BAB 4
IMPLEMENTASI DAN PENGUJIAN

4.1. Implementasi Sistem

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

4.1.1. Implementasi Perangkat Keras

Perangkat keras yang digunakan dalam pembangunan sistem *Coreference Resolution* menggunakan *Support Vector Machine* pada novel berbahasa Indonesia dengan pertimbangan ketersediaan dan kemampuannya yang dapat mengakses program *Coreference Resolution* dengan spesifikasi sebagai berikut.

Tabel 4. 1 Spesifikasi Perangkat Keras untuk Pembangunan

No	Perangkat Keras	Spesifikasi
1	Processor	Intel Core i3-4005U, 1.7GHz
2	Memori (RAM)	2 GB
3	*Hard disk*	500 GB

4.1.2. Implementasi Perangkat Lunak

Spesifikasi perangkat lunak yang digunakan dalam implementasi sistem dengan spesifikasi minimal dapat dilihat pada tabel 4.2 berikut ini.

Tabel 4. 2 Spesifikasi Minimum Perangkat Lunak

No	Perangkat Keras	Spesifikasi
1	Sistem Operasi	Windows 7 - 32 Bit
2	HTML *Editor*	Macromedia Dreamwever 8
3	*Web Server*	XAMPP
4	*Web Browser*	Chrome

4.1.3. Implementasi Antarmuka

Implementasi antarmuka dilakukan untuk setiap tampilan aplikasi yang dibangun. Penjelasan implementasi antarmuka pada aplikasi *Coreference Resolution* menggunakan *Support Vector Machine* dapat dilihat pada tabel 4.3.

Tabel 4. 3 Implementasi Antarmuka

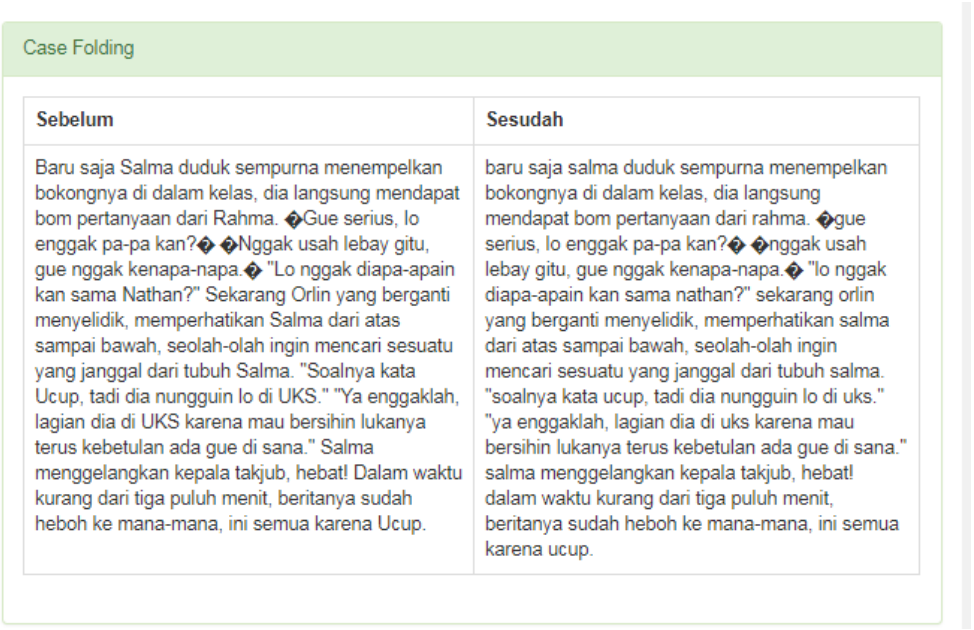
No	Nama Antarmuka	Deskripsi	Nama File
1	*Training*	Menampilkan halaman *Training* yang meliputi tahap *input* data *training*, *preprocessing* (*case folding*, *filtering*, *tokenizing* kalimat, *tokenizing* kata, *stemming*, POS *tagging*), ekstraksi fitur dan *training* SVM.	trainingpanel.php
2	*Testing*	Menampilkan halaman *Testing* yang meliputi tahap *input* data *testing*, *preprocessing* (*case folding*, *filtering*, *tokenizing* kalimat, *tokenizing* kata, *stemming*, POS *tagging*), ekstraksi fitur, hasil klasifikasi SVM dan akurasi.	testingpanel.php

