akhir yang akan diperiksa. Hasil dari proses perhitungan menggunakan algoritma LSA akan menampilkan nilai kemiripan tertinggi dari dokumen proposal tugas akhir pada data latih dengan dokumen proposal tugas akhir yang diuji.

Berikut merupakan contoh kasus sederhana yang dilakukan:

- Q : Era globalisasi ditandai dengan berkembangnya ilmu pengetahuan dan teknologi yang sangat pesat.  
D1 : Pada saat ini berkembangnya teknologi sangatlah pesat.  
D2 : Era globalisasi dapat dilihat dengan berkembangnya teknologi yang sangat cepat.  
D3 : Berkembangnya teknologi pada saat ini berawal dari ilmu pengetahuan yang sangat maju.  
D4 : Berkembangnya ilmu pengetahuan dan teknologi saat ini dapat membantu pekerjaan dengan baik.



#### 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.

D5 : Dewasa ini berkembangnya teknologi dapat mempermudah dan membantu pekerjaan.

Keterangan:

Q = *Query*

D = Dokumen

Diketahui:

Q = kueri yang diinputkan

Dn = 5 (total dokumen)

Proses yang dilakukan adalah proses *preprocessing* sebagai berikut:

1. Memisahkan rangkaian kata pada kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal. Pada tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan merubah semua token menjadi huruf kecil. Untuk lebih jelas dapat dilihat di bawah ini.

D1 : pada saat ini berkembangnya teknologi sangatlah pesat .

D2 : era globalisasi dapat dilihat dengan berkembangnya teknologi yang sangat cepat

D3 : berkembangnya teknologi pada saat ini berawal dari ilmu pengetahuan yang sangat maju

D4 : berkembangnya ilmu pengetahuan dan teknologi saat ini dapat membantu pekerjaan dengan baik

D5 : dewasa ini berkembangnya teknologi dapat mempermudah dan membantu pekerjaan

2. Melakukan proses *stopword*. Hasilnya sebagai berikut:

D1 : berkembangnya teknologi pesat .

D2 : era globalisasi dilihat berkembangnya teknologi cepat

D3 : berkembangnya teknologi berawal ilmu pengetahuan maju

D4 : berkembangnya ilmu pengetahuan teknologi membantu pekerjaan baik

D5 : dewasa berkembangnya teknologi mempermudah membantu pekerjaan

3. Merubah kata yang ada menjadi kata dasar (*stemming*). Hasilnya sebagai berikut:

- D1 : berkembang teknologi pesat .
- D2 : era global lihat berkembang teknologi cepat
- D3 : berkembang teknologi awal ilmu pengetahuan maju
- D4 : berkembang ilmu pengetahuan teknologi bantu pekerjaan baik
- D5 : dewasa berkembang teknologi mudah bantu pekerjaan

4. Memberikan bobot pada kata dan membangun *index*.

Pada tahapan ini, kata yang ada di dalam dokumen akan diberi nilai bobot sebelum diindeks berdasarkan jumlah kemunculan pada tiap-tiap dokumen agar dapat dianalisa kemiripannya dengan dokumen yang ada. Pemberian nilai bobot menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*). Nilai bobot yang diberikan pada masing-masing kata dapat didefinisikan melalui perhitungan frekuensi kemunculan dokumen yang mengandung sebuah kata, yaitu DF (*document frequency*). Perhitungan frekuensi kemunculan kata di dalam dokumen yaitu TF (*term frequency*) dan perhitungan jumlah dokumen yang mengandung sebuah kata yang dicari dari kumpulan dokumen yang ada yaitu IDF (*inverse document frequency*).

Berikut merupakan perhitungan tahap perhitungan pada kata di tiap-tiap dokumen untuk memperoleh nilai TF dapat dilihat pada tabel 4.1 di bawah ini.

**Tabel 4.1 Nilai TF (*term frequency*)**

Token	tf					
	Q	D1	D2	D3	D4	D5
era	1	0	1	0	0	0
bantu	0	0	0	0	1	1
tanda	1	0	0	0	0	0
dewasa	0	0	0	0	0	1
kerja	0	0	0	0	1	1
baik	0	0	0	0	1	0
teknologi	1	1	1	1	1	1
maju	0	0	0	1	0	0
awal	0	0	0	1	0	0
ilmu	1	0	0	1	1	0
berkembang	1	1	1	1	1	1
mudah	0	0	0	0	0	1

Token	tf					
	Q	D1	D2	D3	D4	D5
pengetahuan	1	0	0	1	1	0
pesat	1	1	0	0	0	0
cepat	0	0	1	0	0	0
global	1	0	1	0	0	0
lihat	0	0	1	0	0	0

Dari tabel di atas dapat dilihat pemberian nilai bobot pada masing-masing kata yang muncul pada setiap dokumen yang telah diproses melalui tahapan *preprocessing*. Nilai TF merupakan nilai yang didapat dari kata yang muncul pada dokumen yang telah dipecah menjadi kata dasar dan dihitung frekuensi kemunculannya pada setiap dokumen.

