

Pengaruh *Brill Tagger* Terhadap Hasil Klasifikasi Analisis Sentimen Menggunakan Algoritma *Multinomial Naïve Bayes*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

R. Astero Nandito
NIM : 09021381621095

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021**

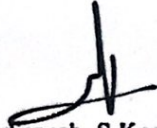
LEMBAR PENGESAHAN TUGAS AKHIR

PENGARUH *BRILL TAGGER* TERHADAP HASIL KLASIFIKASI ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *MULTINOMIAL NAÏVE BAYES*

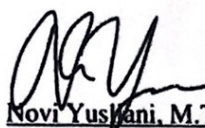
Oleh:

R. Astero Nandito
NIM : 09021381621095

Pembimbing I,


Dr. Abdiansah, S.Kom, M.Cs
NIP. 196602202006041001

Palembang, Januari 2021
Pembimbing II,


Novi Yushani, M.T
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,


Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari senin tanggal 4 Januari 2021 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : R. Astero Nandito
NIM : 09021381621095
Judul : Pengaruh *Brill Tagger* Terhadap Hasil Klasifikasi Analisis Sentimen Menggunakan Algoritma *Multinomial Naive Bayes*

1. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005



2. Pembimbing II

Novi Yusliani, M.T.
NIP. 198211082012122001




3. Penguji I

Yunita, M.Cs.
NIP. 198306062015042002



4. Penguji II

Kanda Januar Miraswan, M.T.
NIP. 199001092019031012



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : R. Astero Nandito
NIM : 09021381621095
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh *Brill Tagger* Terhadap Hasil Klasifikasi Analisis Sentimen Menggunakan Algoritma *Multinomial Naïve Bayes*

Hasil Pengecekan Software *iThenticate/Turnitin* : 11%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Januari 2021



R. Astero Nandito
NIM. 09021381621095

MOTTO DAN PERSEMBAHAN

“You can’t control luck, you can’t control timing, you can’t control what other people say and do. But you can always control how hard you work.”

Kupersembahkan karya tulis ini kepada :

- *Kedua orangtua*
- *Keluarga besar*
- *Teman - teman Seperjuangan*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

THE EFFECT OF BRILL TAGGER ON THE CLASSIFICATION RESULTS OF SENTIMENT ANALYSIS USING MULTINOMIAL NAIVE BAYES ALGORITHM

By:

R. Astero Nandito
09021381621095

ABSTRACT

Twitter is a social media that is often used by researchers as an object of research to conduct sentiment analysis. Twitter is also a good indicator for influence in research, the problem that arises in research in the field of sentiment analysis is the large number of factors such as the use of informal or colloquial language and other factors that can affect the results of sentiment classification. To improve the results of sentiment classification, an information extraction process can be carried out. One part of the information extraction feature is a part of speech tagging, which is the giving of word classes automatically. The results of part of speech tagging are used for weighting words based on part of speech. This study examines the effect of Part of Speech Tagging with the method Brill Tagger in sentiment analysis using the Naive Bayes Multinomial algorithm. Testing were carried out on 500 twitter tweet texts and obtained the results of the sentiment classification with implementing part of speech tagging precision 73,2%, recall 63,2%, f-measure 67,6%, accuracy 60,7% and without implementing part of speech tagging precision 65,2%, recall 60,6%, f-measure 62,4% accuracy 53,3%. From the results of the accuracy obtained, it shows that the application of part of speech tagging in sentiment analysis using the Multinomial Naïve Bayes algorithm has an effect with an increase in classification performance.

Keywords: Sentiment Analysis, Part of Speech Tagging, Brill Tagger, Multinomial Naive Bayes, Weighting Part of Speech

PENGARUH *BRILL TAGGER* TERHADAP HASIL KLASIFIKASI
ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA
MULTINOMIAL NAIVE BAYES

Oleh:

R. Astero Nandito
09021381621095

ABSTRAK

Twitter adalah media sosial yang sering digunakan oleh peneliti sebagai objek riset untuk melakukan analisis sentimen. Twitter juga merupakan indikator yang baik untuk memberi pengaruh dalam penelitian, masalah yang sering muncul dalam penelitian di bidang analisis sentimen adalah banyaknya faktor-faktor seperti penggunaan bahasa yang tidak formal atau bahasa sehari-hari dan faktor lainnya yang dapat mempengaruhi hasil klasifikasi sentimen. Untuk meningkatkan hasil klasifikasi sentimen, dapat dilakukan proses ekstraksi informasi. Salah satu bagian dari fitur ekstraksi informasi adalah part of speech tagging, yaitu pemberian kelas kata secara otomatis. Hasil part of speech tagging digunakan untuk pembobotan kata berdasarkan part of speech. Penelitian ini menguji pengaruh part of speech tagging dengan Brill Tagger dalam analisis sentimen menggunakan algoritma Multinomial Naive Bayes. Pengujian dilakukan pada 500 teks kicauan twitter dan didapatkan hasil klasifikasi sentimen dengan menerapkan part of speech tagging presisi 73,2%, recall 63,2%, f-measure 67,6%, akurasi 60,7% dan tanpa menerapkan part of speech tagging presisi 65,2%, recall 60,6%, f-measure 62,4%, akurasi 53,3%. Dari hasil klasifikasi analisis sentimen yang didapatkan menunjukkan bahwa penerapan part of speech tagging dalam analisis sentimen menggunakan algoritma Multinomial Naive Bayes memberikan pengaruh dengan adanya peningkatan kinerja klasifikasi.

