

BAB 4

IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Sistem

Pada tahap ini dilakukan penerjemahan perancangan berdasarkan analisis ke dalam bahasa pemrograman Python. Adapun implementasi sistem terdiri dari implementasi perangkat keras, implementasi perangkat lunak, dan implementasi antarmuka.

4.2 Implementasi Perangkat Keras

Perangkat keras yang digunakan dalam pembangunan sistem deteksi kesamaan dokumen dengan menggunakan metode *Vector Space Model* dan *Clustering* pada teks berbahasa Indonesia mempunyai spesifikasi sebagai berikut.

Tabel 4.1 Spesifikasi Perangkat Keras

No	Perangkat Keras	Spesifikasi
1	Processor	Intel Core i5-3317U, 1.7 Ghz
2	Memori (RAM)	4 GB
3	HDD	500 GB

4.3 Implementasi Perangkat Lunak

Spesifikasi perangkat lunak yang digunakan dalam implementasi sistem dapat dilihat pada table 4.2.

Tabel 4.2 Spesifikasi Perangkat lunak

No	Perangkat Lunak	Spesifikasi
1	Sistem Operasi	Window 7 Enterprise 64-bit
2	Project Interpreter	Python 2.7
3	Project Editor	PyCharm 2018.1.2 (Community Edition)
4	Package (Python)	- nltk Version 3.3 - Sastrawi Version 1.0.1

4.4 Implementasi Antarmuka

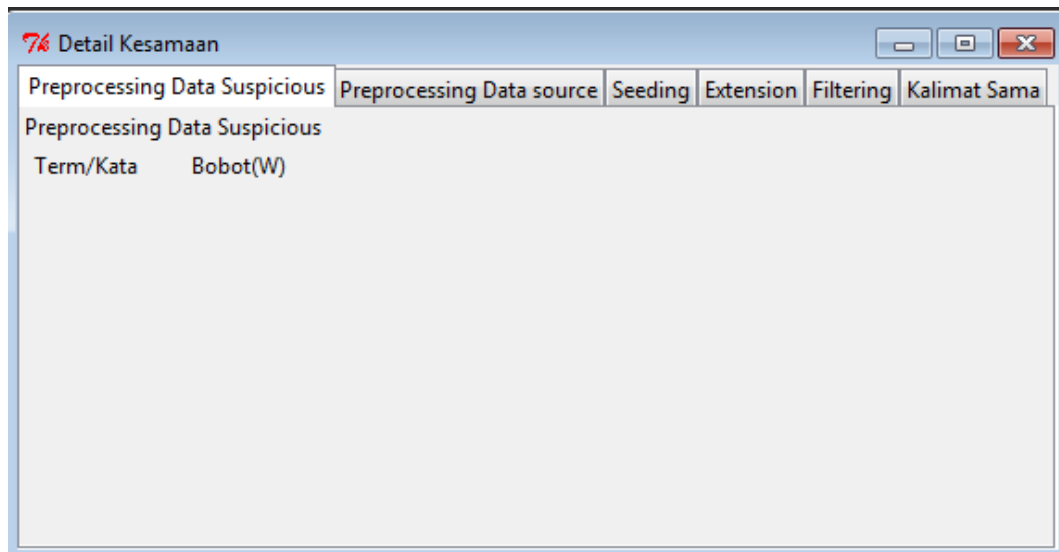
Implementasi antarmuka dilakukan untuk setiap tampilan aplikasi yang dibangun. Penjelasan implementasi antarmuka pada aplikasi deteksi kesamaan dokume menggunakan VSM dan *clustering* dapat dilihat pada tabel 4.3.