Proses selanjutnya yaitu mencari jumlah dokumen yang mengandung sebuah kata yang akan dicari melalui kumpulan dari dokumen yang ada dengan melakukan perkalian antara nilai TF dan DF. Untuk lebih jelasnya dapat dilihat pada tabel 4.2 di bawah ini.

**Tabel 4.2 Nilai TFxDF (term frequency x document frequency)**

idf log(n/df)	tf x idf					
	q	d1	d2	d3	d4	d5
0.39794	0.39794	0	0.39794	0	0	0
0.39794	0	0	0	0	0.39794	0.39794
0.69897	0.69897	0	0	0	0	0
0.69897	0	0	0	0	0	0.69897
0.39794	0	0	0	0	0.39794	0.39794
0.69897	0	0	0	0	0.69897	0
-0.07918	-0.07918	-0.07918	-0.07918	-0.07918	-0.07918	-0.07918
0.69897	0	0	0	0.69897	0	0
0.69897	0	0	0	0.69897	0	0
0.221849	0.221849	0	0	0.221849	0.221849	0
-0.07918	-0.07918	-0.07918	-0.07918	-0.07918	-0.07918	-0.07918
0.69897	0	0	0	0	0	0.69897
0.221849	0.221849	0	0	0.221849	0.221849	0
0.39794	0.39794	0.39794	0	0	0	0
0.69897	0	0	0.69897	0	0	0
0.39794	0.39794	0	0.39794	0	0	0
0.69897	0	0	0.69897	0	0	0





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.

Dari tabel di atas dapat dilihat hasil dari perhitungan nilai TF-IDF yang akan dijadikan sebuah matrik asal. Nilai matrik yang diambil adalah seluruh bobot dokumen tidak termasuk bobot kueri dan dapat disebut dengan matrik A. Untuk lebih jelasnya dapat dilihat pada matrik di bawah ini.

0	0,39794	0	0	0
0	0	0	0,39794	0,39794
0	0	0	0	0
0	0	0	0	0,69897
0	0	0	0,39794	0,39794
0	0	0	0,69897	0
-0,07918	-0,07918	-0,07918	-0,07918	-0,07918
0	0	0,69897	0	0
0	0	0,69897	0	0
0	0	0,221849	0,221849	0
-0,07918	-0,07918	-0,07918	-0,07918	-0,07918
0	0	0	0	0,69897
0	0	0,221849	0,221849	0
0,39794	0	0	0	0
0	0,69897	0	0	0
0	0,39794	0	0	0
0	0,69897	0	0	0

Matrik A di atas akan didekomposisikan menjadi matriks U, S, V *transpose*. Pendekomposisian matrik A dilakukan sesuai dengan proses yang ada dalam algoritma *Latent Semantic Analysis* (LSA) menggunakan teknik aljabar linear yaitu *Singular Value Decomposition* (SVD). Hasil dari pendekomposisian matrik A dapat dilihat pada matrik dibawah ini.

1. Matriks U

0,11829	0,10009	0,14015	0,26927	0,43467
0,15664	0,09179	0,50926	0,48478	0,56035
0,00000	0,00000	0,00000	0,00000	0,00000
0,10135	0,03295	0,29561	0,61636	0,56650
0,15664	0,09179	0,50926	0,48478	0,56035
0,17379	0,12827	0,59888	0,23514	0,41774
-0,26556	-0,10916	-0,14930	-0,17311	-0,23520



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.

0,16816	0,56774	0,11025	0,13192	0,25679
0,16816	0,56774	0,11025	0,13192	0,25679
0,10853	0,22091	0,22507	0,11650	0,21409
-0,26556	-0,10916	-0,14930	-0,17311	-0,23520
0,10135	0,03295	0,29561	0,61636	0,56650
0,10853	0,22091	0,22507	0,11650	0,21409
0,96397	0,03352	0,03818	0,04082	0,04081
0,20777	0,17581	0,24617	0,47296	0,76348
0,11829	0,10009	0,14015	0,26927	0,43467
0,20777	0,17581	0,24617	0,47296	0,76348

Nilai matriks U di atas diperoleh dari perkalian antara matriks A dengan matriks A transpose. Kemudian dicari nilai *eigenvalues* dan *eigenvectors* dari matriks  $AA^T$ . Untuk mendapatkan nilai *eigenvectors* digunakan rumus  $A\vec{v} = \lambda\vec{v}$  pada matriks  $AA^T$ . Kemudian dicari nilai *eigenvalues* dengan menghitung determinan sama dengan nol ( $D=0$ ).

2. Matriks S dan S inverse

0,41277	0	0	0	0
0	1,22751	0	0	0
0	0	1,14246	0	0
0	0	0	1,04817	0
0	0	0	0	0,84080

2,42267	0	0	0	0
0	0,81466	0	0	0
0	0	0,87531	0	0
0	0	0	0,95404	0
0	0	0	0	1,18934

Nilai matriks S diperoleh dari nilai *eigenvalue* yang diakarkan. Kemudian nilai tersebut akan diubah menjadi matriks *orthogonal* dengan menggunakan proses Gram-Schmidt *orthonormalization* pada vektor kolom.



