

PART OF SPEECH TAGGING MENGGUNAKAN BAHASA JAWA DENGAN METODE CONDITION RANDOM FIELDS

Atik Zilziana¹, Arie Ardiyanti Suryani², Ibnu Asror³

Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹atikzilziana@students.telkomuniversity.ac.id, ²ardiyant@telkomuniversity.ac.id @telkomuniversity.ac.id,

³iasror@telkomuniversity.ac.id@telkomuniversity.ac.id

Abstrak—Part of Speech Tagging (POS Tagging) adalah proses memberi label pada setiap kata dalam kalimat dengan POS atau tag yang sesuai dengan kelas kata seperti kata kerja, kata benda kata keterangan, kata sifat, dan lainnya. Bahasa yang digunakan ialah Bahasa Jawa. Bahasa Jawa adalah salah satu Bahasa daerah yang digunakan sebagian besar di pulau Jawa Indonesia. Digunakan di provinsi Jawa Tengah, DIY, Jawa Timur dan sekitarnya. Menggunakan data dari berita online berbahasa Jawa Krama. Metode yang digunakan yaitu CRF (*Conditional Random Fields*). CRF adalah suatu model probabilistik yang banyak digunakan pada proses segmentasi dan pelabelan suatu sekuen data. Pada penelitian ini menggunakan data corpus sebanyak 3000 kata dan menggunakan cross validasi untuk proses training. Pada POS Tagging Bahasa Jawa ini mendapatkan akurasi sebesar 67%.

Keywords—Part of Speech, Conditional Random Fields, Javanese

Abstract

Javanese is the one of regional languages that is used mostly on the Indonesian island of Java. Used in the provinces of Central Java, DIY, East Java, and surrounding areas. Using data from online news in the Javanese language. The method used is CRF (*Conditional Random Fields*). CRF is a structure for building probabilistic models. This paper describes the part of speech for Javanese using Conditional Random Fields (CRF) method. We used a corpus of size 3000 words and cross-validation for training. Our model gave an overall accuracy of 67%.

Keywords—Part of Speech, Conditional Random Fields, Bahasa Jawa

I. Pendahuluan

Part-of-speech tagging (POS), juga disebut gramatikal tagging adalah proses penugasan bagian dari tag ucapan untuk kata-kata dalam teks. Part-of-speech adalah kategori tata bahasa, umumnya termasuk kata kerja, kata benda, kata sifat, kata keterangan, dan sebagainya. Penandaan sebagian ucapan merupakan alat penting dalam banyak aplikasi pemrosesan bahasa alami seperti disambiguasi kata, penguraian, penjawaban pertanyaan, dan terjemahan mesin. *Part of Speech tagging* (POS tagging) memiliki peran penting dalam berbagai bidang pemrosesan bahasa alami (NLP) termasuk terjemahan mesin untuk memproses sentimen analisis. Tagset umum yang terdapat pada *Penn Treebank POS Tag* memiliki 37 tag[1]. POS Tagger ini merupakan tools yang penting dalam pemrosesan Bahasa sehingga perlu untuk dikembangkan dengan akurasi yang baik.

Indonesia adalah negara multibahasa yang luas dengan beragam budaya. Ini memiliki banyak bahasa dengan bentuk tertulis dan bahasa yang digunakan di masing-masing daerah. Dan bahasa Jawa adalah salah satu bahasa daerah di Indonesia.

Penelitian bahasa Jawa ini menggunakan dataset dari media berita online. Karakteristik berita online terletak pada panjangnya kalimat. Penelitian POS tagging menggunakan bahasa Jawa juga sudah ada sebelumnya tetapi menggunakan metode basis aturan (*Rule base*).

Dalam penelitian ini mengembangkan POS Tagging menggunakan *Conditional Random Fields* (CRF). CRF telah diterapkan pada sejumlah besar domain lain, termasuk pemrosesan teks, visi komputer, dan bioinformatika. Sejak itu, *linear-chain* CRF telah diterapkan pada banyak masalah dalam pemrosesan bahasa alami, termasuk pengenalan entitas bernama, fitur induksi untuk NER, dekomposisi dangkal, segmentasi alamat pada halaman Web [2].