Tabel 4.3 Implementasi Antarmuka

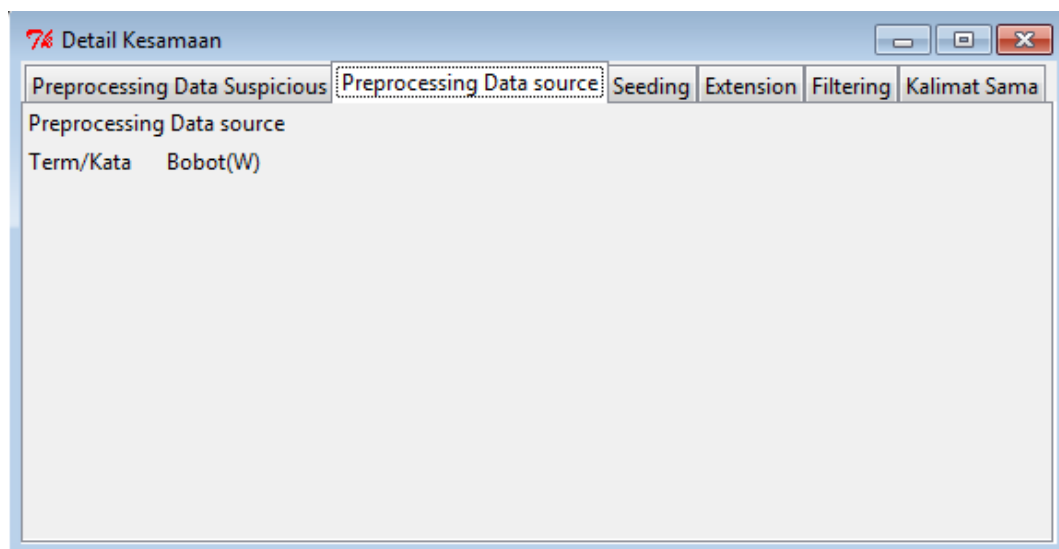
No	Nama Antarmuka	Deskripsi	Nama File
1	Menu Utama	Menampilkan halaman utama aplikasi yang berupa input data suspicious dan source, input nilai parameter, dan informasi hasil proses perbandingan yang berupa persentase kesamaan dan waktu eksekusi.	DeteksiKesamaan.py
2	Tampil Detail	Menampilkan hasil dari proses TF-IDF, <i>seeding, extension, filtering</i> dan kalimat sama.	DeteksiKesamaan.py

Implementasi antarmuka dapat dilihat pada gambar gambar berikut.