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. Matriks  $V$  transpose

0,99989	0,10341	0,10962	0,10751	0,08623
0,12270	0,30875	0,40236	0,70925	0,91840
0,09930	0,99704	0,18020	0,19783	0,30889
0,10263	0,22527	0,97886	0,35262	0,50251
0,05985	0,05786	0,48318	0,92428	0,68145

Nilai dari matriks  $V$  transpose didapat dari proses pencarian nilai *eigenvalues* dan *eigenvectors*. Prosesnya dilakukan dengan perkalian matriks  $A$  transpose dengan matriks  $A$ . Kemudian barulah dicari nilai *eigenvalues* dan *eigenvectors* dari  $A^T A$ . Proses selanjutnya sama dengan mencari nilai matriks  $U$ .

4. Kueri transpose

Untuk mencari nilai *similarity* (kemiripan) memerlukan matriks kueri dan kueri transpose. Untuk lebih jelasnya dapat dilihat pada matrik di bawah ini.

0,39794	0	0,69897	0	0	0	-0,0792	0	0	0,22185	-0,0792	0	0,22185	0,3979	0	0,39794	0
---------	---	---------	---	---	---	---------	---	---	---------	---------	---	---------	--------	---	---------	---

Setelah mendapatkan kueri transpose, kemudian mencari panjang vektor kueri dan panjang vektor dokumen. Untuk mencari panjang vektor kueri dan panjang vektor dokumen digunakan rumus:

$$\bar{q} = q^T U_k \sum_k 1$$

Keterangan:

- $\bar{q}$  : Vektor kueri
- $q^T$  : Transpose matrik vektor kueri
- $U_k$  : Matrik Singular kiri dalam ruang  $k$
- $\sum_k 1$  : Invers Matrik Singular dalam ruang  $k$

Tabel 4.3 Vektor Kueri

Vektor Kueri	
1,375972	0,169698



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 4.4 Vektor Dokumen**

Vektor Dokumen		
d1	0,999888	0,103412
d2	0,122696	0,308748
d3	0,099304	0,99704
d4	0,102629	0,225266
d5	0,05985	0,057865

Langkah selanjutnya adalah mencari nilai *similarity* (kemiripan) menggunakan *cosine similarity* dengan menggunakan rumus:

$$\text{Cos}(q, d) = \frac{\bar{q} \cdot \bar{d}}{\|\bar{q}\| \|\bar{d}\|} = \frac{\sum_{i=1}^n \bar{q}_i \times \bar{d}_i}{\sqrt{\sum_{i=1}^n (\bar{q}_i)^2} \times \sqrt{\sum_{i=1}^n (\bar{d}_i)^2}}$$

Keterangan:

$\bar{q}$  : Vektor kueri

$\bar{d}$  : Vektor dokumen

$\bar{q} \cdot \bar{d}$  : Dot produk antara vektor kueri dan dokumen

$\|\bar{q}\|$  : Panjang vektor kueri

$\|\bar{d}\|$  : Panjang vektor dokumen

$\|\bar{q}\| \|\bar{d}\|$  : Cross produk antara  $\|\bar{q}\|$  dan  $\|\bar{d}\|$

**Tabel 4.5 Hasil Cosine Similarity**

Cosine Similarity	
cos(Q,D1)	0,540465
cos(Q,D2)	0,886867
cos(Q,D3)	0,975862
cos(Q,D4)	0,866395
cos(Q,D5)	0,697707

Dari tabel 4.5 dapat dilihat nilai kedekatan antara kueri dengan dokumen yang ada pada database. Dapat dilihat nilai persentase dari seluruh dokumen di atas 50%, hal ini membuktikan bahwa kueri yang diinput adalah plagiat terhadap 5 dokumen yang ada dan nilai persentase tertinggi terdapat pada dokumen 3.

## 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.

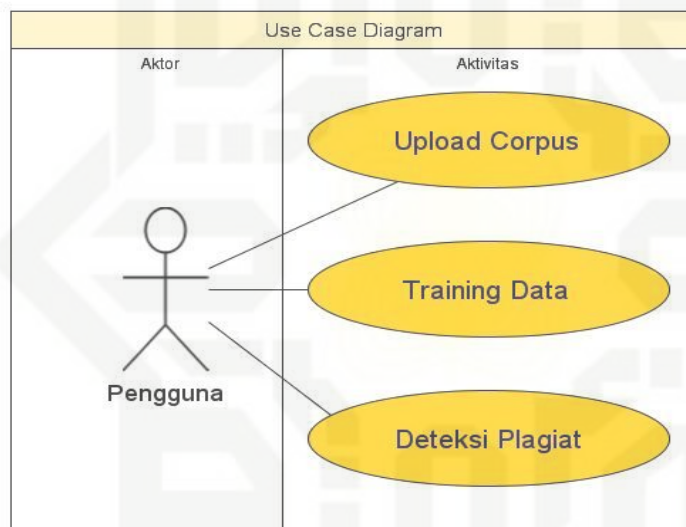
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 4.3 Perancangan Sistem

Tahap perancangan sistem dilakukan dengan merancang *Unified Modeling Language* (UML) yang meliputi *use case diagram*, *sequence diagram* dan *class diagram*.