Studi sebelumnya telah mengevaluasi metode CRF menggunakan tweet Indonesia di Twitter dengan hasil akurasi yang cukup bagus yaitu 71%[3]. Namun, dalam penelitian itu teks yang digunakan menggunakan teks tweet. Dalam penelitian ini, teknik yang sama untuk POS Tagger dalam bahasa Jawa akan digunakan dengan harapan bahwa hasilnya juga akan menghasilkan akurasi yang relatif baik. Adapun penelitian bahasa Jawa ini menggunakan tag set pada media berita

online. Karakteristik media berita online sangat berbeda dibandingkan dengan menggunakan tweet. Perbedaannya terletak pada panjang kalimat, di mana teks dalam tweet relatif lebih pendek atau lebih pendek. Masalah utama dari penelitian ini bagaimana metode CRF dapat memberikan akurasi paling baik untuk *POS Tag* Bahasa Jawa.

II. Literature Survey

Pada penelitian "Pengaruh Bagian Berdasarkan Aturan Tagging Pidato dan Distribusi Probabilitas Entropi Maksimum untuk Krama Jawa". Menguji *POS Tagging* berdasarkan Basis Aturan dan Distribusi Probabilitas Entropi Maksimum di Krama hasilnya adalah akurasi tertinggi 97,67% [2]. Penelitian dengan Bahasa Jawa lainnya yaitu "Part of Speech Tagging for Javanese Ngoko Language with Hidden Markov Model" dengan mendapat akurasi 92.6% [15].

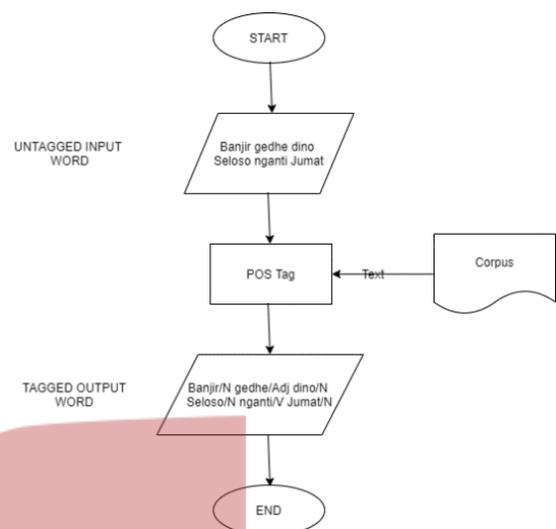
Penelitian selanjutnya "Perbandingan Algoritma Medan Acak Bersyarat dan Model Markov Tersembunyi dalam Posting Tagging Indonesia". Pengumpulan data dilakukan dengan mengumpulkan tweet pada tag #UNBK satu per satu. Penelitian ini menggunakan 10 tweet dan algoritma Hidden Markov Model mendapatkan tingkat akurasi 62,7% sedangkan algoritma Conditional Random Field mendapatkan 71% [3].

CRF dengan bahasa kannada menggunakan corpus dengan ukuran 3000 kalimat. Model ini memberikan akurasi keseluruhan 96,86% menggunakan tag set *hierarkis* yang terdiri dari 19 tag [4]. Penelitian lain dengan Conditional Random Fields (CRF). Hasil percobaan untuk teks Filipina dengan CRF menunjukkan akurasi 90,59% [5]. Penelitian dengan bahasa non-Inggris mendapat akurasi yang baik. Pada bahasa-bahasa seperti Telugu dan Hindi dengan akurasi masing-masing 78,66% dan 77,37% [7]. Dan POS Tagging dengan Bahasa Inggris mendapat akurasi 97% [6]. Tagging dan kata-kata POS adalah fitur dasar dari model CRF, kombinasi dari dua fitur di atas telah meningkatkan akurasi sistem untuk penandaan POS adalah 77,37% dan 79,15% (data uji Telugu) untuk chunking [7].

III. Metode

Sistem yang dibangun dalam penelitian ini adalah POS Tagging untuk Bahasa Jawa dapat dilihat pada gambar 1. Proses atau cara *kerja POS Tagging* itu sendiri. Setiap kata input menghasilkan

label untuk setiap kata yang di inputkan menggunakan metode Conditional Random Filed.