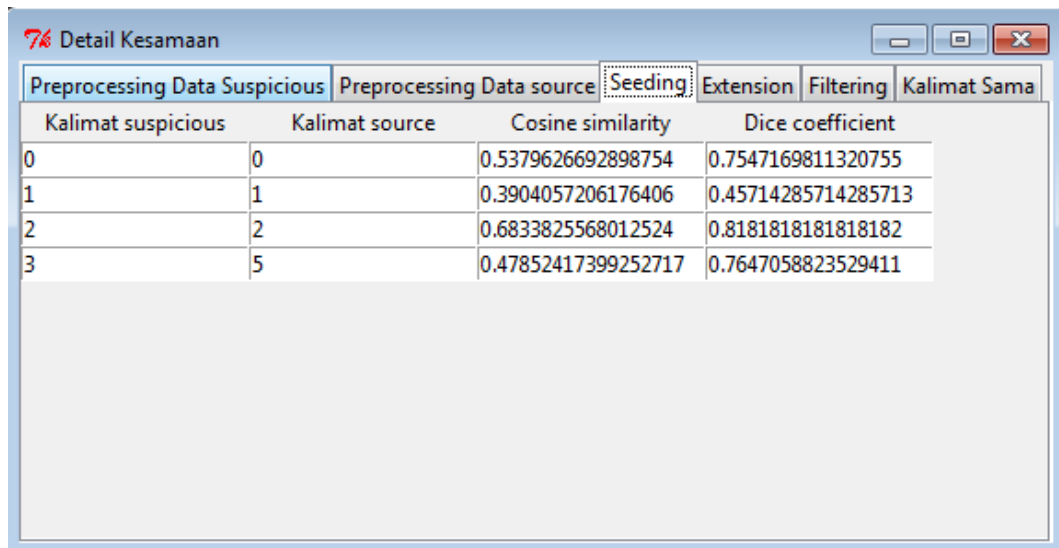
Gambar 4.1 Halaman Utama



Gambar 4.2 Hasil TF-IDF pada dokumen suspicious

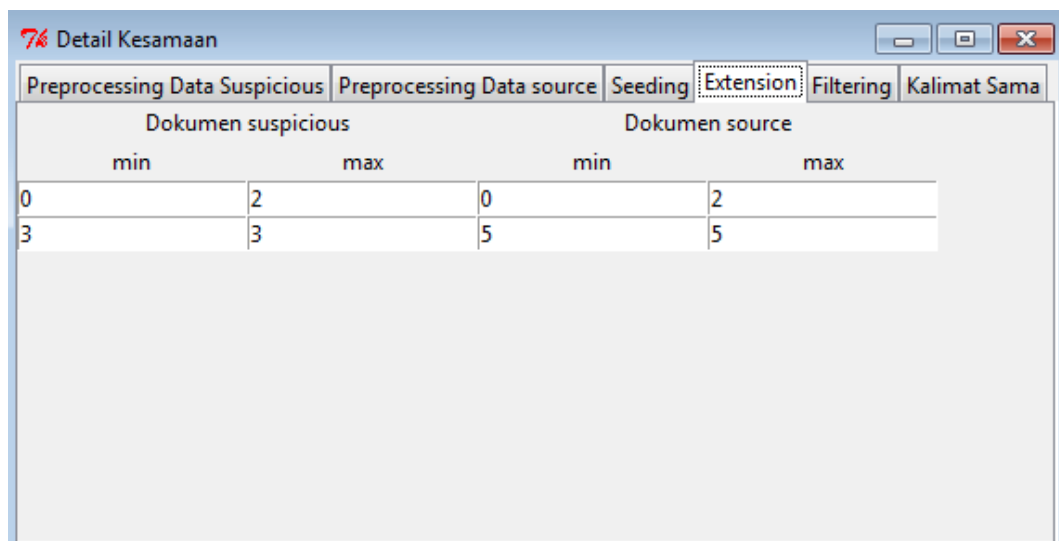


Gambar 4.3 Hasil TF-IDF pada dokumen source



Kalimat suspicious	Kalimat source	Cosine similarity	Dice coefficient
0	0	0.5379626692898754	0.7547169811320755
1	1	0.3904057206176406	0.45714285714285713
2	2	0.6833825568012524	0.8181818181818182
3	5	0.47852417399252717	0.7647058823529411

Gambar 4.4 Hasil proses seeding



Dokumen suspicious		Dokumen source	
min	max	min	max
0	2	0	2
3	3	5	5

Gambar 4.5 Hasil Proses Extension

Dokumen suspicious		Dokumen source	
min	max	min	max
0	2	0	2
3	3	5	5

Gambar 4.6 Hasil proses filtering

No	Dokumen suspicious	Dokumen source
1	plagiat adalah perbuat:	plagiat adalah perbuat:
2	jenis-jenisnya diantara	terdapat beberapa jenis
3	plagiat merupakan tinc	plagiat dapat dianggap
4	adapun cara untuk me	selain itu langkah lain u

Gambar 4.7 Kalimat yang terindikasi sama

4.5 Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan bahwa sistem yang dibuat sesuai dengan rancangan sistem yang telah dibangun sebelumnya. Pada tahap pengujian sistem terdiri dari pengujian fungsionalitas dan pengujian akurasi.

4.5.1 Pengujian Fungsionalitas

Pengujian fungsionalitas yang dilakukan menggunakan metode *black box*, output yang dihasilkan metode ini berdasarkan input dan kondisi eksekusi yang dipilih untuk menampilkan output atau proses dari setiap tahapan. Adapun pengujian fungsionalitas dapat dilihat pada table 4.4.

Tabel 4.4 Pengujian Fungsionalitas

No	Nama Proses	Poin Pengujian	Jenis Pengujian
1	<i>Preprocessing</i>	<i>TF-IDF</i>	<i>Black Box</i>
2	<i>Seeding</i>	<i>Indeks kalimat sama</i>	<i>Black Box</i>
		Cosine similarity	<i>Black Box</i>
		<i>Dice coefficient</i>	<i>Black Box</i>
3	Extension	Banyak cluster	<i>Black Box</i>
		Nilai min dan max pada setiap cluster	<i>Black Box</i>
4	Filtering	Banyak cluster hasil filtering	<i>Black Box</i>
		Nilai min dan max pada setiap cluster hasil filtering	<i>Black Box</i>
5	Kalimat sama	Banyak kalimat sama	<i>Black Box</i>

4.5.2 Hasil Pengujian Fungsionalitas

Pengujian black box berfokus pada persyaratan fungsional dari sebuah sistem yang dibangun dan menemukan kesalahan program. Hasil pengujian untuk antarmuka dapat dilihat pada Tabel.

