

BAB 4

IMPLEMENTASI DAN PENGUJIAN

4.1 Perancangan Sistem

Tahap implementasi merupakan tahap pembangunan perangkat lunak Ekstraksi Informasi menggunakan CRF sesuai dengan hasil dari tahap analisis dan perancangan yang sudah dilakukan sebelumnya.

4.1.1 Implementasi Perangkat Keras

Perangkat keras yang digunakan untuk melakukan pengujian ditentukan berdasarkan percobaan sebelumnya terhadap beberapa perangkat keras yang sekiranya mampu untuk menangani proses ekstraksi informasi menggunakan CRF. Perangkat keras yang digunakan dapat dilihat pada tabel berikut.

Tabel 4. 1 Implementasi Perangkat Keras

No	Perangkat keras	Spesifikasi
1	<i>Processor</i>	Intel® Core (TM) i5-8250U CPU @ 1.60GHz
2	<i>Installed memory (RAM)</i>	4,0 GB

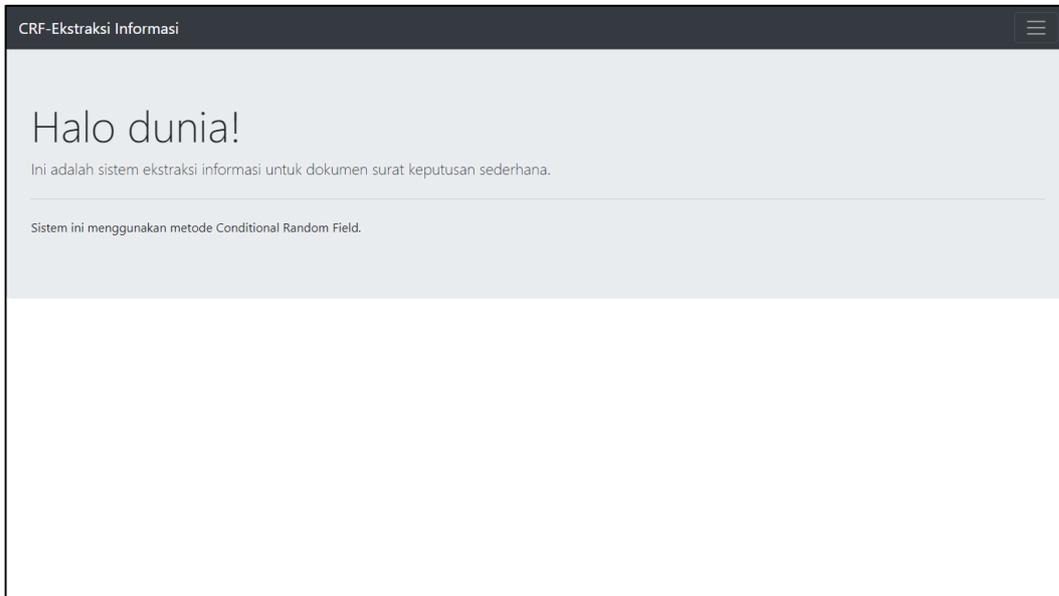
4.1.2 Implementasi Perangkat Lunak

Sistem ekstrakssi informasi menggunakan CRF dibangun pada platform web melalui beberapa perangkat lunak pilihan. Perangkat lunak yang digunakan dapat dilihat pada tabel berikut.

