

LEXICAL RULE DAN PENGARUH PENGGUNAAN LEXICON PADA POS TAGGING BAHASA MADURA

Nindian Puspa Dewi¹, Ubaidi²

(1)Universitas Madura; Pamekasan Madura, (Contact : 087750452008, nindianpd@unira.ac.id)

(2)Universitas Madura; Pamekasan Madura, (Contact : 081939320423, ubed@unira.ac.id)

Abstrak

POS Tagging adalah dasar untuk pengembangan Text Processing suatu bahasa. Dalam penelitian ini kita meneliti pengaruh penggunaan lexicon dan perubahan morfologi kata dalam penentuan tagset yang tepat untuk suatu kata. Aturan dengan pendekatan morfologi kata seperti awalan, akhiran, dan sisipan biasa disebut sebagai lexical rule. Penelitian ini menerapkan lexical rule hasil learner dengan menggunakan algoritma Brill Tagger. Bahasa Madura adalah bahasa daerah yang digunakan di Pulau Madura dan beberapa pulau lainnya di Jawa Timur. Objek penelitian ini menggunakan Bahasa Madura yang memiliki banyak sekali variasi afiksasi dibandingkan dengan Bahasa Indonesia. Pada penelitian ini, lexicon selain digunakan untuk pencarian kata dasar Bahasa Madura juga digunakan sebagai salah satu tahap pemberian POS Tagging. Hasil ujicoba dengan menggunakan lexicon mencapai akurasi yaitu 86.61% sedangkan jika tidak menggunakan lexicon hanya mencapai akurasi 28.95 %. Dari sini dapat disimpulkan bahwa ternyata lexicon sangat berpengaruh terhadap POS Tagging.

Keyword : lexical rule, lexicon, bahasa madura, POS tagging

I. PENDAHULUAN

POS Tagging adalah dasar untuk pengembangan Text Processing suatu bahasa. POS Tagging sendiri merupakan proses pemberian tag yang berupa jenis/kelas kata pada sebuah kata, kalimat atau kumpulan kalimat. Part-of-Speech Tagging memiliki peranan penting sebagai dasar penelitian Text Processing, seperti dalam Word Sense Disambiguation, stemming dalam Information Retrieval, Text Summarization dan Question and Answering.

Penelitian mengenai Part-of-Speech Tagging di Indonesia sudah banyak dilakukan dengan menggunakan berbagai metode antara lain POS Tagging Bahasa Indonesia dengan HMM dan Rule Based oleh Kathryn Widhiyanti dan Agus Harjoko dengan nilai akurasi tertinggi 92,2% , Probabilistic Part Of Speech Tagging for Bahasa Indonesia oleh Femphy Pisceldo, Mirna Adriani & Ruli Manurung dengan nilai akurasi 95,19% dengan menggunakan 37 tagset , Implementasi Brill Tagger untuk memberikan POS-Tagging pada Dokumen Bahasa Indonesia

oleh Vinny Cristanti M, Jenny Pragantha dan Endah Purnasari dengan nilai akurasi mencapai 99,75% .

Bahasa Madura adalah bahasa daerah sehari-hari yang digunakan oleh Masyarakat Madura, baik yang tinggal di Pulau Madura maupun di luar Pulau Madura. Bahasa ini merupakan sarana komunikasi sehari-hari yang penggunaannya tidak hanya terbatas di Pulau Madura tetapi juga daerah lain di luar pulau tersebut yang masyarakatnya mayoritas dihuni oleh suku Madura. Salah satu bentuk pengembangan kebudayaan daerah, perlu adanya pembinaan dan pengembangan Bahasa Madura sebagai bahasa daerah di Indonesia, sehingga nantinya dapat mendukung kebudayaan nasional.[1]

Part-of Speech memiliki peranan penting dalam pengembangan text processing dan dapat dijadikan sebagai awal penelitian bahasa salah satunya Bahasa Madura. Karena itulah dalam penelitian ini, penulis melakukan penelitian Part-of Speech Tagging pada Bahasa Madura dengan

lexicon dan pendekatan perubahan morfologi kata dengan menggunakan leksikal rule.