Tabel 4.5 Hasil pengujian fungsionalitas

Aktivitas yang Dilakukan	Data Masukan	Keluaran yang Diharapkan	Pengamatan	Kesimpulan
<i>Prerocessing</i>	Suspicious dan source	TF-IDF	Menampilkan TF-IDF	[√] Diterima [] Ditolak
<i>Seeding</i>	Hasil preprocessing	Indeks kalimat sama beserta nilai cosinus dan dice	Menampilkan Indeks kalimat sama beserta nilai cosinus dan dice	[√] Diterima [] Ditolak
<i>Extension</i>	Hasil seeding	Nilai min dan max dari setiap cluster	Menampilkan Nilai min dan max dari setiap cluster	[√] Diterima [] Ditolak
<i>Filtering</i>	Hasil extension	Nilai min dan max dari setiap cluster	Menampilkan Nilai min dan max dari setiap cluster	[√] Diterima [] Ditolak
Kalimat sama	Hasil filtering	Banyak kalimat sama	Menampilkan banyak kalimat sama	[√] Diterima [] Ditolak

4.5.3 Pengujian Akurasi

Pengujian akurasi pada penelitian ini dilakukan pengukuran akurasi kesamaan dengan menggunakan metode evaluasi instrinsik ROUGE [9] untuk mengetahui nilai akurasi untuk dapat menentukan apakah *Vector Space Model* dan *Clustering* cocok digunakan untuk deteksi kesamaan dalam dokumen berbahasa Indonesia.

Jumlah data yang digunakan adalah dua abstrak jurnal yang digunakan sebagai data *suspicious* dan 30 abstrak jurnal yang digunakan sebagai data *source*. Nilai parameter yang digunakan untuk percobaan ini sama dengan yang digunakan Sanchez-perez [7] yaitu :

Nilai Parameter :

Parameter constanta : $minsize = 1$ $maxgap_least = 0$

Parameter inputan : $mincose = 0,30$ $maxgap = 1$

$Mindice = 0,33$ $minplaglen = 150$ $th_validation = 0,34$

Tabel 4.6 Data Suspicious dan Source

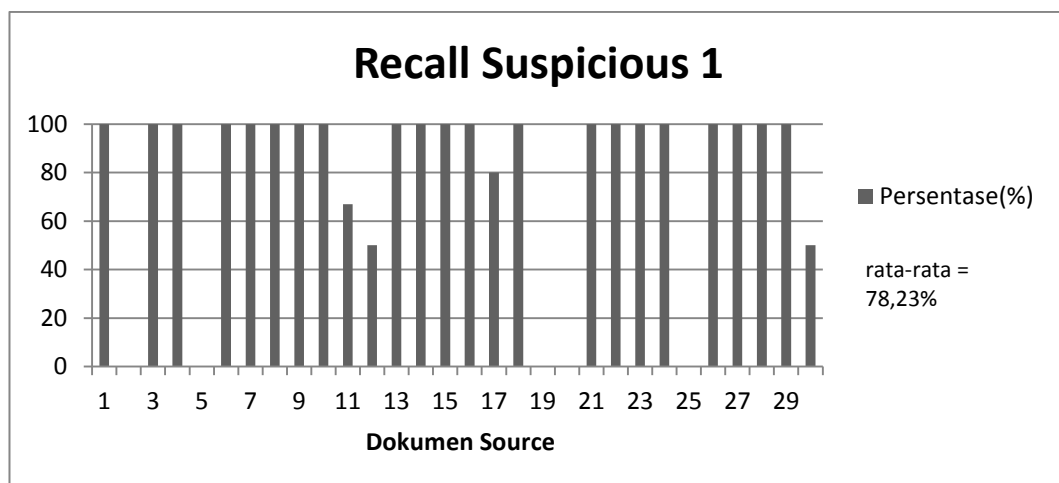
No	Data	Tipe	Jumlah kata
1	jbptunikompp-gdl-donibayuak-40064-11-20.unik-a	Suspicious 1	169
2	jbptunikompp-gdl-luthfiaulv-36482-7-unikom_1-a	Suspicious 2	214
3	jbptunikompp-gdl-adangsaepu-38301-6-unikom_a-a	Source 1	226
4	jbptunikompp-gdl-addriansap-39851-7-20.unik-a	Source 2	252
5	jbptunikompp-gdl-agiesudyan-39056-6-unikom_a-a	Source 3	249
6	jbptunikompp-gdl-ariefifand-36481-10-20.unik-a	Source 4	206
7	jbptunikompp-gdl-cahyadipri-38512-7-unikom_c-a	Source 5	245
8	jbptunikompp-gdl-herdizulfi-40070-10-20.10113-a	Source 6	190
9	jbptunikompp-gdl-indrahandi-39287-4-unikom_i-a	Source 7	189
10	jbptunikompp-gdl-indraharya-38433-7-unikom_i-a	Source 8	241
11	jbptunikompp-gdl-ingeulizzi-37775-7-20.uniko-a	Source 9	210
12	jbptunikompp-gdl-iqbalsyahi-36852-11-20.uniko-a	Source 10	305
13	jbptunikompp-gdl-jihanansha-38603-10-20.10112-a	Source 11	167
14	jbptunikompp-gdl-kikiirawan-39474-6-unikom_k-a	Source 12	208