Berikut adalah tampilan antar muka sistem *coreference resolution.* Di bawah ini merupakan tampilan input data training dimana pengguna harus mengunduh terlebih dahulu data teks novel dengan format .txt.



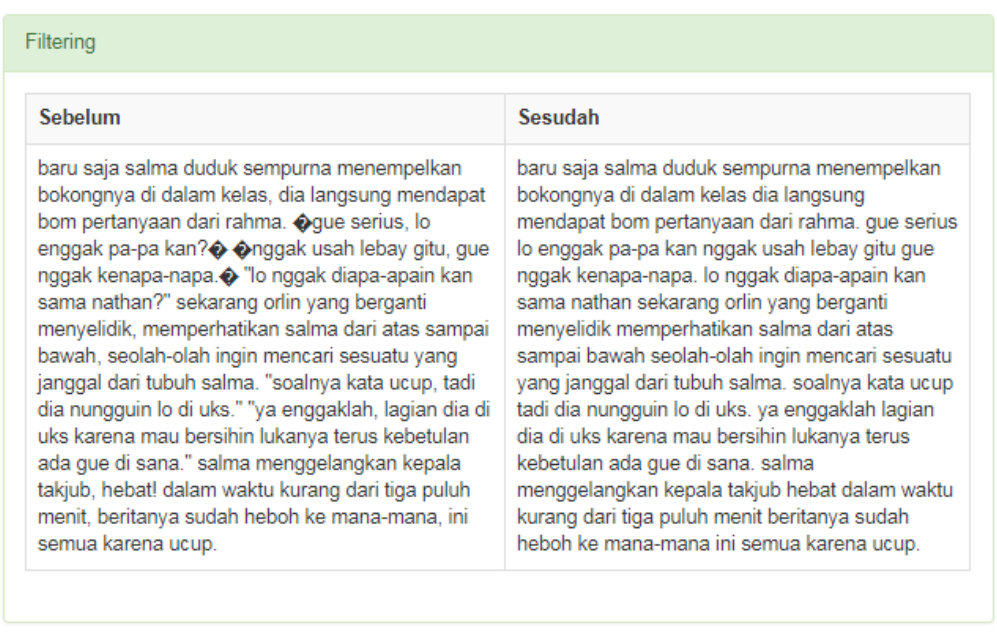
Gambar 4. 1 Antarmuka Input Data Training

Tahap selanjutnya adalah memasuki tahap preprocessing. Tahap yang pertama adalah tahap *case folding* dimana sistem menyeragamkan teks menjadi huruf kecil.