Adapun dalam penelitian ini digunakan sejumlah kajian pustaka yaitu sebagai berikut :

A. Morfologi Kata

Satuan gramatikal yang merupakan satuan dasar bahasa diidentifikasi dalam cabang linguistik yang disebut dengan Morfologi. Morfologi mempelajari bentuk kata dan hal-hal yang mempengaruhi perubahan pada bentuk kata. Perubahan bentuk kata mempengaruhi terhadap golongan, arti kata dan fungsi kata itu, baik fungsi gramatik maupun fungsi semantik.

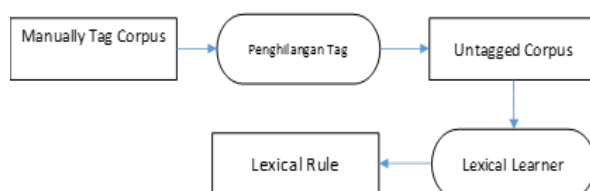
Morfem adalah bagian terkecil dari sebuah kata. Morfem adalah satuan terkecil dari kata yang sudah tidak bisa terbagi lagi. Sekalipun sudah merupakan bagian terkecil dari sebuah kata, setiap morfem tetap memiliki makna baik gramatikal maupun leksikal.[2]

Proses pembentukan morfen sebagai berikut :[3]

1. Satu paragraf dapat dipecah menjadi beberapa kalimat.
2. Kalimat dapat dipecah menjadi bagian makna terkecil yaitu beberapa kata.

menggunakan proses learner dari Brill Tagger. Brill Tagger adalah algoritma untuk melakukan pemberian tag pada kata dengan dua aturan, yaitu : lexical rule untuk memberi tag pada unknown word dan Contextual Rule untuk meningkatkan akurasi dari perbaikan skor.

Lexical rule diperoleh melalui lexical learner. Rule ini digunakan untuk melabeli kata dengan memperhatikan perubahan bentuk kata karena adanya perubahan morfologi kata. Berikut proses leksikal learner untuk menghasilkan leksikal rule dapat dilihat pada gambar 1.



3. Kata dapat terdiri dari beberapa morfem, sebagai contoh kata membacakan dipecah menjadi me-baca-kan. Terdapat juga kata yang sudah merupakan satu morfen seperti nama benda dan kata asal, misalnya buku, rumah, sakit, pergi, masuk.

Dalam kaitannya dengan kebahasaan, yang dipelajari dalam morfologi ialah bentuk kata. Selain itu, perubahan bentuk kata dan makna (arti) yang muncul serta perubahan kelas kata yang disebabkan perubahan bentuk kata itu.

Proses morfologis adalah peristiwa pembentukan kata. Ada tiga macam proses morfologi, yaitu afiksasi, reduplikasi, dan pemajemukan/komposisi. Afiksasi merupakan proses pemberian atau penambahan afiks (afiksasi) atau imbuhan, baik awalan, akhiran dan sisipan. Reduplikasi merupakan proses pembentukan kata ulang sedangkan pemajemukan atau komposisi merupakan proses penggabungan dua kata atau lebih sehingga membentuk kata majemuk atau kata yang memiliki arti baru.

B. Lexical Rule

Leksikal rule ada rule yang terbentuk dari proses pelatihan leksikal learner[4]. Leksikal learner yang akan dibahas pada penelitian ini

Gambar 1. Blok Diagram Proses Training pada Lexical Learner

C. Lexicon

Lexicon berasal dari bahasa Yunani yakni, lexikón atau lexikós yang berarti kata, ucapan, atau cara bicara. Istilah lexicon lazim digunakan untuk mewadahi konsep kumpulan leksem dari suatu bahasa, baik kumpulan secara keseluruhan, maupun secara sebagian.[5] Dalam penelitian ini lexicon diartikan sebagai kumpulan kata dan kelas katanya. Lexicon ini biasa digunakan dalam penelitian bahasa alami.