No	Data	Tipe	Jumlah kata
15	jbptunikompp-gdl-megapuspit-36395-8-unikom_m-a	Source 13	260
16	jbptunikompp-gdl-mohamadbur-39283-5-unikom_m-a	Source 14	252
17	jbptunikompp-gdl-mohammadsa-36898-8-unikom_m-a	Source 15	177
18	jbptunikompp-gdl-mranggamah-39923-6-unikom_m-a	Source 16	196
19	jbptunikompp-gdl-muhamadiqb-39110-6-unikom_m-a	Source 17	219
20	jbptunikompp-gdl-muhammadal-39924-9-2110113-a	Source 18	201
21	jbptunikompp-gdl-noviyantos-39170-7-unikom_n-a	Source 19	201
22	jbptunikompp-gdl-ogiesyaput-39046-9-20.unik-a	Source 20	209
23	jbptunikompp-gdl-rahmatsyap-38503-5-unikom_r-a	Source 21	185
24	jbptunikompp-gdl-rakaginipe-39261-7-unikom_r-a	Source 22	199
25	jbptunikompp-gdl-rezanurulf-38434-6-unikom_r-a	Source 23	209
26	jbptunikompp-gdl-riarizkiaw-36386-7-unikom_r-a	Source 24	248
27	jbptunikompp-gdl-romiwadisa-39256-5-unikom_r-a	Source 25	215
28	jbptunikompp-gdl-sutrianisa-37762-11-20.uniko-a	Source 26	198
29	jbptunikompp-gdl-tresnadwig-39289-4-	Source 27	259

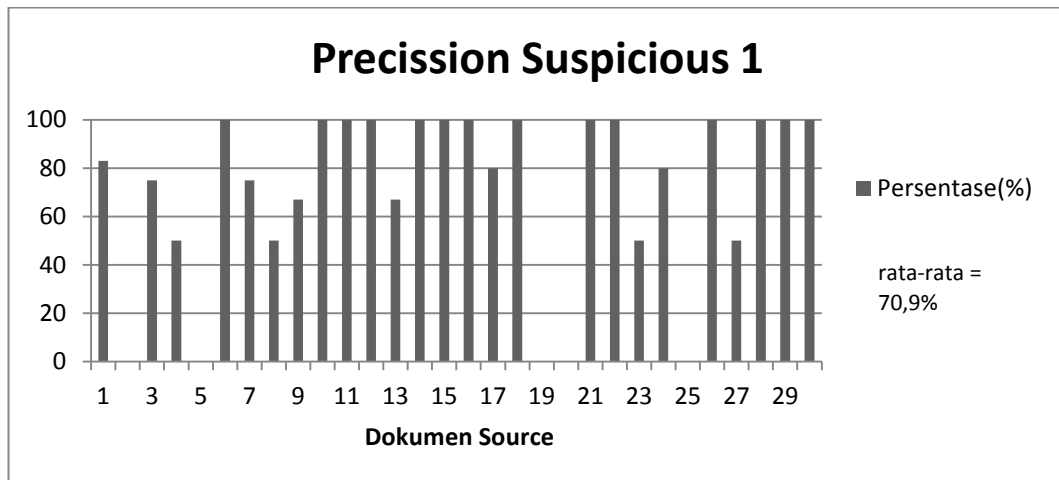
No	Data	Tipe	Jumlah kata
	unikom_t-a		
30	jbptunikompp-gdl-wanidinnim-40050-6-unikom_w-a	Source 28	240
31	jbptunikompp-gdl-yantielfri-37807-7-unikom_y-a	Source 29	244
32	jbptunikompp-gdl-yudiahmadr-37778-9-unikom_y-a	Source 30	200