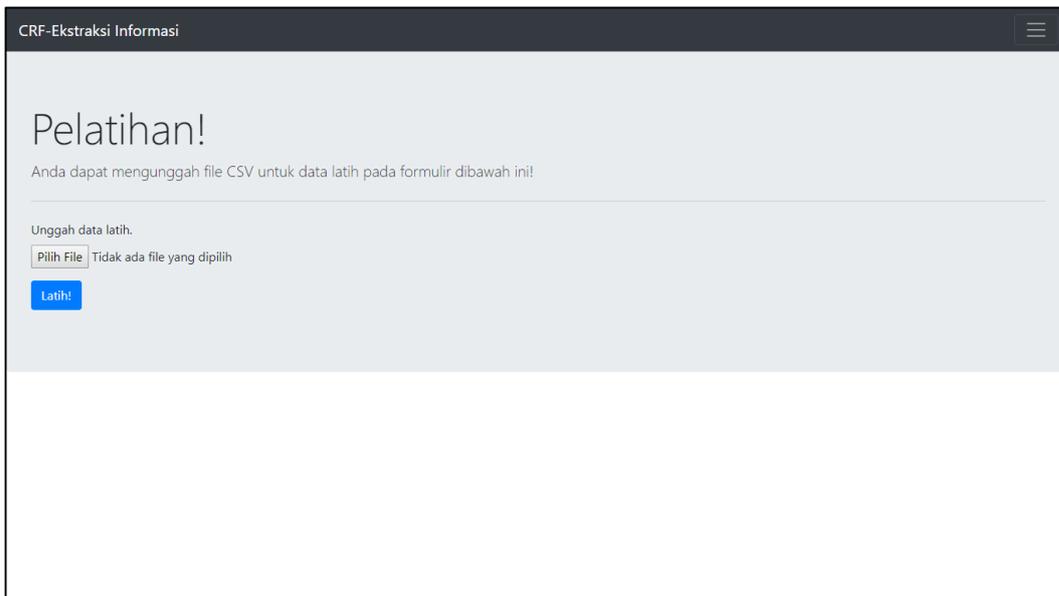
Tabel 4. 2 Implementasi Perangkat Lunak

No	Jenis Perangkat Lunak	Perangkat Lunak
1	Sistem operasi	Windows 10 Home
2	<i>Web browser</i>	Google Chrome
3	<i>Web server</i>	XAMPP
4	<i>Teks Editor</i>	Sublime Text
5	<i>Optical Character Recognition</i>	ABBY OCRsdk
6	Bahasa Pemrograman	<i>PHP 7.1.9</i>

4.1.3 Implementasi Antarmuka



Gambar 4. 1 Halaman Utama



Gambar 4. 2 Halaman Unggah Data Latih

CRF-Ekstraksi Informasi

Ekstraksi Fitur Node

Kata	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15
AKADEMISEKRETARI	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STATUS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SURAT	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Nomor	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Ekstraksi Fitur Edge

Kata	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	f19	f20
AKADEMISEKRETARI	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STATUS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SURAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Nomor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Node Potensial

Kata	f1	f2	f3	f4	f5	f6
AKADEMISEKRETARI	0.99936960579086	1	1	1	1	1
STATUS	0.99936960579086	1	1	1	1	1

Gambar 4. 3 Halaman Hasil Pelatihan

CRF-Ekstraksi Informasi

Tekan **F11** untuk keluar dari mode layar penuh

Pengujian!

Anda dapat mengunggah file CSV untuk data uji pada formulir dibawah ini!

Unggah data uji.

Tidak ada file yang dipilih

Gambar 4. 4 Halaman Unggah Data Uji

CRF-Ekstraksi Informasi															
Ekstraksi Fitur Node															
Kata	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15
SURAT	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Nomor	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Menimbang	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Madrasah	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Ekstraksi Fitur Edge																					
Kata	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	f19	f20	f21
SURAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Nomor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Menimbang	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Madrasah	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Node Potensial									
Kata	f1	f2	f3	f4	f5	f6	f7	f8	f9
SURAT	1	0.99943956410303	1		1	1	1		1
Nomor	1	1		0.99939359094622	1	1	1		1

Gambar 4. 5 Halaman Hasil Data Uji

4.2 Pengujian Sistem

Pengujian yang dilakukan sistem melalui 3 tahapan, yang pertama adalah membuat skenario pengujian, kemudian melakukan pengujian berdasarkan skenario yang telah dibuat. Pengujian dilakukan terhadap fungsionalitas sistem, nilai parameter, serta nilai akurasi.

4.2.1 Skenario Pengujian

Berikut ini merupakan skenario pengujian yang dilakukan pada sistem ekstraksi informasi dokumen surat keputusan menggunakan CRF.