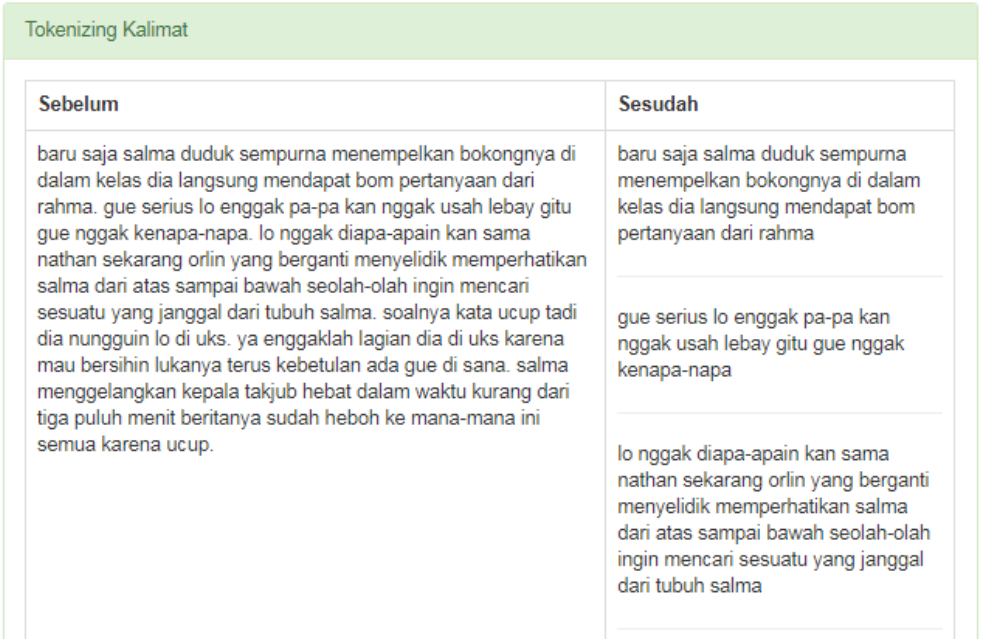
Gambar 4. 2 Antarmuka Case Folding

Tahap yang kedua adalah tahap *filtering* dimana sistem membuang karakter tidak penting di dalam teks.



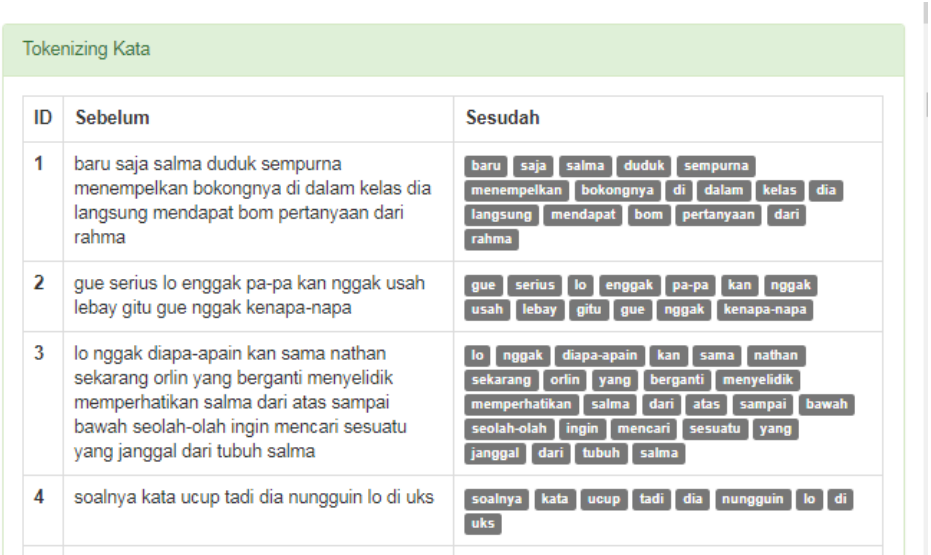
Gambar 4. 3 Antarmuka Filtering

Tahap yang ketiga adalah tahap *tokenizing* kalimat dimana sistem memecah teks menjadi kumpulan kalimat.



Gambar 4. 4 Antarmuka Tokenisasi Kalimat

Tahap yang keempat adalah tahap *tokenizing* kata dimana sistem memecah kalimat menjadi kumpulan kata.



Gambar 4. 5 Antarmuka Tokenisasi Kata

Tahap yang kelima adalah tahap *stemming* dimana sistem mengubah kata berimbuhan menjadi kata dasar.


Stemming

Kalimat 1

Sebelum	Sesudah
baru	baru
saja	saja
salma	salma
duduk	duduk
sempurna	sempurna
tempel	tempel
bokong	bokong
di	di

Gambar 4. 6 Antarmuka Stemming

Tahap yang terakhir adalah tahap POS tagging dimana sistem memberi tag pada setiap kata.