4.5.4 Hasil Pengujian Akurasi

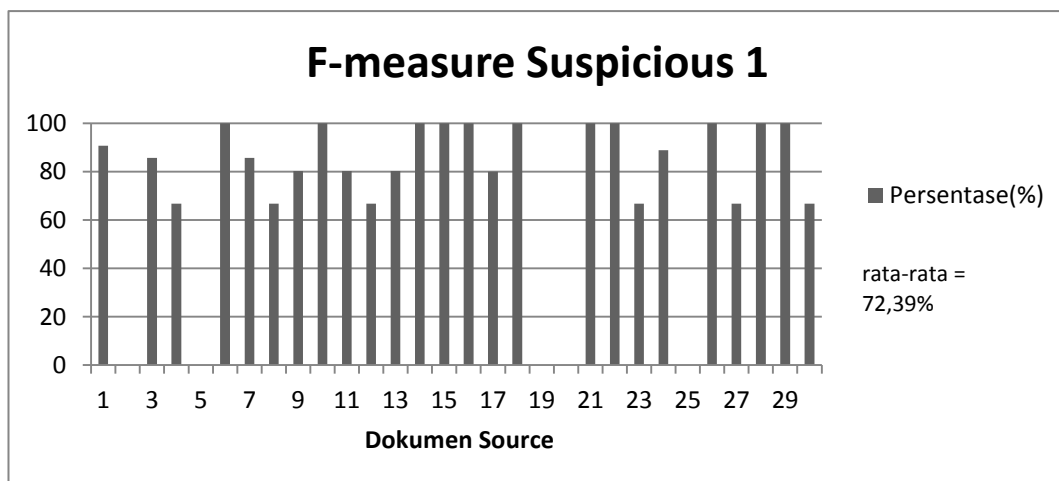
Pengujian nilai akurasi deteksi kesamaan dokumen dengan metode *Vector Space Model* dan *Clustering* dapat dilihat sebagai berikut.