4.2.1.1 Skenario Pengujian Fungsionalitas

Metode yang digunakan untuk membuat skenario pengujian fungsionalitas adalah metode *black box*. Metode ini berfokus pada nilai *ouput* yang diberikan oleh *input*, maupun terhadap proses yang terjadi. Berikut skenario pengujian fungsionalitas dapat dilihat pada tabel berikut.

Tabel 4. 3 Skenario Pengujian *Black Box*

Nama Skenario : Pengujian Praproses Data <i>Training</i>		
No	Nama Proses	Poin Pengujian
1	Pemisahan data	Memisahkan data dengan label data per baris.

	dan label kata	
2	Tokenisasi Kata	Memisahkan kata dengan spasi sebagai pemisah antar kata per baris data.
Nama Skenario : Pengujian Ekstraksi Fitur Data <i>Training</i>		
No	Nama Proses	Poin Pengujian
1	Ekstraksi fitur data <i>training</i>	Melakukan proses ekstraksi fitur <i>node</i> dan <i>edge</i> sesuai dengan label kelas kata yang dimiliki.
Nama Skenario : Pengujian Proses Testing		
No	Nama Proses	Poin Pengujian
1	Hitung <i>node</i> dan <i>edge</i> potensial	Menghitung <i>node</i> dan <i>edge</i> potensial.
2	Hitung <i>forward</i> dan <i>backward pass</i>	Menghitung <i>forward</i> dan <i>backward pass</i> .
3	Menghitung probabilitas <i>node</i> dan <i>edge</i>	Menghitung probabilitas <i>node</i> dan <i>edge</i> .
4	Hitung gradien	Menghitung gradien untuk fitur <i>node</i> dan <i>edge</i> .
Nama Skenario : Pengujian Praproses Data <i>Testing</i>		
No	Nama Proses	Poin Pengujian
1	Pemisahan data dan label kata	Memisahkan data dengan label data per baris.
2	Tokenisasi Kata	Memisahkan kata dengan spasi sebagai pemisah antar kata per baris data.
Nama Skenario : Pengujian Ekstraksi Fitur Data <i>Testing</i>		
No	Nama Proses	Poin Pengujian
1	Ekstraksi fitur data <i>testing</i>	Melakukan proses ekstraksi fitur <i>node</i> dan <i>edge</i> sesuai dengan data <i>training</i> .
Nama Skenario : Pengujian Proses <i>Testing</i>		
No	Nama Proses	Poin Pengujian
1	Hitung <i>node</i> dan <i>edge</i> potensial	Menghitung <i>node</i> dan <i>edge</i> potensial.
2	Hitung <i>maximal forward pass</i>	Menghitung menghitung <i>mmaximal forward pass</i> .
3	Hitung <i>backtracking</i>	Menghitung nilai <i>backtracking pass</i> .

4.2.1.2 Skenario Pengujian Nilai Parameter dan Akurasi

Pada tahap ini akan dilakukan pengujian pengaruh nilai parameter terhadap akurasi yang dihasilkan. Adapun nilai parameter yang diuji adalah inialisasi nilai awal parameter (λ), nilai standar deviasi (σ) serta nilai *learning rate* (ω). Untuk inialisasi nilai awal nilai yang akan diuji adalah 0 dan 1, nilai

standar deviasi rentang nilainya adalah antara 2 sampai 8, sedangkan nilai *learning rate* 0,000001.

Pengujian akurasi dilakukan untuk mengetahui akurasi yang dihasilkan sistem ekstraksi informasi menggunakan CRF. Pengujiannya dilakukan terhadap hasil dari proses *training* yang menggunakan data *training* sebanyak 20 dokumen surat keputusan. Adapun untuk pengujiannya dilakukan menggunakan 6 dokumen surat keputusan.

Adapun metode perhitungan akurasi yang akan dilakukan yaitu dengan metode *Confusion Matrix*. *Confusion Matrix* merupakan suatu matriks yang berisi data aktual dan prediksi yang diklasifikasikan oleh sistem. Adapun ketentuan *confusion matrix* dapat dilihat pada tabel berikut.

