

**PELABELAN KELAS KATA TEKS BERBAHASA INDONESIA
BERBASIS BRILL TAGGER**

Skripsi



oleh

STEPHAN KENT DIORIZKY WIJAYA

71140017

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA

2019

PELABELAN KELAS KATA TEKS BERBAHASA INDONESIA BERBASIS BRILL TAGGER

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

STEPHAN KENT DIORIZKY WIJAYA

71140017

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA

2019

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PELABELAN KELAS KATA TEKS BERBAHASA INDONESIA BERBASIS BRILL TAGGER

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 5 Agustus 2019



STEPHAN KENT DIORIZKY
WIJAYA
71140017

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

Judul Skripsi : PELABELAN KELAS KATA TEKS BERBAHASA
INDONESIA BERBASIS BRILL TAGGER
Nama Mahasiswa : STEPHAN KENT DIORIZKY WIJAYA
N I M : 71140017
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Genap
Tahun Akademik : 2018/2019

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 5 Agustus 2019

Dosen Pembimbing I



Lucia Dwi Krisnawati, Dr. Phil.

Dosen Pembimbing II



Gloria Virginia, S.Kom., MAI, Ph.D.

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

PELABELAN KELAS KATA TEKS BERBAHASA INDONESIA BERBASIS BRILL TAGGER

Oleh: STEPHAN KENT DIORIZKY WIJAYA / 71140017

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 26 Juli 2019

Yogyakarta, 5 Agustus 2019
Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. Restyandito, S.Kom., MSIS, Ph.D
4. Lukas Chrisantyo, S.Kom., M.Eng.



Dekan

(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji Tuhan, Segala puji syukur kepada Tuhan Yang Maha Esa, berkat rahmat dan pertolongan-Nya sehingga penulis dapat menyelesaikan skripsi ini. Berkat serta penyertaan-Nya selalu tercurah kepada umat-Nya sepanjang masa. Penulis menyadari sepenuhnya bahwa begitu banyak kesempatan ini, dengan segala kerendahan hati, penulis mengucapkan banyak – banyak terima kasih kepada :

1. Bapak Budi Susanto, S.Kom., M.T. Selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia S.Kom., MAI. Selaku Ketua Program Studi Informatika Universitas Kristen Duta Wacana.
3. Ibu Dr. Lucia Dwi Krisnawati selaku dosen pembimbing I yang telah memberikan waktunya secara rutin dan menghadapi mahasiswa pembimbing dengan sabar untuk melakukan konsultasi, memberikan solusi dalam hal pemograman dan saran dalam setiap menemui masalah.
4. Ibu Gloria Virginia S.Kom., MAI. selaku dosen pembimbing II yang telah memberikan waktu luangnya dan menghadapi mahasiswa pembimbing dengan sabar untuk melakukan konsultasi, dan telah memberikan banyak saran dalam hal penulisan supaya menjadi baik.
5. Untuk kedua orang tua dan adik, Penulis mengucapkan terima kasih atas dukungan dalam bentuk doa, semangat dan biaya selama perkuliahan.
6. Seftika Dyah Palupi yang telah ikut serta membantu dalam pengerjaan skripsi dan dukungan moral kepada penulis.
7. Peter Jimie Kusuma, Anom Bayu Pradnyana, anggota ALX Yk selaku sahabat-sahabat penulis serta teman-teman seperjuangan skripsi, yang telah turut mendukung dan menyemangati dalam proses skripsi penulis.
8. Pihak-pihak lain yang ikut membantu dalam proses skripsi, namun tidak dapat disebutkan satu persatu oleh penulis. Semoga Tuhan senantiasa melimpahkan berkat bagi kita semua dan terima kasih untuk bantuannya selama ini, Tuhan Memberkati, Amin.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi perkembangan ilmu pengetahuan khususnya dibidang teknologi informasi.

©UKDW

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena pertolongan dan rahmat-Nya, sehingga penulis dapat menyelesaikan pembuatan sistem dan laporan tugas akhir dengan judul “*PELABELAN KELAS KATA TEKS BERBAHASA INDONESIA BERBASIS BRILL TAGGER*” dengan baik.

Penulisan laporan tugas akhir kuliah ini diajukan sebagai salah satu syarat guna mencapai gelar Sarjana Strata Satu (S1) di Fakultas Teknologi Informasi Program Studi Informatika Universitas Kristen Duta Wacana Yogyakarta.