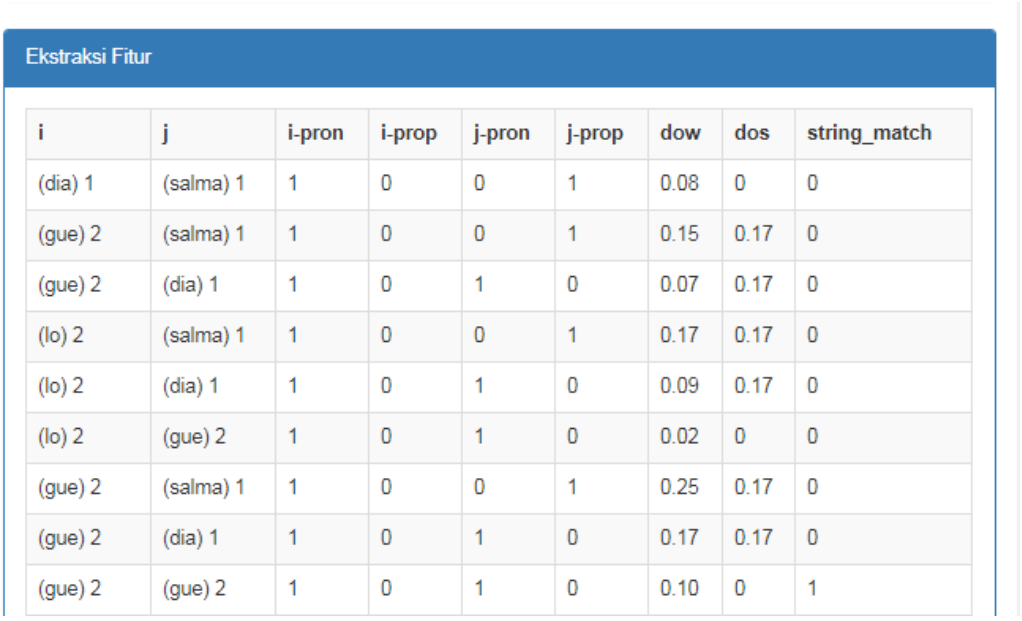

POS tag

Kalimat 1

Kata	POS Tag
baru	JJ
saja	RB
salma	NNP
duduk	VBI
sempurna	JJ
tempel	NN
bokong	NN
di	IN

Gambar 4. 7 Antarmuka POS Tagging

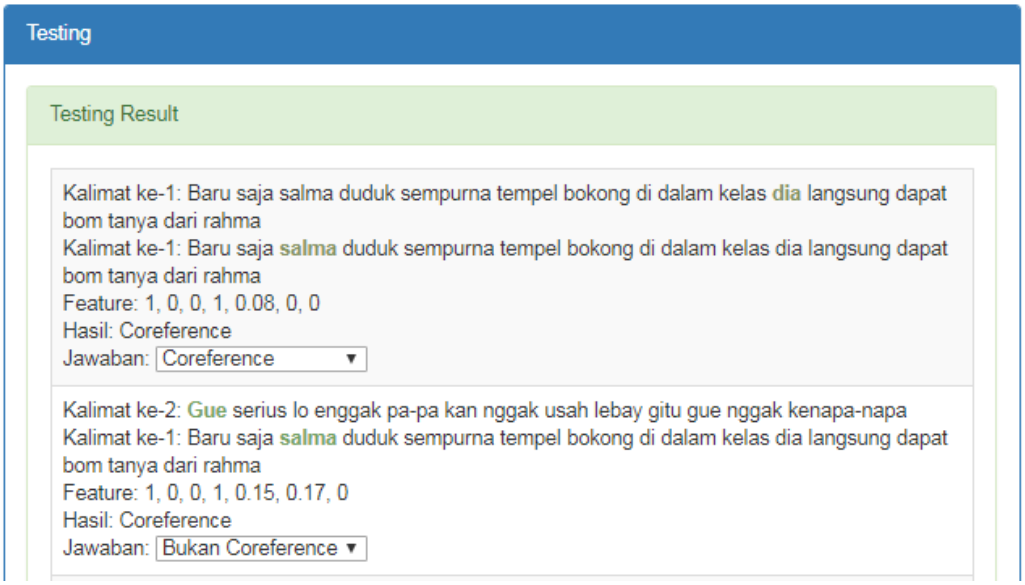
Selanjutnya memasuki tahap ekstraksi fitur dimana sistem mengekstrak informasi yang ada menjadi nilai-nilai yang kemudian akan digunakan untuk tahap klasifikasi.