#### 4.3.1 Use Case Diagram

Pada *use case diagram* digambarkan aktor dan aktivitas yang dapat dilakukannya di dalam sistem. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.



Gambar 4.2 Use Case Diagram

Pada gambar 4.2 dapat dilihat terdapat satu aktor yang ada pada sistem, yaitu pengguna. Pengguna dapat melakukan proses *upload corpus*, *training data* dan deteksi plagiat. Proses *upload corpus* dilakukan dengan memilih *file* dokumen yang akan di-*upload* dengan ekstensi *file* \*.doc/docx. Setelah melakukan *upload corpus*, pengguna harus melakukan *training data* yang berfungsi untuk membuat matriks dari *corpus* yang telah di-*upload*. Proses terakhir, yaitu deteksi plagiat dengan memilih *file* yang akan diuji kemiripannya.



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 4.6 Use Case Specification Upload Corpus**

Nama Use Case : <i>Upload Corpus</i>		
Aktor	:	Pengguna
Kondisi Awal	:	Pengguna masuk ke dalam sistem Pendeteksian Plagiarisme Proposal Tugas Akhir
Kondisi Akhir	:	<i>Corpus</i> berhasil di- <i>upload</i>
Main Scenario Success	:	<ol style="list-style-type: none"> <li>1. <i>Use case</i> dimulai ketika pengguna memilih menu <i>upload corpus</i></li> <li>2. Sistem menampilkan <i>interface</i> untuk <i>Upload Corpus</i></li> <li>3. Pengguna memasukkan <i>file</i> dokumen untuk dijadikan <i>corpus</i></li> <li>4. Sistem menyimpan <i>file</i> dokumen ke dalam <i>corpus</i> dan menampilkan data <i>corpus</i> berhasil di-<i>upload</i></li> </ol>
Exeption	:	<ol style="list-style-type: none"> <li>1. Pengguna tidak meng-<i>input file</i> dokumen atau tidak sesuai ekstensi <i>file</i> yang dianjurkan, yaitu *.doc/docx</li> <li>2. Sistem mengeluarkan pesan bahwa ekstensi <i>file</i> salah</li> <li>3. Pengguna memperbaiki <i>file</i> dokumen masukan sesuai ekstensi</li> <li>4. Sistem menyimpan <i>file</i> dokumen ke dalam <i>corpus</i> dan menampilkan data <i>corpus</i> berhasil di-<i>upload</i></li> </ol>

**Tabel 4.7 Use Case Specification Training Data**

Nama Use Case : <i>Training Data</i>		
Aktor	:	Pengguna
Kondisi Awal	:	Pengguna masuk ke dalam sistem Pendeteksian Plagiarisme Proposal Tugas Akhir
Kondisi Akhir	:	<i>Training data</i> selesai
Main Scenario Success	:	<ol style="list-style-type: none"> <li>1. <i>Use case</i> dimulai ketika pengguna memilih menu <i>training data</i></li> <li>2. Sistem menampilkan <i>interface</i> untuk <i>training data</i></li> <li>3. Pengguna memilih tombol <i>training data</i></li> <li>4. Sistem melakukan <i>training data</i> dan menampilkan <i>training data</i> selesai</li> </ol>

**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 4.8 Use Case Specification Deteksi Plagiat**

Nama <i>Use Case</i> : Deteksi Plagiat		
Aktor	:	Pengguna
Kondisi Awal	:	Pengguna masuk ke dalam sistem Pendeteksian Plagiarisme Proposal Tugas Akhir
Kondisi Akhir	:	Sistem menampilkan nilai kemiripan tiap dokumen <i>corpus</i> terhadap dokumen uji
<i>Main Scenario Success</i>	:	<ol style="list-style-type: none"> <li>1. <i>Use case</i> dimulai ketika pengguna memilih menu deteksi plagiat</li> <li>2. Sistem menampilkan <i>interface</i> untuk deteksi plagiat</li> <li>3. Pengguna memasukkan <i>file</i> dokumen untuk dilakukan deteksi plagiat</li> <li>4. Sistem melakukan deteksi plagiat terhadap <i>file</i> dokumen yang di-<i>upload</i> dan menampilkan nilai kemiripan tiap dokumen <i>corpus</i> terhadap dokumen uji</li> </ol>
<i>Exeption</i>	:	<ol style="list-style-type: none"> <li>1. Pengguna tidak meng-<i>input file</i> dokumen uji atau tidak sesuai ekstensi <i>file</i> yang dianjurkan, yaitu *.doc/docx</li> <li>2. Sistem mengeluarkan pesan bahwa ekstensi <i>file</i> salah</li> <li>3. Pengguna memperbaiki <i>file</i> dokumen uji masukan sesuai ekstensi</li> <li>4. Sistem melakukan deteksi plagiat terhadap <i>file</i> dokumen yang di-<i>upload</i> dan menampilkan nilai kemiripan tiap dokumen <i>corpus</i> terhadap dokumen uji</li> </ol>