D. Part-of Speech

Part-of Speech (POS) biasa dikenal sebagai jenis kata dalam sebuah kalimat seperti kata kerja (verb), kata sifat (adjective), kata benda (noun) dan sebagainya.[6]

E. POS Tagging

Part-of Speech (POS) biasa dikenal sebagai jenis kata dalam sebuah kalimat seperti kata kerja (verb), kata sifat (adjective), kata benda (noun) dan sebagainya. Part-of Speech (POS) Tagger (tagging) merupakan suatu proses memberi tanda (mark) kelas kata atau Part-of Speech pada setiap kata dalam kalimat.[7]

Tagging dapat dimanfaatkan pada aplikasi bahasa alami lainnya. Beberapa penggunaan POS Tagging adalah membantu dalam membedakan antar kata yang memiliki bentuk yang sama tetapi arti yang berbeda sesuai (ambigu), dapat digunakan pada proses stemming dalam information retrieval, dan untuk proses pencarian kata benda. POS-Tagging dapat dilakukan secara manual maupun otomatis. POS Tagging secara manual dilakukan dengan meminta bantuan satu atau beberapa ahli bahasa untuk memberikan tag yang sesuai pada kata.[8] Berbeda dari yang manual, POS-Tagging secara otomatis dilakukan oleh sistem dengan menerapkan metode matematika, sejumlah rule dan lainnya.

Keberadaan part-of speech tagging memiliki peran yang sangat penting dalam beberapa aplikasi Natural Language Processing modern, seperti:[9]

1. Word Sense Disambiguation

Salah satu contoh kata-kata yang ambigu dalam bahasa Indonesia adalah kata "bisa". Kata "bisa" dalam bahasa Indonesia memiliki arti "dapat" dan juga memiliki arti "racun ular". Selain memiliki perbedaan dalam arti kata, kedua kata "bisa" tersebut juga memiliki perbedaan dalam kelas kata atau part-of speech. Kata "bisa" yang berarti "dapat" memiliki kelas kata verb (kata kerja), sedangkan kata "bisa" yang berarti "racun ular" memiliki kelas kata noun (kata benda). Hal ini membuktikan bahwa part-of speech dapat membantu dalam membedakan antar kata yang memiliki bentuk yang sama tetapi arti yang berbeda sesuai dengan konteks kalimatnya.

2. Text to Speech

Keberadaan Part-of Speech Tagging juga membantu dalam peningkatan akurasi dari pelafalan pada speech synthesis system dan akan menambah akurasi dari speech recognition

system. Sebagai contoh, kata "object" memiliki memiliki 2 kelas kata yaitu, noun dan verb. Saat kata "object" memiliki kelas kata noun, maka kata "object" tersebut memiliki pelafalan "OBject". Sedangkan kata "object" yang memiliki kelas kata verb, akan memiliki pelafalan "obJECT".

3. Stemming dalam Information Retrieval

Part-of Speech juga dapat digunakan pada proses stemming dalam information retrieval, karena dengan mengetahui part-of speech dari sebuah kata maka sekaligus juga membantu mengenali bentuk morfologi suatu kata.

II. METODOLOGI

Perubahan kelas kata dapat disebabkan perubahan bentuk kata itu. Perubahan bentuk kata dapat terjadi dari tiga proses yaitu proses afiksasi, reduplikasi dan komposisi. Adapun pendekatan morfologi yang digunakan saat ini dibatasi hanya menggunakan proses afiksasi dan reduplikasi.

Dalam penelitian ini, proses POS Tagging dimulai dengan mempelajari dan mengumpulkan proses afiksasi dalam Bahasa Madura, penyusunan lexicon, training dan testing. Berikut macam-macam afiksasi dalam Bahasa Madura dapat dilihat pada Tabel 1.[10]

Tabel 1. Macam – Macam Afiksasi Dalam Bahasa Madura

Prefik	Suffix	Konfix	Infix
n-	-a	N-è	-al-
a-	-na	N-aghi	-ar-
ta-	-an	N-an	-èn-
ma-	-è	a-aghi	-om-
ka-	-aghi	a-an	
ka-		è-è	
sa-		èka-è	
pa-		è-aghi	
nga-		ma-an	
è-		ma-aghi	
èpa-		ka-an	
		ka-è	
		ka-aghi	
		sa-na	
		paN-an	
		pa-aghi	
		Nga-è	

Masing-masing proses afiksasi dapat menimbulkan perubahan bentuk kata dan kelas kata. Contoh perubahan bentuk kata dapat dilihat pada tabel 2 berikut ini.