Dalam pembuatan laporan ini, penulis menyadari masih ada kekurangan, baik dari materi maupun teknik penyajiannya. Oleh karena itu, penulis mengharapkan adanya saran dan kritik dari pembaca. Akhir kata penulis mohon maaf apabila dalam penulisan laporan ini, ada kata atau kalimat yang kurang berkenan. Semoga hasil dari pengerjaan tugas akhir kuliah ini dapat bermanfaat bagi banyak pihak.

Yogyakarta, 21 Juli 2019

Penulis

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR	viii
INTISARI.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian	3
1.5. Metodologi Penelitian	4
1.6. Sistematika Penulisan	5
BAB 2 TINJAUAN PUSTAKA	7
2.1. Tinjauan Pustaka	7
2.2. Landasan Teori.....	9
2.2.1. Kelas Kata	9
2.2.1.1 Kelas Kata Terbuka	9
2.2.1.2 Kelas Kata Tertutup.....	13
2.2.2. POS Tangging.....	18

2.2.3.	N-gram	19
2.2.4.	Tagset	20
2.2.5.	Smoothing	22
2.2.6.	Algoritma Brill Tagger.....	23
2.2.7.	Evaluasi	24
BAB 3 PERANCANGAN SISTEM		25
3.1	Tahapan Pembangunan Sistem	25
3.1.1.	Pra-pemrosesan	25
3.1.1.1.	Pengumpulan Dokumen Data Latih	25
3.1.1.2.	Pembuatan Label Kelas Kata Bahasa Indonesia	25
3.1.1.3.	Perancangan Basis Data Data Latih	26
3.1.1.4	Prapemrosesan Dokumen Latih	27
3.1.2.	Pemrosesan.....	27
3.1.3.	Pasca Pemrosesan.....	28
3.2.	Analisis Kebutuhan Sistem	28
3.3.	Rancangan Sistem	28
3.3.1	Alur Kerja Sistem Secara Umum.....	28
3.3.1.1.	Alur Kerja Pembuatan Data Latih.....	29
3.3.1.2.	Alur Kerja Pembuatan Sistem.....	30
3.3.1.3.	Alur Kerja Evaluasi.....	31
3.4.	User Interface	32
3.5.	Rancangan Pengujian	33
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....		35
4.1.	Implementasi Sistem	35
4.1.1.	Implementasi Pembuatan Dokumen Latih	35

4.1.1.1. Tahap Normalisasi	36
4.1.1.2. Tahap Tokenisasi	37
4.1.1.3. Tahap Perhitungan Frekuensi Kelas Kata	38
4.1.2. Implementasi Basis Data.....	41
4.1.3. Implementasi Antar Muka Sistem.....	45
4.1.4. Implementasi Pembuatan Dokumen Evaluasi.....	46
4.2. Implementasi Program	46
4.2.1. Proses Prapemrosesan	47
4.2.2. Proses Perhitungan Probabilitas Menggunakan N-GRAM.....	49
4.2.3. Pembuatan Aturan kelas kata Bahasa Indonesia.....	52
4.3. Analisis Sistem.....	55
4.3.1. Presisi	56
4.3.2. Akurasi	57
4.3.3. Pembahasan Analisis.....	58
4.3.4. Kekurangan Sistem	62
BAB 5 KESIMPULAN DAN SARAN	64
5.1 Kesimpulan	64
5.2 Saran.....	65
DAFTAR PUSTAKA	66
LAMPIRAN A	1
LAMPIRAN B	1

DAFTAR TABEL

Tabel 2.1. Tagset II Bahasa Indonesia	21
Tabel 3.1. Label Kelas kata Bahasa Indonesia.....	26
Tabel 3.2. Tabel Confusion Matrix	34
Tabel 4.1. Daftar Pembuatan Aturan Kelas Kata Bahasa Indonesia.....	68
Tabel 4.2. Pengujian Presisi Dokumen Uji 1 dengan Dokumen Evaluasi 1	56
Tabel 4.3. Hasil Pengujian Sistem Dengan 30 Dokumen Uji.....	58