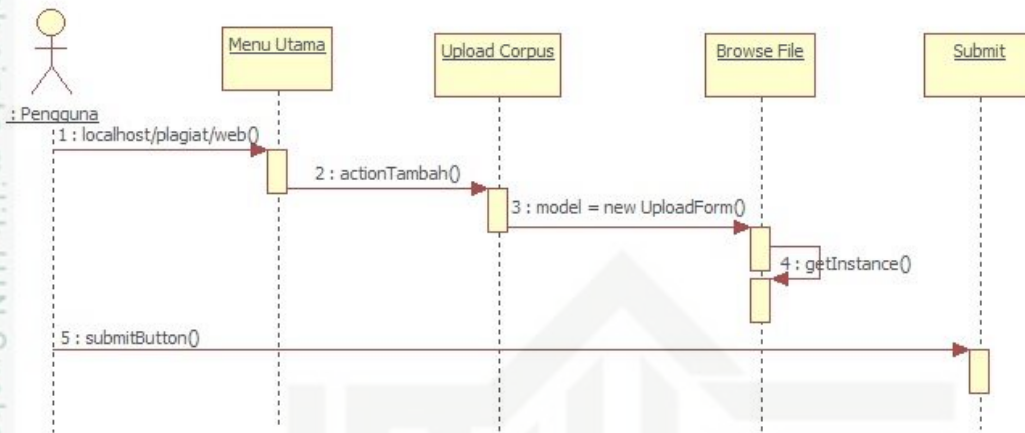
**4.3.2 Sequence Diagram**

*Sequence Diagram* digunakan untuk menggambarkan kelakuan/perilaku objek pada use case dengan mendeskripsikan waktu hidup objek dan pesan yang dikirimkan serta pesan yang diterima antar objek. Terdapat 5 *sequence diagram* yang sesuai dengan *use case specification*, yaitu *sequence main success* dan *exception* pada *upload corpus*, *main success training data* dan *main success* dan *exception* pada deteksi plagiat. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.