Tabel 2. Perubahan Bentuk Kata karena Afiksasi

Prefiks	Contoh
N- (N-, Ng-, M-, Ny)	Kakan (makan) - ngakan (memakan)
a-	main (main) - amain (bermain)
ma-	Tao (Tau) - Matao (memberi tau)
ta-	Porop (tukar) - Taporop (tertukar)
ka-	sarong (sarung) - kasarong (jadikan sarung)
pa-	Robbhu (robok) - Parobbhu (robokkanlah)
nga-	Pele (pilih) - Ngapele (tinggal pilih)
è-	Tèmpak (tendang) - Ètèmpak (ditendang)
èka-	Jhámo (jamu) - Èkajhámo (dijadikan jamu)
èpa-	Jháu (jauh) - Èpajháu (dijauhkan)
Sufiks	Contoh
-è	Salebbár (celana) - Salebbárr (celanailah)
-aghi	Ghibá (bawa) - Ghibáághi (bawakanlah)
Konfiks	Contoh
N-è	Kolè' (kulit) - Ngolè'è (memberi kulit)
N-aghi	Buwang (buang) - Mowangngaghi (membuangkan)
N-ana	Ngosod (hapus) - Ngosodána (akan menghapuskan)
a-è	Jhálán (jalan) - Ajhálánè (menjalani')
a-aghi	Jhálán (jalan) - Ajhálánnaghi (menjalankan)
a-an	Kalambi (baju) - Akalambian (memakai baju)

Selain perubahan bentuk kata, afiksasi juga menyebabkan perubahan kelas kata. Tabel 3 berikut adalah contoh perubahan kelas kata yang disebabkan karena perubahan bentuk kata.

Tabel 3. Perubahan Kelas Kata karena Afiksasi

No	Ubah Ke-		Kondisi
	Kata Dasar	Ke-	
1	VBI	VB	Prefix N- berupa N-, Ng-, M-, Ny-
2	VBI	VBI	Prefix N-
3	N	VBI	Prefix N-
4	N	VB	Prefix N-
5	N	ADJ	Prefix N-
6	ADJ	VBI	Prefix N-
7	NPT	VBI	Prefix N-
8	VBI	VBI	Prefix a-
9	N	VB	Prefix a-
10	N	VBI	Prefix a-
11	VBI	VBI	Prefix ta-
12	N	VBI	Prefix ta-
13	ADJ	ADJ	Prefix ta-
14	VBI	VB	Prefix ma-
15	ADJ	VB	Prefix ma-
16	ADJ	VBI	Prefix ma- dan bereduplikasi
17	ADJ	VB	Prefix ma-
18	VBI	VBI	Prefix ka-
19	N	VBI	Prefix ka-
20	ADJ	VBI	Prefix ka-
21	NPT	VBI	Prefix ka-
22	N	N	Prefix sa-
23	N	N	Prefix sa- dan sufiks -na
24	VBI	NTT	Prefix sa- + konfiks {paN-an}
25	VBI	VBI	Prefix pa-
26	ADJ	VBI	Prefix pa-
27	NPT	VBI	Prefix pa-
28	VB	VB	Prefix pa- + N-
29	VBI/N	N	Prefix pa- atau pè-
30	ADJ	N	Prefix pa-

Selanjutnya melakukan penyusunan lexicon yaitu proses pengumpulan kurang lebih 10.000 kata dalam Bahasa Madura yang diambil dari sejumlah artikel dan cerita Bahasa Madura. Semua kata ini kemudian diberi kelas kata secara manual. Berikut adalah contoh kalimat sebelum dan sesudah dilabeli :

Sebelum dilabeli :

Kabhupatèn Songennep
kadháddhiyán daratan jhughán
polo. Bádá atosán polo è
kabhupatèn palèng tèmor è
Madhurá panèka. Artèna,
kabhupatèn panèka jhughán
ghádhuwán tasè' sè cè' lèbárrá
ètèmbháng tello' kabhupatèn laèn
è Madhurá.