Gambar 1 Proses POS Tagging

1. Regi/ N Lombok/N ing/Prp Tlatah/N Klaten/N ./, Jawa/N Tengah /N ./, Ngantos/V samangke/Knj taksih/ADV awis/ADJ ./, Regi/N Lombok/N rawit/Adj abrit/ADJ dereng /ADV mandhap/V ./, kepara /ADJ lon-lonan/ADJ mindhak/V malih/V Miturut/ADV Para/Knj bakul/N lombok/N ing/Prp Peken/N Klaten/N Kota/N ./, Regi/N lombok/N samangke/KNJ mindhak/V sekedhik/ADJ ./.
2. Jawah/N dipunmaweni/ADV angin/N kencang/ADJ ingkang/KNJ nempuh/V hampir/ADV sedaya/NUM kabupaten/N kitha/N ing/KNJ DIY/N Ngakibataken/V macem-macem/N bencana/N ./, Ing/Knj antawisipuN/Knj yaiku/Knj wit/N ambruk/V ingkang/KNJ kathah/ADJ kedadosan/N ing/KNJ kabupaten/N sleman/N ./, Dina/N Kamis/N sonten/N (/Kurung Buka 5/NUM -/Tanda pisah1/NUM -/Tanda pisah 2017/NUM)/Kurung Tutup ./.

Berdasarkan Table I, dari pengamatan data tagset Bahasa Indonesia definisikan dengan 37 tag untuk kata-kata bahasa Indonesia termasuk tanda baca[10].

Table 1 TAGSET BAHASA INDONESIA [10]

No	Tag	Desripsi	Example
1	(Opening Parenthesis	(([
2)	Closing Parenthesis)]]
3	,	Coma	,
4	.	Sentence Terminator	. ? !
5	:	Colon or Ellipsis	:
6	-	Dash	-
7	“	Opening Quotation Mark	“ “
8	”	Closing Quotation Mar	” ”
9	\$	Dollar	\$
10	Rp	Rupiah	Rp
11	SYM	Symbols	\$ % & @
12	NNC	Countable Common Nouns	Buku, Rumah, Karyawan
13	NNU	Uncountable Common Nouns	Air, Gula, nasi, Hujan
14	NNG	Genitive Common Nouns	Idealnya, komposisinya, fungsinya, jayanya
15	NNP	Proper Nouns	Jakarta, Australia, Soekarno-Hatta
16	PRP	Personal Pronouns	Saya, aku, dia, kami
17	PRN	Number Pronouns	Kedua-duanya, Ketigatiganya
18	PRL	Locative Pronouns	Sini, Situ, Sana
19	WP	WH-Pronouns	Apa, Siapa, Mengapa, Bagaimana
20	VBT	Transitive Verbs	Berbicara, Mengangkat, Menyanyi
21	VBI	Intransitive Verbs	Bermain, terdiam, berputar-putar
22	MD	Modal or Auxiliaries Verb	Sudah, boleh, harus, mesti

23	JJ	Adjectives	Mahal, kaya, besar, malas
24	CDP	Primary Cardinal Numerals	Satu, juta, milyar
25	CD O	Ordinal Cardinal Numerals	Pertama, kedua, ketiga
26	CDI	Irregular Cardinal Numerals	Beberapa, segala, semua
27	CDC	Collective Cardinal Numerals	Bertiga, bertujuh, berempat
28	NEG	Negations	Bukan, tidak, belum, jangan
29	IN	Prepositions	Di, ke, dari, pada, dengan
30	CC	Coordinate Conjunction	Dan, atau, tetapi
31	SC	Subordinate Conjunction	Yang, Ketika, Setelah
32	RB	Adverbs	Sekarang, nanti, sementara, sebab, sehingga
33	UH	Interjections	Wah, aduh, astaga, oh
34	DT	Determiners	Para, ini, masing masing, itu
35	WD T	WH Determiners	Apa, siapa, barangsiapa
36	RP	Particles	Kan, kah, lah, pun
37	FW	Foreign Word	Semua kata kecuali Bahasa