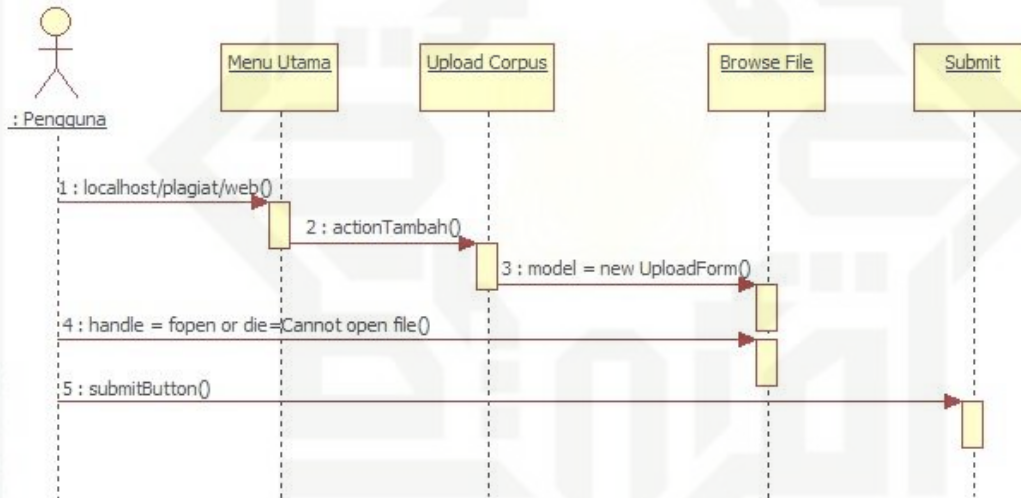


**Hak Cipta Dilindungi Undang-Undang**

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



**Gambar 4.3 Sequence Diagram Main Success Upload Corpus**

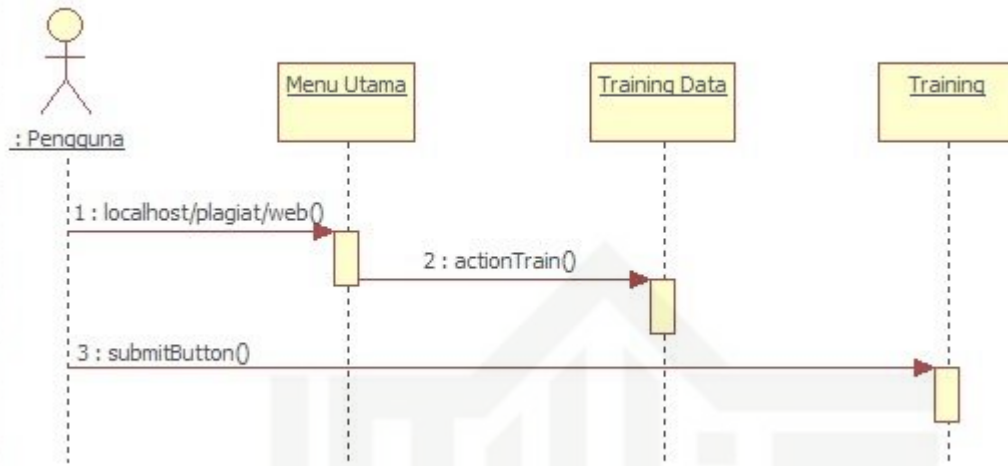


**Gambar 4.4 Sequence Diagram Exception Upload Corpus**

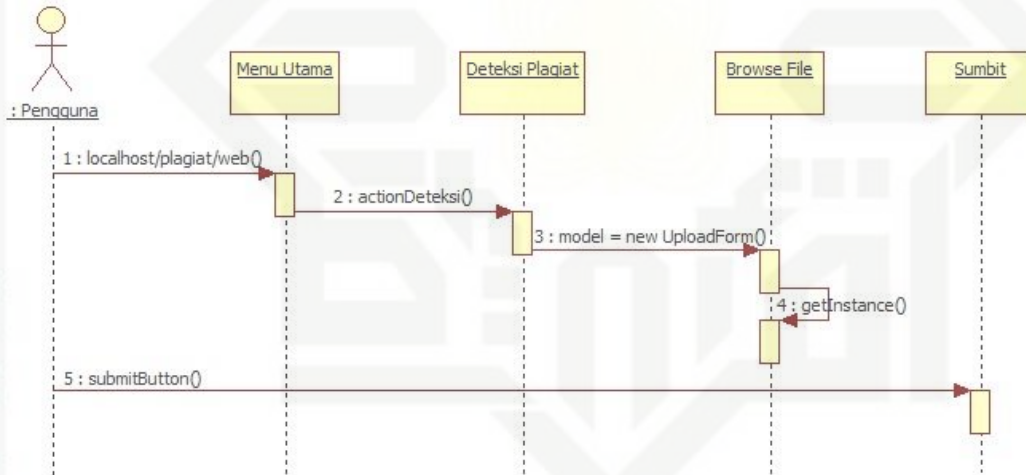


**Hak Cipta Dilindungi Undang-Undang**

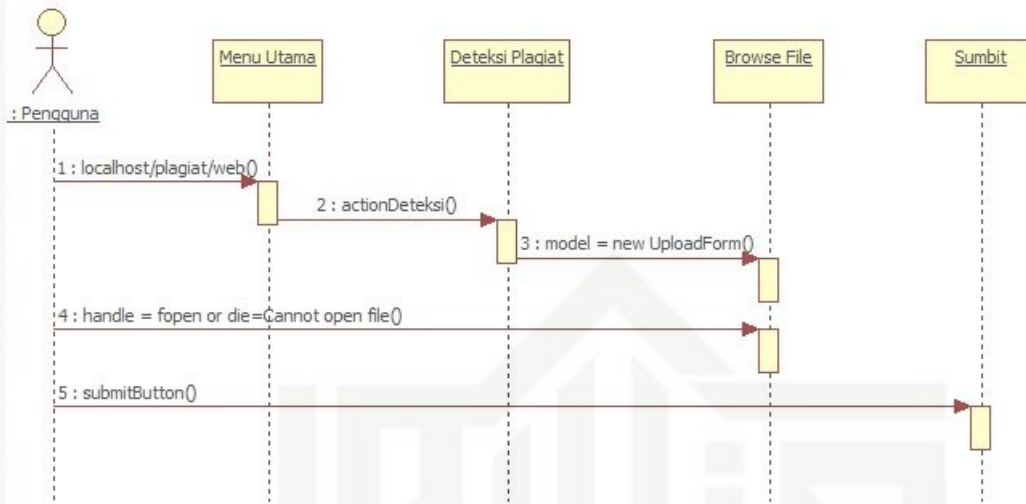
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah,
  - b. Pengutipan tidak merugikan kepentingan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



**Gambar 4.5 Sequence Diagram Main Success Training Data**



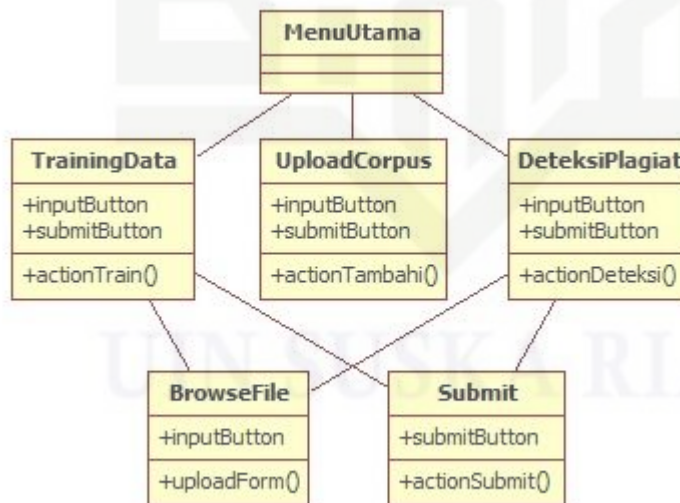
**Gambar 4.6 Sequence Diagram Main Success Deteksi Plagiat**



Gambar 4.7 Sequence Diagram Exception Deteksi Plagiat

### 4.3.3 Class Diagram

Class diagram digunakan untuk menggambarkan class pada sistem pendeteksian plagiarisme proposal tugas akhir. Dalam class diagram akan memperlihatkan aktifitas yang terjadi pada class di sistem pendeteksian plagiarisme proposal tugas akhir. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.