Setelah dilabeli :

Kabhupatèn/NNP Songennep/NNP
kadháddhiyán/VBI daratan/NN
jhughán/CC polo/NN ./.. Bádá/VBI
atosán/CDC polo/NN è/IN
kabhupatèn/NN palèng/RB tèmor/NN
è/IN Madhurá/NNP panèka/DT ./..
Artèna/NNG ,/, kabhupatèn/NN
panèka/DT jhughán/RB
ghádhuwán/VBT tasè'/NN sè/SC
cè'/RB lèbárrá/JJ ètèmbháng/SC
tello'/CDP kabhupatèn/NN laèn/JJ
è/IN Madhurá/NNP ./..

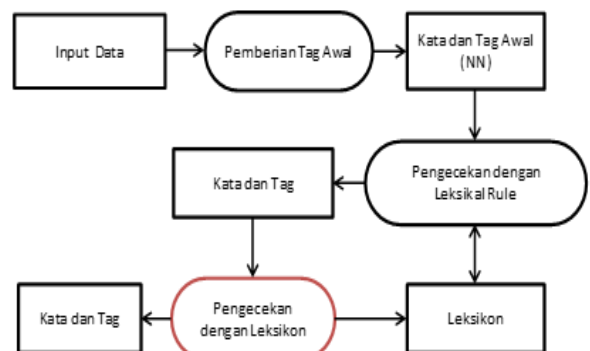
Dari hasil penyusunan dataset diketahui bahwa tagset NN memiliki jumlah terbanyak yaitu sebanyak 2284 yang artinya ada 2284 kata yang ditag dengan common noun (NN). Tagset tertinggi berikutnya adalah VBT sebanyak 933 dan diikuti dengan IN sebanyak 888. Sedangkan jumlah tagset terendah adalah Dash, AR, CDD dan Elipsis (...) yang hanya berjumlah 1. Berikut grafik sebaran tagset yang digunakan dalam penelitian dapat dilihat pada gambar 2.



Gambar 2. Grafik Sebaran Tagset

Proses POS Tagging dimulai dengan menginputkan data berupa kata, kalimat atau paragraf yang akan dicari Tagsetnya. Semua kata yang diinputkan kemudian akan diberi tag

awal sebagai kata benda (NN/NNP). Setelah itu semua input data akan dicek afiksasinya sesuai dengan rule yang ada dalam lexical rule. Pada proses ini juga dilakukan pengecekan kata hasil pemotongan dan penambahan afiksasi dibandingkan dengan lexicon, lalu dicek kelas kata dilexicon sehingga dapat diketahui kelas kata yang tepat sesuai dengan perubahan bentuk kata. Jika perubahan kata tidak ditemukan dalam lexicon maka kelas kata diubah berdasarkan perkiraan afiksasi yang sesuai. Terakhir semua input kata dibandingkan dengan lexicon. Jika kata ditemukan dalam lexicon maka ubah kelas kata sesuai dengan kelas kata dalam lexicon. Gambar 3 menunjukkan proses POS Tagging dalam penelitian ini.



Gambar 3. Blok Diagram Proses Tagging

III. HASIL DAN PEMBAHASAN

Uji coba dilakukan dengan menggunakan tagset Bahasa Madura yang terdiri dari 36 tagset yang dapat dilihat pada tabel 4. Penyusunan data set Bahasa Madura dilakukan dengan mengumpulkan artikel berbahasa Madura sejumlah 10.443 kata dan diberi tagset secara manual dengan menggunakan Tagset Bahasa Madura yang telah ditentukan. Hasil pelabelan ini kemudian dijadikan file lexicon.