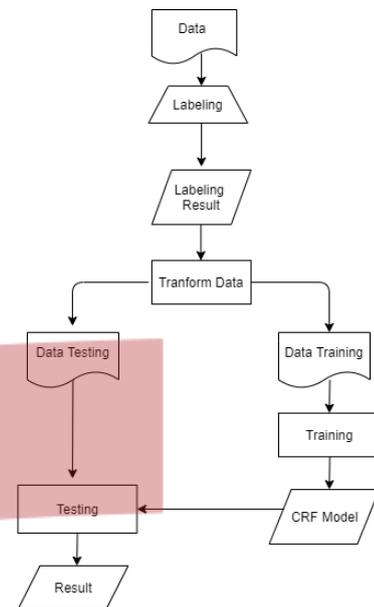
Pada table II menjelaskan label atau kelas kata yang akan digunakan pada penelitian ini. Dengan label Bahasa Jawa yang berjumlah 19 label [2]. Modifikasi diberikan pada pelabelan tertentu seperti pada pelabelan \$ dan Rp yang dimasukkan dalam symbol dan pelabelan tanda baca seperti yang ditunjukkan pada Tabel nomor 2 sampai dengan 19. Tagset dalam penelitian ini mengikuti referensi dari Buku Kosakata Bahasa Jawa yang memiliki 10 kategori tagset untuk bahasa Jawa [13].

TABLE II TAGSET BAHASA JAWA [3]

NO	Tag	Main Tag	Example
1	N	Noun	Bayu, Wedhus
2	V	Verb	Tindak, Tumbas
3	Adj	Adjective	Apik, Ayu, Bagus
4	Adv	Adverb	Mangkih, kadung, bablas
5	Knj	Conjunction	Lamun, wiwt
6	Prp	Preposition	Marang
7	Kh	Pengkhusus	Banget
8	So	Subordinator	Nalika
9	Em	Emotif	Eh, Aduh
10	Pr	Pronominal	Niki, sampean, kuwi, kulo
11	Sym	Symbol	Rp \$ % & @
12	{ { [Opening parenthesis	{ { [
13) }]	Closing parenthesis) }]
14	,	Comma	,
15	“ “	Quotation	“ “
16	.,?;!	Sentence termonator	. ? !
17	--,-	Dash	- --
18	:	Colon	:
19	;	Semicolon	;

Gambar 2 adalah gambaran umum dari proses *POS Tagging* dalam sistem dengan metode *Conditional Random Fields*. Dataset tersebut berasal dari berita Bahasa Jawa online. Dengan berbagai berita, seperti berita kecelakaan, bencana alam, ekonomi, dan kesehatan. Dengan total dataset yaitu sekitar 3000 kata. Namun pada

dataset ini tidak semua menggunakan tagset Bahasa Jawa. Sekitar 5% merupakan Bahasa Indonesia yang harus diberi label menggunakan tagset Bahasa Indonesia



Gambar 2 Perancangan Sistem

A. Pelabelan

Dari data yang sudah terkumpul pada dataset atau corpus maka akan diberi label manual oleh beberapa orang yang asli Jawa. Proses pemberian label sendiri disebut labeling. Tagset sendiri mengacu pada tagset Bahasa Jawa yang terdapat 19 tagset.

B. Transform Data

Setelah proses pelabelan maka sistem akan melakukan proses transform data atau pemisahan antara kata dan label. Proses Conditional Random Fields merupakan model probabilitas untuk menghitung $P(y|x)$.

TABEL III CONTOH TRANFORM DATA

Kata	Label
Kanggo	PRP
sangu	N
sapa	PR
ngerti	V
mengko	N
yen	KNJ
wis	ADV
lulus	V

ora	ADV
mung	ADJ
dadi	V
guru	N

C. Conditional Random Fields

CRF adalah struktur untuk membangun model probabilistik untuk segmentasi sekuensial dan data pelabelan [9]. Conditional Random Fields dapat dimodelkan dalam bentuk gambar seperti pada Gambar 1. Dengan gambar terstruktur paling sederhana dan paling umum di NLP, digunakan untuk memodelkan urutan adalah gambar di mana node yang sesuai dengan elemen Y membentuk rantai yang dapat dipilih, seperti yang diilustrasikan pada Gambar 1 [10].