Tabel 4. 4 *Confusion Matrix*

<i>True Label</i>	<i>Predicted Label</i>	
	0	1
0	TP	FN
1	FP	TN

Adapun persamaan untuk perhitungan akurasi pada metode *Confusion Matrix* adalah sebagai berikut.

$$\text{Akurasi} = \frac{\sum_{i=1}^L \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}}{L} \times 100\%$$

Keterangan:

TP : *True Positive* ke-i

TN : *True Negative* ke-i

FP : *False Positive* ke-i

FN : *False Negative* ke-i

L : Jumlah Keseluruhan Data Yang Diuji

4.2.2 Pengujian

Pengujian dilakukan terhadap fungsionalitas, nilai parameter, serta nilai akurasi pada sistem ekstraksi informasi dokumen surat keputusan menggunakan CRF.

4.2.2.1 Pengujian Fungsionalitas

Adapun untuk hasil pengujian *black box* untuk sistem ekstraksi informasi dokumen surat keputusan menggunakan CRF adalah sebagai berikut.

Tabel 4. 5 Hasil Pengujian *Black Box*

Nama Proses	Data Masukan	Keluaran yang Diharapkan	Pengamatan	Kesimpulan
Pemisahan data dan label kelas <i>data training</i>	File CSV <i>data training</i>	Data dan label kelas data dipisahkan	Menampilkan data dan label kelas data yang sudah dipisahkan	[<input checked="" type="checkbox"/>] Diterima [<input type="checkbox"/>] Ditolak
Tokenisasi Kata <i>data training</i> .	Hasil pemisahan data dengan label <i>data training</i>	Teks data <i>training</i> menjadi per kata dengan pemisah spasi.	Menampilkan teks data <i>training</i> per kata	[<input checked="" type="checkbox"/>] Diterima [<input type="checkbox"/>] Ditolak
Ekstraksi fitur <i>data training</i>	Hasil tokenisasi kata <i>data training</i>	Fitur <i>node</i> dan <i>edge</i> sesuai dengan label kelas yang dimiliki kata pada data <i>training</i>	Menampilkan hasil ekstraksi fitur yang sesuai dengan label kelas yang dimiliki pada data <i>training</i> .	[<input checked="" type="checkbox"/>] Diterima [<input type="checkbox"/>] Ditolak
Hitung <i>node</i> dan <i>edge</i> potensial	Hasil ekstraksi fitur data <i>training</i> , nilai parameter	Nilai <i>node</i> dan <i>edge</i> potensial yang benar, dan sesuai dengan hasil ekstraksi fitur	Menampilkan Nilai <i>node</i> dan <i>edge</i> potensial yang benar, dan sesuai dengan hasil ekstraksi fitur	[<input checked="" type="checkbox"/>] Diterima [<input type="checkbox"/>] Ditolak
Nama Proses	Data	Keluaran yang	Pengamatan	Kesimpulan

	Masukan	Diharapkan		
Hitung <i>forward</i> dan <i>backward pass</i>	<i>Node</i> dan <i>edge</i> potensial	Nilai <i>forward</i> dan <i>backward pass</i> yang benar, yang mengacu pada nilai <i>node</i> dan <i>edge</i> potensial yang tepat	Menampilkan Nilai <i>forward</i> dan <i>backward pass</i> yang benar, yang mengacu pada nilai <i>node</i> dan <i>edge</i> potensial yang tepat	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Ditolak
Menghitung probabilitas <i>node</i> dan <i>edge</i>	<i>Node</i> dan <i>edge</i> potensial, Nilai <i>forward</i> dan <i>backward pass</i> .	Nilai probabilitas <i>node</i> dan <i>edge</i> yang benar serta mengacu pada nilai <i>node</i> dan <i>edge</i> serta nilai <i>forward</i> dan <i>backward pass</i> yang tepat	Menampilkan Nilai probabilitas <i>node</i> dan <i>edge</i> yang benar serta mengacu pada nilai <i>node</i> dan <i>edge</i> serta nilai <i>forward</i> dan <i>backward pass</i> yang tepat	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Ditolak
Hitung gradien	Hasil ekstraksi fitur, probabilitas <i>node</i> dan <i>edge</i>	Nilai gradien dan parameter yang benar, berdasarkan hasil ekstraksi fitur serta nilai probabilitas <i>node</i> dan <i>edge</i> yang tepat	Menampilkan Nilai gradien dan parameter yang benar, berdasarkan hasil ekstraksi fitur serta nilai probabilitas <i>node</i> dan <i>edge</i> yang tepat	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Ditolak
Pemisahan data dan label kelas <i>data testing</i>	File CSV data <i>testing</i>	Data dan label kelas data dipisahkan	Menampilkan data dan label kelas data yang sudah dipisahkan	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Ditolak
Tokenisasi Kata <i>data testing</i> .	Hasil pemisahan data dengan label <i>data testing</i>	Teks data <i>testing</i> menjadi per kata dengan pemisah spasi.	Menampilkan teks data <i>testing</i> per kata	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Ditolak