Tabel 4. Tagset Bahasa Madura

No	Tagset	Contoh	Simbol
1	Verba Transitif	Matoron	VBT
2	Verba Intransitif	Entar	VBI
3	Adjective	Alpo'	JJ

4	Adverb	Ghi'ta'	RB
5	Common Noun	Buku	NN
6	Proper Common Noun	Hasan	NNP
7	Genetive Common Noun	Klambina h	NNG
8	Personal Pronoun	Sengko'	PRP
9	Locative Pronoun	Diyá	PRL
10	Primary Numeral	Settong	CDP
11	Collective Numeral	ebuen	CDC
12	Distributive Numeral	Wa'-daduwa'	CDD
13	Irregular Numeral	Kabbhi, banyak	CDI
14	Kata bantu bilangan	kg, bhutol	CDB
15	WH Pronoun	Sapa, apa	WPRP
16	WH Adverb	Bárámma, arapa	WRB
17	Determiner	Areya, reya	DT
18	Article	Se, para ,	AR
19	Preposition	sang Dari, e,	IN
20	Coordinate Conjunction	polana Ban,	CC
21	Subordinate Conjunction	otaba Samarena	SC
22	Particle	Ya, ra, la,	RP
23	Interjection	kan e, he, sst,	UH
24	Positive Modal	oy Pola,	MD
25	Negative Modal	ma'pola Enjá', ta', enjá'	NEG
26	Symbol	@#^&	Sym
27	Sentence Terminator	?!.	ST
28	Comma	,	,
29	Ellipsis	...	...
30	Colon	:	:
31	Semi Colon	;	;

32	Open Paranthesis	(	OP
33	Close Paranthesis	)	CP
34	Quotation	' , “	QT
35	Dash	—	DASH
36	Slash	/	GM

Untuk perhitungan nilai akurasi digunakan dua jenis perhitungan yaitu perhitungan tanpa proses pengecekan dengan file lexicon dan perhitungan akurasi dengan pengecekan lexicon. Dari hasil percobaan lexical learner untuk threshold 10 menghasilkan 48 rule, threshold 20 sampai dengan 40 mengalami penurunan jumlah rule yaitu hanya 32 rule. Begitu juga untuk threshold 50 menghasilkan rule yang sama yaitu sebanyak 13 rule. Hasil ini menunjukkan bahwa semakin tinggi nilai threshold maka semakin kecil jumlah rule yang dihasilkan. Begitu juga sebaliknya, semakin besar nilai threshold semakin kecil jumlah rule yang dihasilkan. Beberapa contoh Hasil lexical rule dapat dilihat pada tabel 5 di bawah ini.

Tabel 5. Contoh Lexical Rule

Rul e	Lexical Rule	Tag Awal	Tag Baru	Kondisi
1	è haspref VBT		VBT	Jika berawalan è
2	na hassuf NNG		NNG	Jika berakhir na
3	NN a fhaspref VBT	NN	VBT	Jika berawalan a
4	sè goodonright VBT		VBT	Jika kata disebelah kiri adalah “sè”
5	NN sè fgoodonright JJ	NN	JJ	Jika kata disebelah kiri

adalah "sè" dan tag awal "NN"
--

50	13	87.25
----	----	-------

Rule 1 berarti jika sebuah kata memiliki prefix è- maka akan diubah kelas katanya menjadi VBT. Rule 2 akan dikenai pada kata yang memiliki suffix -na dan diubah kelas katanya menjadi NNG. Rule 3 artinya jika sebuah kata memiliki kelas kata awal NN dan berawalan a, maka akan diubah kelas katanya menjadi VBT.

Dari serangkaian percobaan dengan beberapa perubahan nilai threshold, tabel 6 menunjukkan rata-rata akurasi yang diperoleh dengan hanya menerapkan lexical rule tanpa pengecekan lexicon yaitu 27.57% dengan akurasi tertinggi hanya mencapai 28.95%, sedangkan pada tabel 7 menunjukkan rata-rata akurasi uji coba dengan menggunakan lexical rule dan lexicon mencapai 86.5% dengan akurasi tertinggi 86.61%. Hasil ini menunjukkan bahwa lexicon sangat penting keberadaannya khususnya untuk memberi tagset pada kata asal yang belum atau tidak mengalami perubahan bentuk kata.