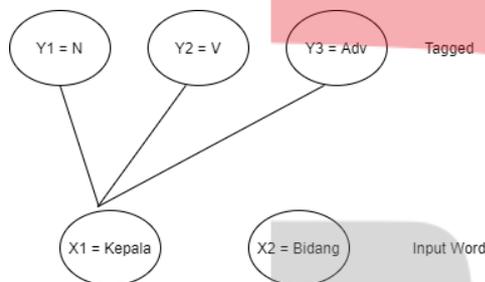


Figure 1 CRF model

Conditional Random Field (CRF) adalah model probabilistik untuk menghitung probabilitas $p(y | x)$ dari output yang memungkinkan $y = (y_1, y_2, \dots, y_n)$ ke input yang diberikan $x = (x_1, x_2, \dots, x_n)$ yang disebut juga data observasi [10]. Di mana Y adalah variabel acak untuk urutan label yang sesuai dan X adalah variabel acak untuk input yang akan diberi label. Probabilitas kondisional dari y ke x dapat ditulis dalam persamaan:

$$P(y|x) = \frac{1}{Z(x)} \prod_c \psi_c(y_c, x) \quad (1)$$

$$Z(x) = \sum_y \prod_c \psi_c(y_c, x) \quad (2)$$

Di mana $\psi_c(y_c, x)$ adalah fungsi positif atau fungsi potensial, dan $Z(x)$ adalah fungsi partisi atau fungsi normal.

Dalam penelitian ini, fitur simpul adalah hubungan antara label y dan data input x . Sedangkan fitur edge adalah hubungan antara label kelas kata untuk kata yang sedang diproses dan label kelas kata untuk kata berikutnya. Berikut ini adalah contoh fitur simpul, yaitu:

$$f_t(y_t, x, t) \begin{cases} 1 & \text{if } y_t = CC, x = \text{Saya} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Fitur di atas adalah fitur-t yang memiliki nilai 1 jika label CC dan kata yang sedang diproses adalah saya, jika tidak maka 0.

D. CRF steps

Secara umum, langkah-langkah yang dilakukan dalam metode *Conditional Random Field* (CRF) adalah sebagai berikut [12]:

1. Training

Tahap ini melibatkan mempelajari probabilitas bersyarat untuk memprediksi tag token input. Proses training menggunakan corpus yang ditandai atau sudah diberi label untuk pembelajaran ini. Tag corpus beranotasi diambil dari bahasa Jawa yang terdiri dari 19 tag. Korpus kami berisi 3000 kata oleh berita online Jawa. Dan juga menggunakan *Cross Validation* untuk training. *Cross Validation* (CV) adalah salah satu teknik yang digunakan untuk menguji efektivitas model pembelajaran mesin. Proses CV sendiri akan menghasilkan parameter terbaik yang nantinya akan diujikan untuk proses testing.

2. Testing

Setelah proses training dan mendapatkan hasil terbaik untuk parameter, proses testing sendiri akan melakukan uji coba untuk parameter tersebut. Untuk mendapatkan akurasi terbaik..

IV. Pengujian dan Hasil

Pada penelitian ini data corpus yang digunakan sebanyak 3000 kata. Untuk menentukan besaran data yang digunakan pada proses Training dan Testing bisa di lihat pada Tabel IV. Mencoba beberapa pembagian data agar mendapatkan pembagian yang terbaik. Dan mendapatkan hasil 63% dari pembagian 80:20 untuk data train dan data test.

TABEL IV VALIDASI DATA

Train	Testing	Accuracy
80	20	63%
75	25	62%
70	30	62%
65	35	62%
60	40	61%

Learning rate	Strandar Devition	Accuracy %
0.001	0.0001	67
	0.001	66
	0.01	66
0.01	0.0001	65
	0.001	66,2
	0.01	66
0.1	0.0001	65
	0.001	65,6
	0.01	65
3	0.0001	63
	0.001	63,4
	0.01	63,4
6	0.0001	61,5
	0.001	61,5
	0.01	61

Setelah proses validasi data, kami menggunakan *Cross Validation* (CV) untuk mengevaluasi kinerja algoritma CRF. Dengan menggunakan CV 10 fold yang akan diuji pada data train yaitu 80%. 10 fold adalah salah satu yang direkomendasikan agar setiap lipatan atau fold memiliki ukuran yang sama. *Cross Validation* (CV) digunakan untuk pemilihan model terbaik karena cenderung memberikan perkiraan akurasi yang kurang bias daripada yang biasanya. Dari data yang telah dibagi menjadi 10 fold, satu bagian akan digunakan sebagai data uji dan 9 lainnya untuk pelatihan. Ini dilakukan hingga fold terakhir.