i	j	i-pron	i-prop	j-pron	j-prop	dow	dos	string_match
(dia) 1	(salma) 1	1	0	0	1	0.08	0	0
(gue) 2	(salma) 1	1	0	0	1	0.15	0.17	0
(gue) 2	(dia) 1	1	0	1	0	0.07	0.17	0
(lo) 2	(salma) 1	1	0	0	1	0.17	0.17	0
(lo) 2	(dia) 1	1	0	1	0	0.09	0.17	0
(lo) 2	(gue) 2	1	0	1	0	0.02	0	0
(gue) 2	(salma) 1	1	0	0	1	0.25	0.17	0
(gue) 2	(dia) 1	1	0	1	0	0.17	0.17	0
(gue) 2	(gue) 2	1	0	1	0	0.10	0	1

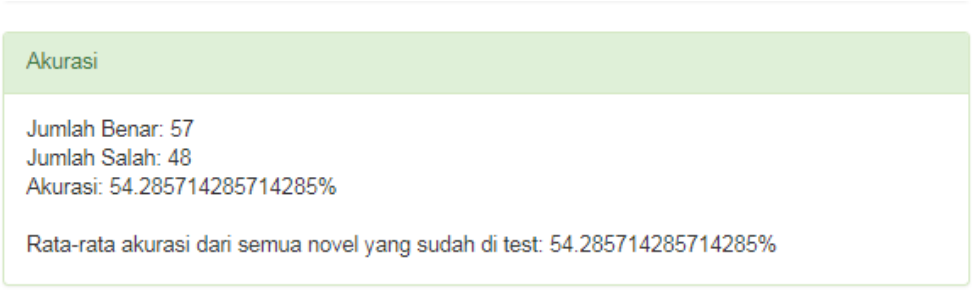
Gambar 4. 8 Antarmuka Ekstraksi Fitur

Berikut adalah tampilan hasil klasifikasi dari SVM.



Gambar 4. 9 Antarmuka Hasil Klasifikasi

Dan yang terakhir adalah hasil akurasi.



Gambar 4. 10 Antarmuka Akurasi

4.2. 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.

4.2.1. Pengujian Fungsionalitas

Pengujian fungsionalitas yang dilakukan menggunakan metode *blackbox*, *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 tabel 4.4 berikut.

Tabel 4. 4 Pengujian Fungsionalitas

No	Nama Proses	Point Pengujian	Jenis Pengujian
1	*Preprocessing*	*Case Folding*	*Black box*
		Filtering	*Black box*
		Tokenisasi Kalimat	*Black box*
		Tokenisasi Kata	*Black Box*

		Stemming	*Black Box*
		POS *tagging*	*Black Box*
2	Ekstraksi Fitur	Penentuan pasangan *coreference*	*Black box*
		Penentuan nilai fitur	*Black box*
3	*Training SVM*	Pelatihan terhadap SVM	*Black Box*
4	*Testing SVM*	Pengklasifikasian pasangan *coreference*	*Black Box*

4.2.2. 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 4.5.