Kata Kunci: Analisis Sentimen, Part of Speech Tagging, Brill Tagger, Multinomial Naive Bayes, Pembobotan Part of Speech.

KATA PENGANTAR

Selama penelitian dan penyusunan laporan penelitian skripsi ini, penulis tidak luput dari kendala dan hambatan. Namun demikian kendala dan hambatan tersebut dapat penulis atasi berkat bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini, penulis ingin menyampaikan rasa terimakasih sebesar-besarnya kepada:

1. Kedua orang tua saya dan keluarga yang telah memberikan dukungan baik moril maupun materil serta doa yang tiada henti-hentinya kepada penulis.
2. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku dosen pembimbing I dan Ibu Novi Yusliani, M.T. selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses pengerjaan Tugas Akhir, hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
3. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Ibu Yunita, M.Cs. selaku dosen penguji I, dan Bapak Kanda Januar Miraswan, M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
5. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Mbak Wiwin dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.

7. Erin dan Melvin, yang telah banyak membantu, memberikan dukungan, informasi dan referensi selama proses pengerjaan Tugas Akhir hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
8. Adi, Irfan, Luthfi, Ridho, Ahmad serta seluruh teman jurusan Teknik Informatika yang tidak dapat disebutkan satu persatu yang telah berbagi keluh kesah, motivasi, semangat, dan hiburan selama masa perkuliahan.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Januari 2021

R. Astero Nandito

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.7.1 Bab I. Pendahuluan	I-5
1.7.2 Bab II. Kajian Literatur	I-6
1.7.3 Bab III. Metodologi Penelitian.....	I-6
1.7.4 Bab IV. Pengembangan Perangkat Lunak	I-6
1.7.5 Bab V. Hasil dan Analisis Penelitian	I-6
1.7.6 Bab VI. Kesimpulan dan Saran.....	I-7
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1

2.2.1	Analisis Sentimen	II-1
2.2.2	Pra-Pengolahan	II-2
2.2.3	<i>POS-Tagging: Brill Tagger</i>	II-3
2.2.4	Pembobotan TF-IDF	II-8
2.2.5	<i>Naïve Bayes Classifier: Multinomial Naïve Bayes</i>	II-10
2.2.6	<i>K-Fold Cross Validation</i>	II-13
2.2.7	<i>Confusion Matrix</i>	II-13
2.2.8	<i>Rational Unified Process</i>	II-16
2.3	Penelitian Lain Yang Relevan	II-17
2.4	Kesimpulan.....	II-19
BAB III METODOLOGI PENELITIAN.....		III-1
2.1	Pendahuluan	III-1
2.2	Pengumpulan Data	III-1
2.2.1	Jenis dan Sumber Data	III-1
2.2.2	Metode Pengumpulan Data.....	III-1
2.3	Tahapan Penelitian	III-2
2.3.1	Menetapkan Kerangka Kerja / Framework	III-2
2.3.2	Menetapkan Kriteria Pengujian	III-7
2.3.3	Menetapkan Format Data Pengujian.....	III-8
2.3.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-9
2.3.5	Melakukan Pengujian Penelitian.....	III-9
2.3.6	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-10
2.4	Metode Pengembangan Perangkat Lunak	III-11
2.4.1	Fase Insepsi	III-11
2.4.2	Fase Elaborasi	III-12
2.4.3	Fase Kontruksi	III-12
2.4.4	Fase Transisi.....	III-13
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-3

4.2.3	Analisis dan Desain.....	IV-5
4.3	Fase Elaborasi.....	IV-34
4.3.1	Pemodelan Bisnis.....	IV-34
4.3.2	Perancangan Data.....	IV-34
4.3.3	Perancangan Antarmuka.....	IV-35
4.3.4	Kebutuhan Sistem.....	IV-35
4.3.5	Diagram Aktivitas.....	IV-36
4.3.6	Diagram Alur.....	IV-40
4.4	Fase Kontruksi.....	IV-46
4.4.1	Kebutuhan Sistem.....	IV-47
4.4.2	Diagram Kelas.....	IV-47
4.4.3	Implementasi.....	IV-49
4.5	Fase Transisi.....	IV-50
4.5.1	Pemodelan Bisnis.....	IV-51
4.5.2	Kebutuhan Sistem.....	IV-51
4.5.3	Rencana Pengujian.....	IV-51
4.5.4	Implementasi.....	IV-53
4.6	Kesimpulan.....	IV-57
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan.....	V-1
5.2	Hasil Penelitian.....	V-1
5.2.1	Hasil Luaran Program.....	V-1
5.2.2	Konfigurasi Percobaan.....	V-3
5.2.3	Data Hasil Konfigurasi.....	V-4
5.3	Analisis Hasil Penelitian.....	V-6
5.4	Kesimpulan.....	V-10
BAB VI KESIMPULAN DAN SARAN.....		VI-1
6.1	Pendahuluan.....	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-3
DAFTAR PUSTAKA.....		xvii
LAMPIRAN.....		L-1