DAFTAR GAMBAR

<i>Gambar 2.1</i> Alur Algoritma Brill Tagger	24
<i>Gambar 3.1</i> Rancangan Basis Data Data Latih	27
<i>Gambar 3.2</i> Struktur Tabel di dalam Basis Data	27
<i>Gambar 3.3</i> Alur Kerja Sistem Secara Umum.....	29
<i>Gambar 3.4</i> Alur Kerja Pembuatan Data Latih	29
<i>Gambar 3.5</i> Alur Kerja Pembangunan Sistem.....	30
<i>Gambar 3.6</i> Alur Kerja Evaluasi	31
<i>Gambar 3.7</i> Tampilan awal Halaman POS Tagging Bahasa Indonesia	32
<i>Gambar 3.8</i> Tampilan awal Halaman POS Tagging Bahasa Indonesia	33
<i>Gambar 4.1</i> Desain Basis Data Data Latih	54
<i>Gambar 4.2</i> Daftar Tabel Data Latih	42
<i>Gambar 4.3</i> Struktur Tabel Unigram	42
<i>Gambar 4.4</i> Struktur Tabel Bigram sampai Pentagram.....	43
<i>Gambar 4.5</i> Contoh Isi tabel <i>unigram</i>	43
<i>Gambar 4.6</i> Contoh Isi tabel bigram dan trigram	44
<i>Gambar 4.7</i> Contoh Isi tabel quadgram dan pentagram	44
<i>Gambar 4.8</i> Tampilan awal sistem	45
<i>Gambar 4.9</i> Tampilan hasil keluaran sistem.....	46
<i>Gambar 4.10</i> Hasil Presisi Pengujian Sistem	57
<i>Gambar 4.11</i> Grafik Rata-Rata Presisi Setiap Kelas	59
<i>Gambar 4.12</i> Grafik Rata-Rata Akurasi Setiap Dokumen.....	60

DAFTAR LAMPIRAN

LAMPIRAN A	1
LAMPIRAN B	1

©UKPDW

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Masyarakat umum pada saat ini lebih cenderung berkomunikasi dengan komputer. Bentuk komunikasi yang dilakukan oleh masyarakat umum dengan komputer seperti mencari artikel menggunakan *google*, melakukan penterjemahan mesin menggunakan aplikasi *google translate* dan ekstraksi informasi tertentu dengan pola khusus. Pemberian pengetahuan kepada komputer sangat diperlukan, sehingga komputer dapat berkomunikasi dengan manusia menggunakan bahasa alami. Untuk menghasilkan komunikasi yang baik antara komputer dengan manusia ada beberapa metode, antara lain metode *rule-based*, statistik dan *transformation based learning*.

Metode-metode tersebut dapat menghasilkan pelabelan kata. Salah satu pelabelan kata adalah *POS Tagging*. *POS Tagging (Part-Of-Speech Tagging)* adalah proses penandaan untuk menempatkan *part-of-speech* atau penanda kelas leksikal lain untuk setiap kata dalam korpus (Jurafsky & Martin, 2000). Contoh dari *POS Tagging*, *saya makan nasi* menjadi *saya/KG makan/V nasi/NN* dimana label KG = kata ganti, V = kata kerja, NN = kata benda (Christanti, Pragantha, & Purnamasari, 2012). Akan tetapi komputer juga belum mengetahui simbol atau *tagset* yang dihasilkan dari *POS Tagging*. Maka dari itu komputer harus diberi metode pembelajaran atau bisa dibilang komputer harus diberi pembelajaran agar komputer mengetahui label kata atau *tagset*.

Metode pembelajaran meliputi *unsupervised*, *semi-supervised*, dan *supervised*. Menurut Stratos & Collins (2015) pengertian dari *semi-supervised* adalah pelabelan dokumen yang memanfaatkan sumber *tagset* seapa-adanya di dalam kamus yang sudah di *POS Tagging*. Sedangkan pengertian *supervised* menurut Karibun (2005) adalah dokumen yang sudah dilabeli

secara manual dimasukkan ke dalam sistem untuk dipelajari. Selanjutnya sistem akan membentuk basis data pola berdasarkan dokumen yang dimasukkan. Dari pengertian tersebut terdapat perbedaan antara metode *supervised* dengan *semi-supervised*. *Supervised* memerlukan pelabelan secara manual, sedangkan *semi-supervised* hanya memerlukan sebagian pelabelan secara manual dan memerlukan *training data* supaya komputer bisa melabelkan secara mandiri. Salah satu contoh dari metode *semi-supervised* adalah algoritma *Brill Tagger*. Menurut Ngai & Florian (2001) algoritma *Brill Tagger* merupakan salah satu algoritma *machine learning* berdasarkan aturan dengan pembelajaran berbasis transformasi. Algoritma ini sangat bagus dan mudah dipakai dalam menyelesaikan permasalahan yang berkaitan dengan pengolahan bahasa alami.