Pada tahap ini nilai parameter akan diuji untuk pengujian yang dihasilkan. Nilai parameter yang dikeluarkan adalah nilai standar deviasi dan nilai learning rate. Untuk inisialisasi learning rate, kisaran nilai dari 0,001 sampai 9, sedangkan nilai standar deviasi adalah dari 0,0001 dan 0,01. Testing data yang digunakan berisi 2.400 kata dari corpus, sedangkan data pengujian yang digunakan berisi 600 kata. Dapat dilihat pada tabel II untuk hasil uji parameter.

Berikut adalah hasil tes dapat dilihat pada tabel V. Nilai parameter yang digunakan dalam

pelatihan adalah nilai yang menghasilkan akurasi terbaik dari proses pengujian nilai parameter. Untuk nilai inisialisasi dari standar deviasi adalah 0,001 dan learning rate adalah 0,01. Dan mendapatkan akurasi 67%.

TABLE V ACCURACY TESTING RESULTS

Tabel VI menunjukkan frekuensi tag pada *POS Tagging* dalam hasil testing dan tag yang memiliki F-Score rendah. Dapat dilihat dari tabel bahwa tag N untuk kata benda memiliki jumlah tes corpus tertinggi dengan frekuensi 253 dan persentase F-Score 77% yang artinya memiliki jumlah kata dan tag terbanyak dalam data test. Selanjutnya diikuti oleh V atau Verb yang untuk kata kerja yang benar memiliki frekuensi 71 tetapi F-Score hanya 42% dikarenakan masih ada beberapa kali salah dalam pelabelan. Dan beberapa tag yang F-Score-nya hanya nol seperti SO dan “ karena frekuensinya hanya 1.

TABLE VI POS TAG RESULT

Tagset	Precision	Recall	F-Score	Support
.,?!	0.972	1.000	0.986	35
N	0.669	0.909	0.771	253
KNJ	0.500	0.245	0.329	53
NUM	1.000	0.793	0.885	29
PRP	0.464	0.488	0.476	41
V	0.444	0.394	0.418	71
,	0.964	1.000	0.982	27
KH	1.000	0.875	0.933	8
-	1.000	1.000	1.000	1
ADJ	1.000	0.029	0.565	35
ADV	0.500	0.208	0.594	24
{{	1.000	1.000	1.000	5
}}	1.000	1.000	1.000	6
PR	0.273	0.214	0.640	14
SYM	1.000	0.667	0.800	3
SO	0.000	0.000	0.000	1

:	1.000	1.000	1.000	1
“	0.000	0.000	0.000	0

Hasil percobaan diatas yaitu terdapat beberapa kesalahan tag kelas kata. Oleh karena itu hasil F-Score beberapa tag cukup rendah. Nilai F-score cukup rendah dan kemunculan kesalahan dalam tag cukup sering ialah PRP, V, KNJ. Juga, ini mungkin karena urutan kata-kata di mana sebagian besar kata-kata V adalah kata kerja dan karena CRF adalah model probabilistik, ia melihat bahwa setelah kata tertentu digantikan oleh V, KNJ atau PRP.

TABLE VII HASIL INPUT KALIMAT

Input Kata	Tag yang Benar
[('Sawilangan', 'N'), ('wit', 'N'), ('ambruk', 'V'), ('dipuntempuh', 'N'), ('angin.', 'V'), ('Setunggal', 'N'), ('dipunantawisipun', 'N'), ('kedadosan', 'N'), ('ing', 'PRP'), ('dusun', 'N'), ('bangunrejo', 'N'), ('tridadi', 'N'), ('Sleman.', 'V')]	[('ambruk', 'N'), ('dipuntempuh', 'V'), ('angin', 'N'), ('dipunantawisipun', 'ADV'), ('Sleman', 'N')]
[('Regi', 'N'), ('lombok', 'N'), ('ing', 'PRP'), ('tlatah', 'N'), ('klaten', 'N'), ('jawa', 'N'), ('tengah', 'N'), ('ngantos', 'N'), ('samangke', 'V'), ('tasih', 'N'), ('awes', 'ADJ'), ('dereng', 'V'), ('mandhap', 'N')]	[('ngantos', 'V'), ('samangke', 'KNJ'), ('dereng', 'ADV'), ('mandhap', 'ADV')]

Pada tabel VII merupakan hasil input beberapa kalimat. Tag yang sering melakukan kesalahan atau menghasilkan hasil tag kurang tepat ialah V, KNJ dapat di lihat pada tabel VII. Hasil testign tersebut juga bisa mempengaruhi hasil F-Score mengapa V, KNJ mendapatkan hasil yang rendah.