Gambar 4.8 Class Diagram

#### 4.4 Perancangan Antarmuka

Pada tahapan ini akan dirancang tampilan dari sistem berupa *prototype* yang kemudian akan diimplementasikan pada sistem. Pada sistem ini akan dibangun 4 *Menu*, yaitu menu beranda, *upload corpus*, *training data* dan deteksi. Pertama terdapat menu beranda yang berisikan informasi mengenai sistem deteksi plagiarisme tugas akhir. Menu yang kedua adalah *upload corpus*. Menu ini merupakan tempat untuk melakukan proses *upload data corpus*. Menu yang ketiga adalah *training data*. Menu ini merupakan proses wajib setelah melakukan *upload corpus*. Pengguna harus melakukan proses *training data* dengan memilih tombol *training data* agar matriks dapat dibuat. Menu yang keempat adalah deteksi plagiat. Pada menu ini, pengguna dapat melakukan proses deteksi dokumen yang dicurigai plagiat.

##### 1. *Prototype* Tampilan Menu Beranda

*Prototype* tampilan beranda merupakan layar awal dari sistem yang akan dibangun. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.



**Gambar 4.9** *Prototype* Tampilan Beranda Sistem

Dari gambar 4.6 di atas dapat dilihat bahwa pada *top bar* terdapat informasi mengenai nama sistem dan menu pada sistem. Menu yang ada di *top bar*, yaitu Beranda, *Upload Corpus*, *Training Data* dan Deteksi Plagiat. Menu Beranda berisi informasi mengenai sistem deteksi plagiarisme. Menu yang kedua, yaitu *Upload*



**Hak Cipta Dilindungi Undang-Undang**

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

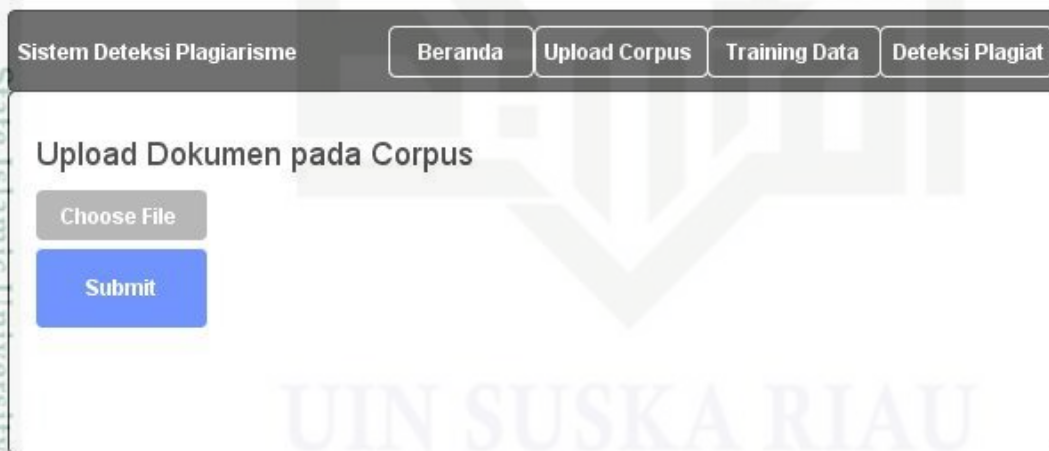
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

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

*Corpus*. Menu ini adalah lokasi untuk melakukan proses meng-*upload file* yang akan dijadikan *corpus*. Menu ini berisi tombol untuk melakukan *browse file* dan tombol *submit* untuk eksekusi *upload file*. Menu yang ketiga, yaitu *Training Data*. Pada menu ini akan dilakukan proses *training data* setelah dilakukan *upload corpus*. Menu ini terdapat tombol *training data* untuk melakukan eksekusi proses. Proses *training data* sangat penting dilakukan untuk mencari nilai matriks dari *file corpus* yang telah di-*upload*. Menu yang keempat, yaitu deteksi plagiat. Pada menu ini akan dilakukan proses pendeteksian plagiarisme dengan meng-*upload file* dokumen yang akan diperiksa. Menu ini berisi tombol untuk *browse* dokumen dan tombol *submit* untuk eksekusi proses.

2. *Prototype* Tampilan Halaman *Upload Corpus*

*Prototype* tampilan halaman *upload corpus* merupakan tampilan antarmuka yang digunakan untuk melakukan proses *upload corpus* pada sistem Pendeteksian Plagiarisme Proposal Tugas Akhir Menggunakan Algoritma *Latent Semantic Analysis* (LSA). Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.