Untuk itu peneliti ingin membangun dan mengembangkan sistem *semi-supervised* pelabelan dokumen dengan cara pelabelan kelas kata berbasis transformasi. Penelitian ini merupakan penelitian lanjutan dan pengembangan dari penelitian *POS tagging* berbasis *rule based* milik Elnatan (2017) dan Kristanto (2017). Penelitian ini juga menggunakan data dari penelitian sebelumnya dan ditambah dengan data baru di penelitian yang dikerjakan peneliti saat ini. Dari latar belakang yang sudah dijelaskan, peneliti mengharapkan hasil dari penelitian ini adalah sebuah sistem *semi-supervised* yang dapat melabelkan dokumen dengan cara pelabelan kelas kata berbasis model transformasi dan dapat menghasilkan *corpus* yang teranotasi untuk penelitian selanjutnya.

1.2. Rumusan Masalah

Dengan latar belakang yang sudah dijelaskan, maka penulis merumuskan beberapa rumusan masalah sebagai berikut:

1. Bagaimana membuat *gold standart* sebagai bahan pembelajaran?
2. Bagaimana menerapkan algoritma *Brill Tagger* berbasis n-gram?

3. Bagaimana menyusun aturan transformasi berdasarkan kesalahan pelabelan otomatis?
4. Bagaimana mengukur tingkat akurasi pelabelan algoritma *Brill Tagger*?

1.3. Batasan Masalah

Peneliti membatasi beberapa masalah untuk memudahkan membangun dan mengembangkan sistem *semi-supervised* pelabelan kelas kata teks berbahasa indonesia berbasis model transformasi sebagai berikut:

1. Dokumen *training* yang dipakai berjumlah 500 dokumen.
2. Dokumen evaluasi yang dipakai berjumlah 30 dokumen. 30 dokumen tersebut diambil dari berita online dan dokumen evaluasi tidak sama dengan dokumen yang dipakai untuk pembuatan basis data.
3. Dokumen yang diujikan adalah dokumen berita online yang diambil dari kompas.com, detik.com, dan liputan6.com karena bahasa yang digunakan dalam berita online bersifat formal dan memiliki struktur bahasa yang jelas.
4. Menggunakan N-gram kata sebagai metode pembelajaran sistem. Dengan $N = 2$ sampai 5.
5. Kelas kata bahasa indonesia yang akan digunakan adalah verba, adjektif, nomina, pronomina, numeralia, adverbial, preposisi dan konjungsi.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk membangun sistem *semi-supervised* yang dapat memberi label pada dokumen dengan cara pelabelan kelas kata berbasis model transformasi.

1.5. Metodologi Penelitian

1. Metode Pengumpulan Data

Pengumpulan dokumen uji dan dokumen latih yang digunakan diambil dari artikel - artikel berita *online* yang menggunakan bahasa Indonesia yang baku. Jenis artikel yang akan digunakan adalah bisnis, olahraga, teknologi, kesehatan, politik, gaya hidup, perjalanan, edukasi, otomotif dan lain-lain. Sumber yang akan digunakan untuk artikel *online* diambil dari kompas.com, dan detik.com. Berikut contoh sumber artikel berita:

- <http://bola.kompas.com/read/2016/10/13/21203498/pemain.barca.yang.bel.um.cetak.gol.ini.akan.diperpanjang.kontraknya>
- <http://news.detik.com/berita/3320383/jokowi-pastikan-menteri-esdm-dari-kalangan-profesional>

2. Implementasi atau coding

Pada tahap ini peneliti akan menggunakan 3 metode untuk membangun sistem. 3 metode itu adalah metode pengumpulan data, metode pengembangan sistem dan metode evaluasi. Metode pengumpulan data terbagi menjadi 2 yaitu pengumpulan dokumen latih dan dokumen uji. Dokumen uji didapatkan dari dokumen uji skripsi Elnatan (2017), karena dokumen uji tersebut sudah menjadi *gold standard*. Metode pengembangan sistem, peneliti memakai algoritma *brill tagger*. Algoritma *brill tagger* ini memiliki 2 tahap, yakni pemrosesan dengan pendekatan statistik menggunakan n-gram sebagai unit dan penggunaan aturan untuk memperbaiki dan disambiguasi luaran tahap pertama. Untuk metode evaluasi, peneliti menggunakan perhitungan presisi, recall dan akurasi.