Tagset untuk tanda baca juga masih belum terlabel dengan benar. Kata pertama pada awal kalimat juga mempengaruhi akurasi, karena beberapa percobaan ternyata kata pertama sering kali mendapatkan label yang kurang tepat. Memungkinkan mengapa akurasi yang di dapat

kurang maksimal pada peneliatian ini dan juga data corpus yang lebih besar akan membawa hasil yang lebih baik. Fitur tambahan perlu diidentifikasi dan ditambahkan untuk meningkatkan hasil evaluasi. Sejauh ini dengan menggunakan CRF untuk POS Tagging memiliki banyak pembahasan yang harus dilakukan untuk meningkatkan hasil akurasi.

V. Kesimpulan

Penelitian ini menjelaskan tentang Part Of Speech Tagging untuk Bahasa Jawa menggunakan metode Conditional Random Fields (CRFs). Kami menggunakan corpus ukuran 3000 kata. Dan menggunakan cross validation untuk proses training. Penelitian ini berfokus pada penerapan metode CRF. Akurasi tertinggi diperoleh dari parameter yang telah di uji coba adalah 67%.

Diharapkan untuk peneliatian selanjut nya dapat melakukan peningkatan akurasi model sehingga akurasi pengujian dapat dioptimalkan. Gunakan metode lain untuk Tag POS dalam bahasa Jawa untuk mendapatkan metode terbaik untuk Tag POS dalam bahasa Jawa sehingga dapat dibandingkan antara metode untuk Tag POS dalam bahasa Jawa.

Referensi

- [1] Dinakaramani and A. , "Designing an Indonesian Part of speech Tagset and Manually Tagged Indonesian Corpus," 2014.
- [2] Sutton Charles, McCallum Andrew, "An Introduction to Conditional Random Fields," pp 306-307, 2011
- [3] H. R. Pramudita, E. Utami and A. Amborowati, "Pengaruh Part of Speech Tagging Berbasis Aturan dan Distribusi Probabilitas Maximum Entropy untuk Bahasa Jawa Krama," 2016.
- [4] S. Briandoko, A. R. Dewi and M. A. Setiawan, "Perbandingan Algoritma Conditional Random Field dan Hidden Markov Model pada Pos Tagging Bahasa Indonesia," 2018.
- [5] S. NM , R. K and S. K. KM, "Part-Of-Speech Tagging And Parsing Of Kannada Text Using Conditional Random Fields", 2017.
- [6] Olivo, Hari, & Fuente, "CRFPOST: Part-of-Speech Tagger for Filipino Texts using Conditional Random Fields," 2018.
- [7] Avinesh, P., and Karthik, G. "Part of Speech Tagging and Chunking using Conditional Random Fields and Transformation Based Learning". Proceedings of IJ-CAI Workshop on "Shallow Parsing for South Asian Languages," 2007
- [8] A. Pvs and G. Karthik, "Part-Of-Speech Tagging Using CRFs and TBL," pp. 1-4, 2002
- [9] F. Pisceldo , M. Adriani and R. Manurung , "Probabilistic Part Of Speech Tagging for Bahasa Indonesia," 2016.
- [10] D. S. Batista, "Conditional Random Field for Sequence Prediction," 13 November 2017. [Online]. Available:

http://www.davidsbatista.net/Conditional_Random_Fields/. [Accessed 7 November 2019].

- [11] F. Saefulloh , "Part of Speech Tagger untuk Bahasa Indonesia Menggunakan metode Conditional Random Field (CRF)," 2016.
- [12] Rusyidi, Mulyanto R.J, Supardiman, and W Sutadi. Kosa Kata Bahasa Jawa. Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan, 1985
- [13] R.Armiditya P, A. Ardiyanti Suryani and W. Maharani "Part of Speech Tagging for Javanese Ngoko Language with Hidden Markov Model," 202