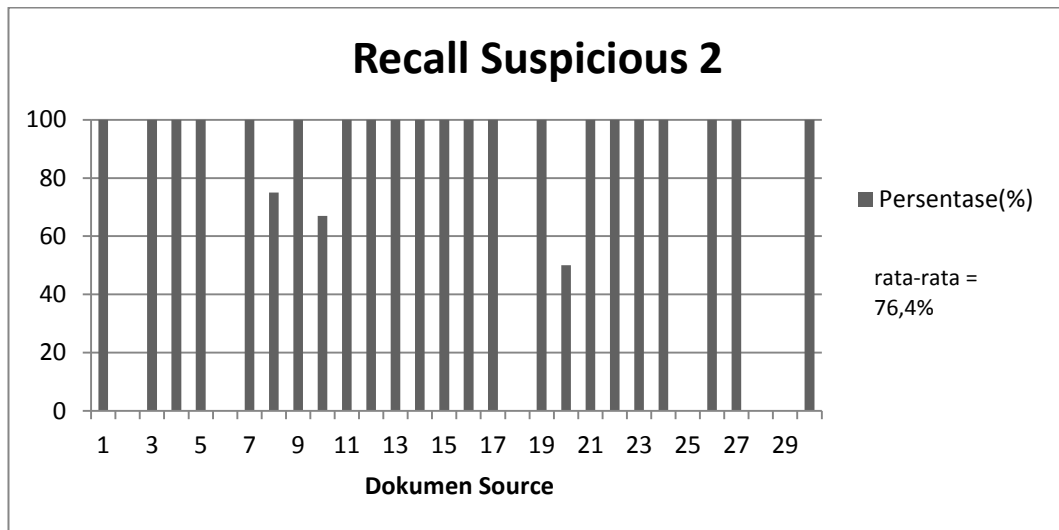
Gambar 4.8 Recall suspicious 1



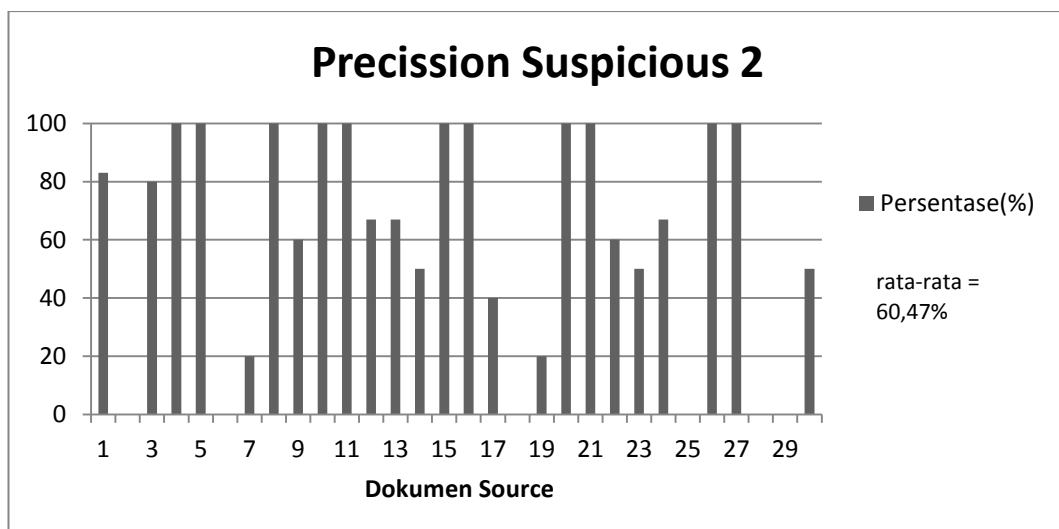
Gambar 4.9 *Precision suspicious 1*



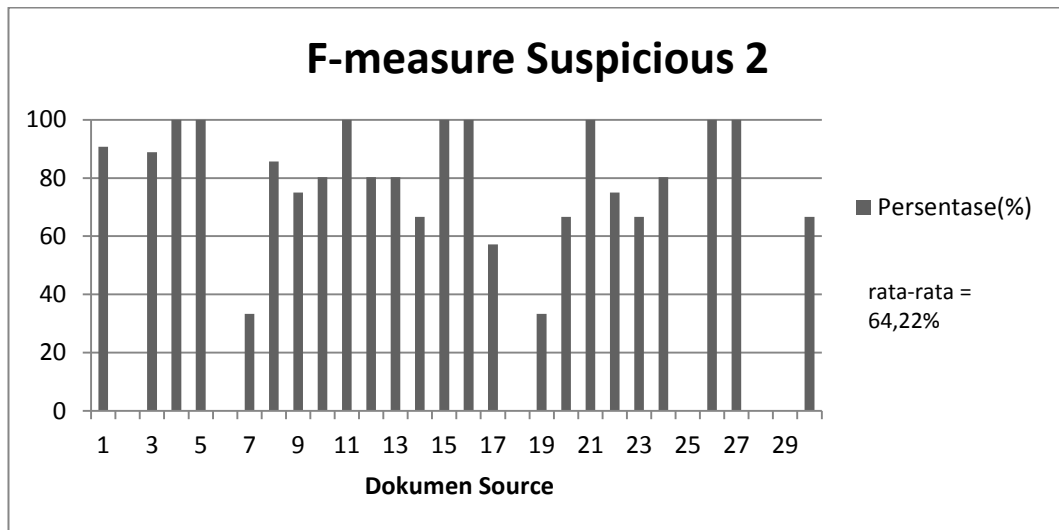
Gambar 4.10 *F-measure suspicious 1*



Gambar 4.11 *Recall Suspicious 2*



Gambar 4.12 *Precision Suspicious 2*



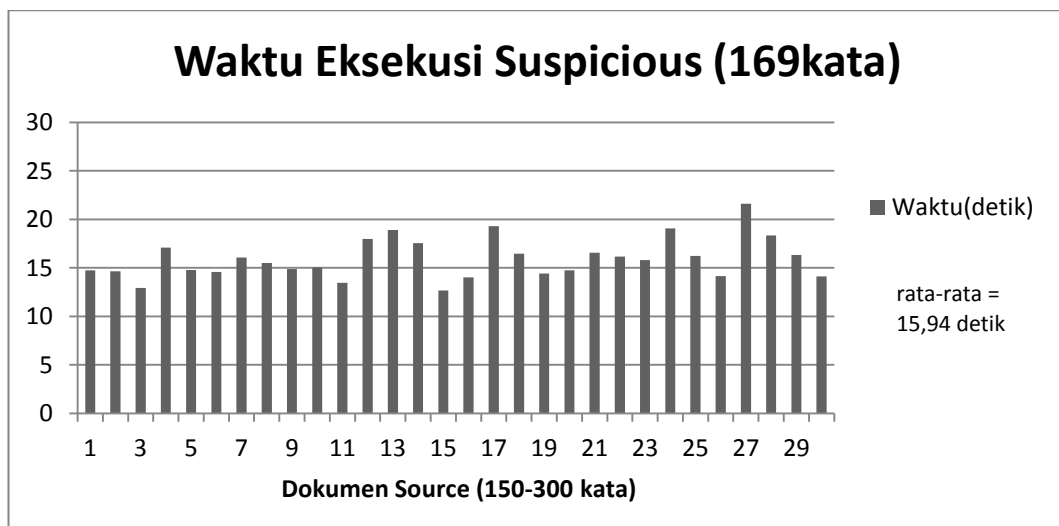
Gambar 4.13 F-measure Suspicious 2

Dari hasil pengujian menggunakan 2 dokumen suspicious dan 30 dokumen source maka dihasilkan rata-rata *recall*, *precision*, dan *f-measure* sebagai berikut.

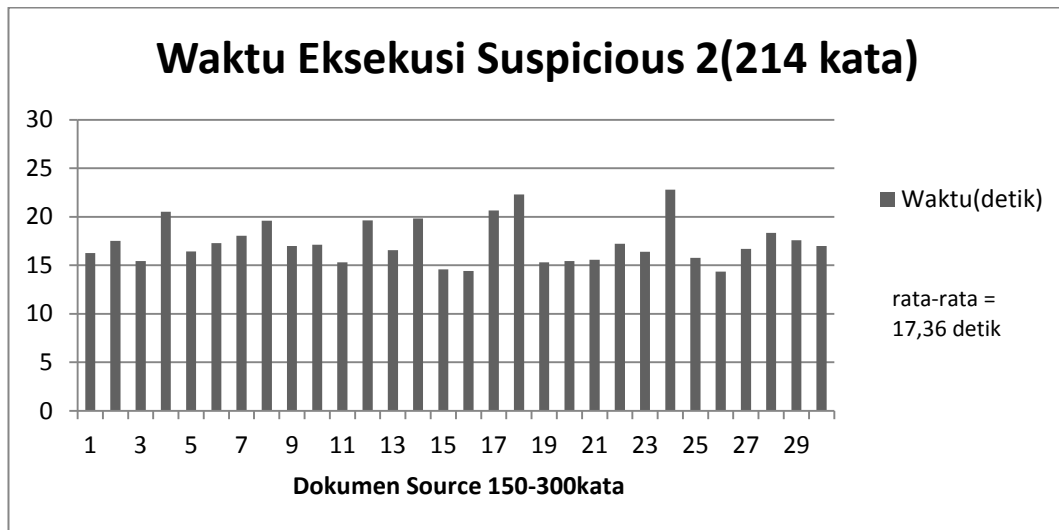
$$recall = \frac{78,23+76,4}{2} \times 100\% = 77,32\%$$

$$precision = \frac{70,9 + 60,47}{2} \times 100\% = 65,68\%$$

$$f-measure = \frac{72,39+64,22}{2} \times 100\% = 68,31\%$$



Gambar 4.14 Waktu Eksekusi Suspicious 1



Gambar 4.15 Waktu Eksekusi *Suspicious 2*

$$\text{Rata-rata waktu eksekusi} = \frac{15,94+17,36}{2} = 16,65 \text{ detik}$$

Pengujian sistem yang dilakukan menunjukkan bahwa aplikasi yang dibuat untuk deteksi kesamaan dokumen sudah memenuhi persyaratan fungsional. Analisis hasil pengujian deteksi kesamaan dokumen menggunakan metode *vector space model* dan *clustering* menghasilkan nilai rata-rata *recall* sebesar 77,32%, *precision* sebesar 65,68% dan *f-measure* sebesar 68,31%. Rata-rata waktu eksekusi yaitu 16,65 detik. Akurasi yang dihasilkan terbilang rendah dan lama, penyebabnya yaitu :

1. Dibutuhkan nilai parameter berbeda untuk setiap perbandingan. Nilai parameter yang satu belum tentu optimal jika digunakan untuk perbandingan lainnya.
2. Kurangnya data set yang di ujikan.
3. Waktu eksekusi lama disebabkan karena satu persatu kalimat pada dua buah dokumen yang dibandingkan akan dibandingkan sebagai contoh pada dokumen A terdapat 20 kalimat, dokumen B terdapat 30 kalimat maka total perbandingan kalimat jika dokumen A dan B dibandingkan yaitu sebesar 600 perbandingan kalimat.