Nama Proses	Data Masukan	Keluaran yang Diharapkan	Pengamatan	Kesimpulan
Ekstraksi fitur data <i>testing</i>	Data <i>training</i>	Fitur <i>node</i> dan <i>edge</i> sesuai dengan label kelas kata yang dimiliki kata pada data <i>training</i>	Menampilkan hasil ekstraksi fitur yang sesuai dengan label kelas kata yang sesuai pada data <i>training</i> .	[√] Diterima [] Ditolak
Hitung <i>node</i> dan <i>edge</i> potensial	Hasil ekstraksi fitur data <i>testing</i> , nilai parameter	Nilai <i>node</i> dan <i>edge</i> potensial yang benar, dan sesuai dengan hasil ekstraksi fitur	Menampilkan Nilai <i>node</i> dan <i>edge</i> potensial yang benar, dan sesuai dengan hasil ekstraksi fitur	[√] Diterima [] Ditolak
Hitung <i>maximal forward pass</i>	<i>Node</i> dan <i>edge</i> potensial	Nilai <i>maximal forward pass</i> yang benar dan mengacu pada nilai <i>node</i> dan <i>edge</i> potensial yang tepat	Menampilkan Nilai <i>maximal forward pass</i> yang benar dan mengacu pada nilai <i>node</i> dan <i>edge</i> potensial yang tepat	[√] Diterima [] Ditolak
Hitung <i>backtracking pass</i>	<i>Node</i> potensial, nilai <i>maximal forward pass</i>	Nilai <i>backtracking pass</i> yang benar, dan mengacu pada nilai <i>node</i> potensial dan <i>maximal forward pass</i> yang tepat	Menampilkan Nilai <i>backtracking pass</i> yang benar, dan mengacu pada nilai <i>node</i> potensial dan <i>maximal forward pass</i> yang tepat	[√] Diterima [] Ditolak

4.2.2.2 Pengujian Nilai Parameter dan Akurasi

Hasil dari proses *training* dengan beberapa kombinasi nilai parameter akan diuji dengan data *testing* yang sama untuk menentukan kombinasi nilai terbaik. Berikut ini merupakan hasil pengujian nilai parameter untuk inisialisasi nilai awal parameter fitur (λ), nilai standar deviasi (σ) serta nilai *learning rate* (ω).

1. Pengujian Akurasi Pertama

Pada pengujian pertama dilakukan pelatihan dengan menggunakan 20 data latih dokumen surat keputusan. Data uji yang digunakan berjumlah 6 data. Nilai parameter yang diuji dengan nilai awal parameter fitur (λ) = 0, nilai standar deviasi (σ) = 2, serta nilai *learning rate* (ω) = 0,00001. Adapun *Confusion Matrix* yang dihasilkan pada pengujian akurasi pertama dapat dilihat pada tabel berikut.