Tabel 4. 5 Pengujian Fungsionalitas

Kasus dan Data Hasil Uji (Data Benar)				
Aktifitas yang dilakukan	**Data Masukan**	**Keluaran yang diharapkan**	**Pengamatan**	**Kesimpulan**
Case Folding	Teks novel	Teks novel dengan huruf kecil	Menampilkan teks novel dengan huruf kecil (*lower case*)	[√] Diterima [ ] Ditolak
Filtering	Teks novel dengan huruf kecil	Teks novel tanpa karakter selain ‘a’-‘z’, ‘-‘, ‘.’ dan spasi ‘ ‘	Menampilkan teks novel tanpa karakter selain ‘a’-‘z’, ‘-‘, ‘.’ dan spasi ‘ ‘	[√] Diterima [ ] Ditolak
Tokenisasi Kalimat	Teks novel tanpa karakter selain ‘a’-‘z’, ‘-‘, ‘.’ dan spasi ‘ ‘	Kumpulan kalimat tanpa delimeter titik	Menampilkan kumpulan kalimat tanpa delimeter titik	[√] Diterima [ ] Ditolak

Tokenisasi Kata	Kumpulan kalimat	Kumpulan kata	Menampilkan kumpulan kata	[√] Diterima [ ] Ditolak
Stemming	Kumpulan kata	Kumpulan kata dasar	Menampilkan kumpulan kata dasar	[√] Diterima [ ] Ditolak
POS *Tagging*	Kumpulan kata dasar	Kumpulan kata dasar beserta masing-masing tag setiap kata	Menampilkan kumpulan kata dasar beserta masing-masing tag setiap kata	[√] Diterima [ ] Ditolak
Ekstraksi Fitur	Kumpulan kata dasar beserta masing-masing tag setiap kata dan label	Kumpulan kumpulan pasangan *coreference* berserta hasil ekstraksi fitur berupa angka	Menampilkan kumpulan pasangan *coreference* berserta hasil ekstraksi fitur berupa angka	[√] Diterima [ ] Ditolak
Training	Kumpulan angka hasil ekstraksi fitur dengan format SVM dan label	Model SVM	Menampilkan notifikasi bahwa training berhasil dan model tersimpan di svm_model.svm	[√] Diterima [ ] Ditolak
Testing	Kumpulan angka hasil ekstraksi fitur dengan format SVM dan data model_svm.svm	Hasil klasifikasi SVM yang berupa pasangan *coreference* serta kelas yang dimiliki masing-masing pasangan	Menampilkan hasil klasifikasi SVM yang berupa pasangan *coreference* serta kelas yang dimiliki masing-masing pasangan	[√] Diterima [ ] Ditolak

4.2.3. Pengujian Akurasi

Pengujian akurasi pada penelitian *Coreference Resolution* ini dilakukan untuk mengetahui nilai akurasi untuk dapat menentukan apakah metode *Support Vector Machine* cocok digunakan untuk *Coreference Resolution* pada novel berbahasa Indonesia.

Jumlah data yang digunakan adalah 2 novel yang digunakan sebagai data latih karena semakin beragam novel yang menjadi data latih maka akan semakin besar akurasinya. Hal ini dikarenakan perbedaan pola yang digunakan setiap novel. Novel pertama yang digunakan sebagai data latih merupakan novel yang berjudul

"Surga Sungsang" sedangkan novel yang kedua adalah novel yang berjudul "5cm". Berikut adalah contoh perbedaan pola pada kedua novel yang berbeda.

"Surga Sungsang"
*SEJAK itu **Abu Jenar** memang tak pernah lagi menyambangi makam. Jika **ia** berkunjung ke kampung itu, paling-paling **ia** mengajak **Kiai Siti** dan penduduk kampung bercakap-cakap di masjid.*

Cuplikan novel di atas menceritakan dengan cara pemaparan, sehingga novel pola yang dihasilkan adalah entitas bernama kemudian kata ganti seseorang.