3. Evaluasi dan Kesimpulan

Pada tahap ini peneliti akan menggunakan metode presisi dan akurasi sebagai metode evaluasi dari sistem yang dibuat oleh peneliti. Menurut Hagerman (2011) Presisi adalah tingkat ketepatan antara informasi yang diminta oleh pengguna

dengan jawaban yang diberikan oleh sistem. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan oleh penliti untuk mengerjakan laporan tugas akhir ini sesuai dengan ketentuan dari buku panduan penulisan tugas akhir. Peneliti akan dibagi menjadi 5 bab, yaitu:

Bab 1: Pendahuluan

Pada bab pertama, peneliti membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan yang dicapai peneliti dan metodologi penelitian yang digunakan oleh peneliti.

Bab 2: Tinjauan Pustaka

Pada bab kedua, peneliti menjelaskan mengenai tinjauan pustaka dan landasan teori. Pada tinjauan pustaka berisi beberapa penelitian yang serupa dan berkaitan dengan penelitian yang dilakukan oleh peneliti. Sedangkan landasan teori berisi dasar teori-teori yang digunakan dan mendukung penelitian yang dibuat oleh peneliti.

Bab 3 : Perancangan Sistem

Pada bab ketiga, peneliti menjelaskan mengenai perancangan sistem yang akan dibuat. Bab ini terdiri dari bagaimana tahapan-tahapan pembangunan sistem, rancangan data pelatihan, rancangan sistem, serta rancangan pengujian dari sistem.

Bab 4: Implementasi dan Analisis Sistem

Pada bab keempat, peneliti menjelaskan mengenai hasil implementasi yang telah dilakukan. Hasil yang akan didapatkan dapat berupa hasil penelitian

yang telah dilakukan oleh peneliti dan dari hasil tersebut, peneliti juga akan menganalisis hasil penelitian.

Bab 5: Kesimpulan

Pada bab kelima, peneliti menjelaskan mengenai kesimpulan dari hasil maupun analisa penelitian yang telah dibuat oleh peneliti dan saran. Kesimpulan berisi tentang hasil keseluruhan yang telah didapatkan selama penelitian berlangsung. Sedangkan saran dapat berupa sesuatu hal yang dapat dikembangkan dari penelitian yang telah dilakukan untuk penelitian yang akan mendatang.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis penelitian yang telah dibahas pada Bab 4, maka dapat ditarik kesimpulan sebagai berikut:

1. Untuk membuat *gold standart* sebagai bahan pembelajaran sistem dibutuhkan orang-orang yang sudah dilatih dalam memberikan label kelas kata Bahasa Indonesia sehingga tidak ada lagi ambiguitas dalam pemberian label kelas kata dan untuk penelitian kedepannya disarankan membuat suatu buku panduan untuk meningkatkan *gold standart*.
2. Untuk menyusun aturan transformasi berdasarkan kesalahan pelabelan otomatis adalah membuat dan menyusun aturan transformasi yang paling detail dan paling penting dalam suatu kelas kata terlebih dahulu. Sehingga dapat meminimalisir pencederaan sesama aturan transformasi.
3. Sistem sebagian besar sudah dapat melakukan pelabelan kelas kata yaitu kelas kata benda, kelas kata entitas benda, kelas kata kerja, kelas kata keterangan, kelas kata sifat, kelas kata tanda baca, kelas kata pengganti orang, kelas kata penghubung, dan kelas kata “url”. Pada Penelitian ini untuk cara mengukur presisi dan akurasi menggunakan metode *confusion matrix*. Sehingga pada pengujian sistem didapatkan rata-rata presisi sebesar 77,5% dan rata-rata akurasi sebesar 86%.
4. Sistem belum bisa memberikan label kelas kata kepemilikan benda dan masih perlu perbaikan dan penambahan aturan yang detail lagi untuk meningkatkan presisi dan akurasi sistem.
5. Penerapan algoritma *Brill Tagger* berbasis n-gram secara keseluruhan mendapatkan hasil yang cukup bagus dengan didasari pengujian sistem mendapatkan nilai rata-rata presisi sebesar 77,59% dan rata-rata akurasi sebesar 86,14%. Tetapi pada perhitungan probabilitas pada algoritma *Brill*