**Gambar 4.10** *Prototype* Tampilan Halaman *Upload Corpus*

Pada gambar 4.7 dapat dilihat tampilan dari halaman *upload corpus* dari sistem Pendeteksian Plagiarisme Proposal Tugas Akhir Menggunakan Algoritma

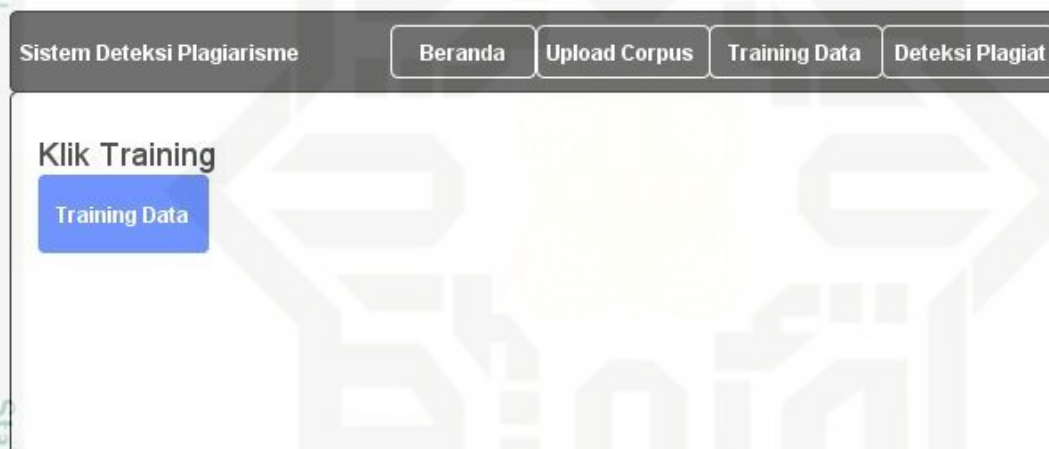
#### Hak Cipta Dilindungi Undang-Undang

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.

*Latent Semantic Analysis* (LSA) yang berisi tombol untuk memilih *file* yang akan dijadikan *corpus* dan tombol *submit* untuk melakukan eksekusi proses *upload corpus*. Proses *upload corpus* harus dilanjutkan dengan melakukan proses *training data* dengan memilih menu *trainig data* untuk mencari nilai matriks dari *file corpus* yang telah di-*upload*.

### 3. *Prototype* Tampilan Menu *Training Data*

*Prototype* Tampilan Menu *Training Data* merupakan tampilan antarmuka untuk melakukan proses *training data* setelah sebelumnya melakukan proses *upload corpus*. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.



**Gambar 4.11** *Prototype* Tampilan Menu *Training Data*

Pada gambar 4.8 di atas merupakan tampilan dari antarmuka menu *training data*. Pada menu ini berisi tombol *training data* untuk melakukan eksekusi proses. Proses *training data* dilakukan setelah selesai melakukan *upload corpus*. Proses ini dilakukan untuk mencari nilai matriks dari *file corpus* yang telah di-*upload*.

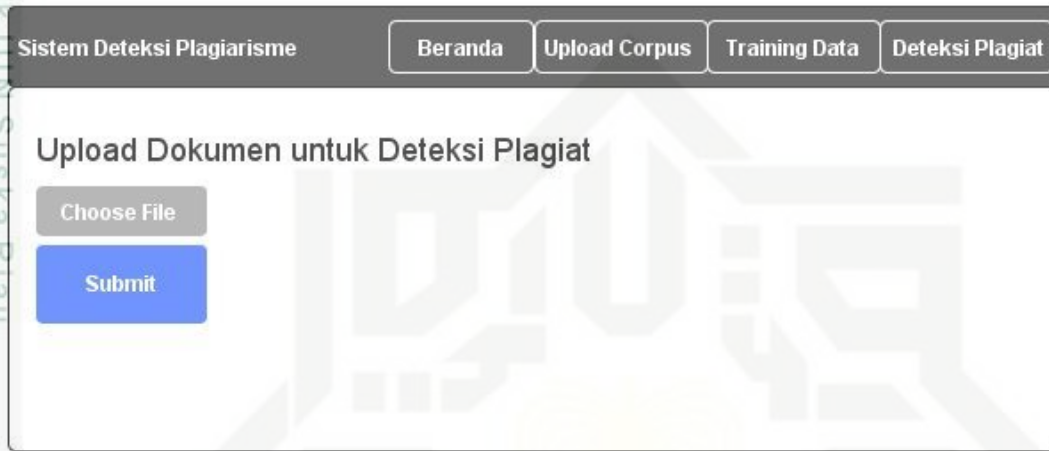


**Hak Cipta Dilindungi Undang-Undang**

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

4. *Prototype* Tampilan *Menu* Deteksi Plagiat

*Prototype* Tampilan *Menu* Deteksi Plagiat merupakan tampilan antarmuka untuk melakukan proses pendeteksian plagiarisme. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini.



**Gambar 4.12** *Prototype* Tampilan *Menu* Deteksi Plagiat

Pada gambar 4.9 dapat dilihat tampilan antarmuka dari menu deteksi plagiat. Pada menu ini berisi tombol untuk memilih *file* dokumen yang akan periksa. Proses deteksi plagiat ini tidak dapat dilakukan jika sebelumnya belum melakukan proses *training data* seperti yang dijelaskan pada bagian sebelumnya.