Tabel 6. Akurasi Hasil Tahap Lexical Tanpa Lexicon

T	Jum Rule	Akurasi (%)
10	48	28.95
20		
30	32	28.95
40		
50	13	24.81

Tabel 7. Akurasi Hasil Tahap Lexical dengan Lexicon

T	Jumlah Rule	Akurasi (%)
10	48	87.43
20		
30	32	87.43
40		

IV. SIMPULAN DAN SARAN

4.1 Simpulan

Nilai threshold pada lexical learner mempengaruhi jumlah rule yang diperoleh dalam proses training. Semakin rendah nilai threshold maka semakin banyak rule yang diperoleh dan begitu juga sebaliknya semakin tinggi nilai threshold maka semakin sedikit jumlah rule yang diperoleh. Hasil percobaan menunjukkan bahwa semakin banyak rule yang diperoleh maka nilai akurasi semakin tinggi.

Pengujian dengan leksikal rule dengan menggunakan pencocokan lexicon menghasilkan rata-rata akurasi yaitu 87.37% dengan akurasi tertinggi 87.25%, sedangkan pengujian tanpa menggunakan lexicon menghasilkan rata-rata akurasi 27.57% dengan akurasi tertinggi 28.95.1%. Hal ini menunjukkan bahwa akurasi dengan pencocokan dengan file lexicon lebih baik daripada tanpa lexicon.

4.2 Saran

1. Penambahan data corpus perlu dilakukan mengingat pentingnya keberadaan lexicon yang merupakan separuh dari jumlah data corpus.
2. Penerapan Contextual Rule sebagai salah satu bagian dari Brill Tagger untuk POS Tagging Bahasa Madura.
3. Perbaikan data corpus terutama mengenai penulisan karena masih ada data dengan kata yang sama namun cara penulisannya berbeda.

V. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada keluarga, teman dan para ahli Bahasa Madura yang telah memberi dukungan terhadap penelitian ini. Penulis juga ucapkan terima kasih kepada semua pihak yang penelitian telah penulis gunakan sebagai acuan dan pembelajaran sehingga penelitian ini bisa terselesaikan.

REFERENSI

- [1] A. Halim, *Politik Bahasa Nasional 1 dan 2*. Jakarta: Aneka Ilmu, 1976.
- [2] A. Chaer, *Morfologi Bahasa Indonesia (Pendekatan Proses)*. Jakarta: Rineka Cipta, 2008.
- [3] A. Carstairs-McCarthy, "An introduction to English morphology: words and their structure," *Linguistics*, p. 160, 2002.
- [4] E. Brill, "A simple rule-based part of speech tagger," *Proc. third Conf. Appl. Nat. Lang. Process.* -, p. 152, 1992.
- [5] A. Chaer, *Linguistik Umum*. Jakarta: Rineka Cipta, 2007.
- [6] Chris Manning and H. Schütze, *Foundation of Statistical Natural Language Processing*. Cambridge: MIT Press Textbook on statistical and probabilistic methods in NLP., 1999.
- [7] V. C. M, J. Pragantha, and E. Purnamasari, "Implementasi Brill Tagger Untuk Memberikan," *Jakarta J. Tek. dan Ilmu Komput.*, pp. 301–315, 2012.
- [8] T. Chandrawati, "Pengembangan Part of Speech Tagger untuk Bahasa Indonesia Berdasarkan Metode Conditional Random Fields dan Transformation Based Learning," p. 465, 2008.
- [9] E. R. Setyaningsih, "Penetapan Tagset dan Modifikasi Brill Tagger untuk Part-of Speech Bahasa Indonesia," Sekolah Tinggi Teknik Surabaya, 2014.
- [10] A. et al Sofyan, *No TitleTata Bahasa Bahasa Madura*. Sidoarjo: Bahasa Surabaya, 2017.