Tagger berbasis n-gram mendapat hasil yang kurang bagus dikarenakan menghasilkan label kelas kata yang sama untuk setiap kata sehingga pada proses pembuatan aturan untuk memerlukan sedikit bantuan leksikon dan memerlukan aturan dalam hal perbaikan hasil perhitungan probabilitas pada algoritma *Brill Tagger* berbasis n-gram.

5.2 Saran

Sistem ini sangat memungkinkan untuk dilakukan pengembangan lebih lanjut sesuai kebutuhan yang terus bertambah, sehingga dapat meningkatkan akurasi sistem. Saran yang diajukan penulis dalam pengembangan sistem kedepannya adalah sebagai berikut:

1. Memperbanyak data latih serta tema dari data latih tersebut.
2. Membuat buku panduan untuk adanya panduan dalam memberikan label kelas kata Bahasa Indonesia pada data latih supaya data latih bisa menjadi *gold data* yang sesungguhnya.

DAFTAR PUSTAKA

- Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for. *Computer Speech and Language*, 359-394.
- Christanti, V. M., Pragantha, J., Purnamasari, E. (2012). IMPLEMENTASI BRILL TAGGER UNTUK MEMBERIKAN POS-TAGGING PADA DOKUMEN BAHASA INDONESIA. *Jurnal Teknik dan Ilmu Komputer*, 1(3), 301-315.
- Elnathan, Y.K., Krisnawati. L.D., Widhiyanti, K., (2017). POS TAGGING BAHASA INDONESIA BERBASIS ATURAN DAN LEXICON (KATA KERJA + KATA KETERANGAN).
- Hagerman, C. (2011). *Evaluating the Performance of Automated Part-of-Speech Taggers on an L2 Corpus*. Jepang.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics*. New Jersey: Alan Apt.
- Karibun H.S., M. (2005). Penerapan Metode Rule-Based Dengan Unsupervised Learning Untuk Pelabelan Dokumen Berbahasa Indonesia. Bogor, Jawa Barat, Indonesia.
- Kristanto, Y.A., Krisnawati. L.D., Widhiyanti, K., (2017). POS – TAGGING BAHASA INDONESIA BERBASIS ATURAN DAN LEKSIKON UNTUK KELAS KATA BENDA DAN ADJEKTIVA
- Kridalaksana, H. (2005). *Kelas Kata dalam Bahasa Indonesia*. Jakarta: PT Gramedia Pustaka Utama.
- Ngai, G., & Florian, R. (2001). *Transformation-Based Learning in the Fast Lane*. Maryland: Baltimore University.

- Pisceldo, F., Adriani, M., & Marunung, R. (2009). *Probalistic Part Of Speech Tagging for Bahasa Indonesia*. Jakarta: Universitas Indonesia.
- Rostianingsih, S., Sugianto, S.A, & Liliana. (2014). *Aplikasi Predictive Text Berbahasa Indonesia Dengan Metode N-Gram*. Surabaya: Universitas Kristen Petra.
- Setyaningsih, E. R. (2017). Part Of Speech Tagger Untuk Bahasa Indonesia Dengan Menggunakan Modifikasi Brill. *Dinamika Teknologi*, 37-42.
- Stratos, K., & Collins, M. (2015). *Simple Semi-Supervised POS Tagging*. New York: Columbia University.
- Trenkle, J. M., & Cavnar, W. B. (1994). *N-Gram-Based Text Categorization*. Di dalam: Lewis D, editor. *Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval*; Las Vegas, 11-13 April 1994. Nevada: UNLV Publications/Reprographics. 161-175.
- Widhiyanti, K., & Harjoko, A. (2012). POS Tagging Bahasa Indonesia Dengan HMM dan Rule Based. *INFORMATIKA*, 8(2), 151-167.