"5cm"
*"**Saya** cuma perantara. **Kamu** sendiri dengan izin dari yang Mahakuasa berhasil membawa diri **kamu** sendiri ke situ dan mengambil keputusan yang tepat." **Ian** tertunduk.*

Sedangkan pada cuplikan novel di atas memiliki pola yang berbeda dengan novel sebelumnya. Novel ini lebih banyak dialog sehingga pola yang dihasilkan adalah kata ganti kemudian entitas bernama. Hal ini akan mempengaruhi hasil klasifikasi karena perbedaan pola yang dihasilkan oleh ekstraksi fitur. Jumlah pasangan *coreference* dari kedua novel tersebut adalah 21736 pasangan. Sedangkan untuk data uji akan digunakan 5 novel yang mengandung kata benda berupa nama seseorang dan kata ganti, masing-masing novel akan diambil sekitar 100 hingga 200 kata.

Tabel 4. 6 Persebaran Data Klasifikasi *Coreference*

Data	Judul Novel	Jumlah Kata	Pasangan Coreference
Training	Surga Sungsang	1057	1535
	5 cm	708	1253
Testing	Garis Waktu	106	12
	Dilan	102	33
	Milea	112	24

	Selamat Datang di Rumah Mati	186	106
	Dear Nathan	106	27

4.2.4. Hasil Pengujian Akurasi

Pengujian nilai akurasi (2.29) yang dihasilkan, dengan menggunakan metode *Support Vector Machine* dapat dilihat pada Tabel 4.7.

Tabel 4. 7 Pengujian Akurasi

Judul Novel	**Jumlah (Pasangan)**	**Hasil Klasifikasi**		**Akurasi**
		Benar	**Salah**	
Garis Waktu	36	14	22	38.8%
Dilan	171	84	87	49%
Milea	136	62	74	45.6%
Selamat Datang di Rumah Mati	528	334	194	63.3%
Dear Nathan	105	57	48	54.29%

Berikut adalah perhitungan untuk akurasi keseluruhan pada klasifikasi novel dengan metode SVM.

$$Rata - rata\ Akurasi = \frac{Total\ akurasi\ setiap\ sampel\ data\ uji}{Jumlah\ sampel\ data\ uji}$$

$$= \frac{38.8\% + 49\% + 45.6\% + 63\% + 54.29\%}{5} = 50.14\%$$

Hasil pengujian detail perancangan merupakan proses dimana sampel data masukan berhasil melewati proses *preprocessing* (*case folding, filtering*, tokenisasi kalimat, tokenisasi kata, stemming dan POS *tagging*), ekstraksi fitur dan pengklasifikasian menggunakan metode *support vector machine* dengan jumlah data sebanyak 5 novel sebagai data uji.

Dari Tabel 4.7 analisis hasil pengujian *Coreference Resolution* menggunakan *Support Vector Machine* menghasilkan nilai rata-rata akurasi sebesar 54%. Berikut ini adalah penjelasan dari penyebab nilai akurasi *Coreference Resolution* yang dihasilkan sistem cukup kecil.

1. Pola data uji yang cukup jauh berbeda dengan data latih seperti novel yang berjudul "Surga Sungsang" memiliki pola mengulang entitas bernama, sedangkan novel yang berjudul "Garis Waktu" hanya membahas tentang kata ganti seseorang sehingga memiliki pola yang jauh berbeda dengan data latih sehingga menghasilkan nilai akurasi yang rendah.
2. Belum adanya pendeteksian mengenai kata ganti majemuk, sedangkan di dalam POS tag kata ganti tersebut masih terdeteksi sebagai PRP. Sehingga belum bisa dikatakan *coreference* dengan siapa pun.

Sehingga dapat diambil kesimpulan bahwa metode SVM ini sebenarnya cocok digunakan untuk *Coreference Resolution* pada novel berbahasa Indonesia, namun dengan syarat mencari data latih yang beragam polanya agar dapat mengklasifikasikan dengan baik dan menambahkan pendeteksian kata ganti majemuk.