Tabel 4. 6 Pengujian Akurasi Pertama

1573	KS	JS	NS	TTG	YMS	MG	MT	MS	NMD	NID	TP	TG	ORG	NYM	TMB
KS	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
JS	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NS	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TTG	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YMS	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MG	0	0	0	0	0	341	0	60	0	0	0	0	0	0	0
MT	0	0	0	0	0	0	265	60	0	0	0	0	0	0	0
MS	0	0	0	0	0	0	0	536	0	0	0	0	0	0	0
NMD	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NID	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TP	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TG	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ORG	5	0	0	0	0	0	0	10	0	0	0	0	0	0	0
NYM	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TMB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82

Dari hasil *confusion matrix* diatas, dapat diukur tingkat akurasi dari prediksi dengan menggunakan persamaan berikut.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\sum_{i=1}^L \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}}{L} \times 100\% \\
 &= \frac{0+0+0+0+0+341+265+536+0+0+0+0+0+0+0+82}{1573} \times 100\% \\
 &= \frac{1224}{1573} \times 100\% = 77,81\%
 \end{aligned}$$

2. Pengujian Akurasi Kedua

Pada pengujian kedua dilakukan pelatihan dengan menggunakan 20 data latih dokumen surat keputusan. Data uji yang digunakan berjumlah 6 data. Nilai parameter yang diuji dengan nilai awal parameter fitur (λ) = 1, nilai standar deviasi (σ) = 2, serta nilai *learning rate* (ω) = 0,00001. Adapun *Confusion Matrix* yang dihasilkan pada pengujian akurasi kedua dapat dilihat pada tabel berikut.

Tabel 4. 7 Pengujian Akurasi Kedua

1573	KS	JS	NS	TTG	YMS	MG	MT	MS	NMD	NID	TP	TG	ORG	NYM	TMB
KS	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JS	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0
NS	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0
TTG	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0
YMS	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0
MG	0	0	0	0	0	341	0	60	0	0	0	0	0	0	0
MT	0	0	0	0	0	0	265	60	0	0	0	0	0	0	0
MS	0	0	0	0	0	0	0	536	0	0	0	0	0	0	0
NMD	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0
NID	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
TP	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0
TG	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0
ORG	0	0	0	0	0	0	0	10	0	0	0	0	5	0	0
NYM	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0
TMB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82

Dari hasil *confusion matrix* diatas, dapat diukur tingkat akurasi dari prediksi dengan menggunakan persamaan berikut.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\sum_{i=1}^L \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}}{L} \times 100\% \\
 &= \frac{9+27+19+19+3+341+265+536+38+3+19+31+5+26+82}{1573} \times 100\% \\
 &= \frac{1443}{1573} \times 100\% = 91,73\%
 \end{aligned}$$

4.2.3 Pembahasan Hasil Pengujian

Dari hasil pengujian terhadap 6 data uji yang telah dilakukan, analisa hasil akurasi yang didapatkan terhadap ekstraksi informasi dokumen surat keputusan adalah sebagai berikut.

1. Akurasi dengan persentase tertinggi didapat pada pengujian kedua yaitu sebesar 91.73% dengan menggunakan nilai parameter awal = 1.
2. Akurasi dengan persentase terendah didapat pada pengujian pertama yaitu sebesar 77.81% dengan menggunakan nilai parameter awal = 0.

Berdasarkan hasil diatas, nilai parameter awal pada proses pelatihan berpengaruh terhadap tingkat akurasi yang didapat. Kesalahan dalam klasifikasi data juga disebabkan oleh kurang-nya fitur yang akurat dalam penentuan *node* awal pada proses pengujian. Hal ini dapat dilihat dari banyaknya kesalahan klasifikasi pada kata yang muncul di banyak bagian terutama pada kelas Mengingat, Menimbang dan Memutuskan.

Pada kelas Mengingat, Menimbang, dan Memutuskan, sering ditemukan kata yang sama pada awal baris seperti “Pertama”, “Kedua”, “a.”, “b.” dan sebagainya. Hal ini berpengaruh pada isi dari kamus kata pertama yang terbentuk, kamus akan memiliki isi yang sama pada label yang berbeda.

Selain itu, hasil dari konversi gambar menuju teks dengan menggunakan OCR juga berpengaruh dalam akurasi yang didapat. Kesalahan seperti “l” menjadi “1” atau “1” akan berpengaruh ketika pencocokan kata dengan kamus kata