DAFTAR TABEL

	Halaman
II-1. Tabel Tagset Bahasa Indonesia yang Akan Digunakan	II-3
II-2. <i>Confusion Matrix</i> Kelas Positif dan Kelas Negatif	II-15
III-1. Tabel Bobot <i>Part of Speech</i>	III-7
III-2. Rancangan Tabel <i>Confusion Matrix</i> Hasil Klasifikasi	III-8
III-3. Rancangan Tabel Hasil Pengujian Klasifikasi	III-11
IV-1. Kebutuhan Fungsional.....	IV-4
IV-2. Kebutuhan Non-Fungsional	IV-4
IV-3. Contoh Data Kicauan	IV-9
IV-4. Hasil Penghapusan <i>Unicode</i> dari Contoh Data Kicauan	IV-10
IV-5. Hasil Perubahan Kata Tidak Baku Menjadi Kata Baku	IV-11
IV-6. Hasil <i>Casefolding</i> dari Contoh Data Kicauan	IV-12
IV-7. Hasil <i>Tokenizing</i> dari Contoh Data Kicauan	IV-12
IV-8. Contoh Data <i>Training POS-Tagging</i>	IV-13
IV-9. Hasil <i>POS-Tagging</i> dari Contoh Data Kicauan.....	IV-14
IV-10. Hasil <i>Stopword Removal</i> tanpa Proses <i>POS-Tagging</i>	IV-16
IV-11. Hasil <i>Stopword Removal</i> dengan Proses <i>POS-Tagging</i>	IV-16
IV-12. Hasil <i>Stemming</i> tanpa Proses <i>POS-Tagging</i>	IV-17
IV-13. Hasil <i>Stemming</i> dengan Proses <i>POS-Tagging</i>	IV-17
IV-14. Hasil TF dan IDF dari Contoh Data Kicauan	IV-18
IV-15. Hasil Pembobotan Kata dari Contoh Data Kicauan	IV-19
IV-16. Hasil TF dan IDF Berdasarkan <i>Part of Speech</i>	IV-20
IV-17. Hasil Pembobotan Kata Berdasarkan <i>Part of Speech</i>	IV-20
IV-18. Hasil Perhitungan Probabilitas Data <i>Training</i> (tanpa <i>POS-Tagging</i>).....	IV-23
IV-19. Hasil Perhitungan Probabilitas Data <i>Training</i> (dengan <i>POS-Tagging</i>).....	IV-23
IV-20. Contoh Data <i>Testing</i>	IV-24
IV-21. Hasil Pra-Pengolahan Data <i>Testing</i> tanpa <i>POS-Tagging</i>	IV-24

IV-22. Hasil Pra-Pengolahan Data <i>Testing</i> dengan <i>POS-Tagging</i>	IV-25
IV-23. Klasifikasi Data <i>Testing</i> tanpa <i>POS-Tagging</i>	IV-26
IV-24. Klasifikasi Data <i>Testing</i> dengan <i>POS-Tagging</i>	IV-26
IV-25. Contoh <i>Confusion Matrix</i>	IV-26
IV-26. Definsi Aktor.....	IV-28
IV-27. Definsi <i>Use Case</i>	IV-29
IV-28. Skenario <i>Use Case</i> Melakukan <i>Load Data</i>	IV-30
IV-29. Skenario <i>Use Case</i> Melakukan Klasifikasi <i>Multinomial Naïve</i> <i>Bayes</i>	IV-31
IV-30. Skenario <i>Use Case</i> Melakukan Klasifikasi <i>Multinomial Naïve</i> <i>Bayes</i> dan <i>POS-Tagging</i>	IV-33
IV-31. Tabel Implementasi Kelas	IV-49
IV-32. Rencana Pengujian <i>Use Case</i> Melakukan <i>Load Data</i>	IV-52
IV-33. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>Multinomial</i> <i>Naïve Bayes</i>	IV-52
IV-34. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>Multinomial</i> <i>Naïve Bayes</i> dan <i>POS-Tagging</i>	IV-52
IV-35. Pengujian <i>Use Case</i> Melakukan <i>Load Data</i>	IV-53
IV-36. Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>Multinomial Naïve</i> <i>Bayes</i>	IV-55
IV-37. Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>Multinomial Naïve</i> <i>Bayes</i> dan <i>POS-Tagging</i>	IV-56
V-1. Contoh Hasil <i>Tagging</i> Teks Kategori Bagus.....	V-2
V-2. Contoh Hasil <i>Tagging</i> Teks Kategori Agak Bagus	V-2
V-3. Contoh Hasil <i>Tagging</i> Teks Kategori Buruk.....	V-3
V-4. Tabel <i>Confusion Matrix</i> Hasil Pengujian Klasifikasi.....	V-4
V-5. Tabel Hasil Pengujian Klasifikasi	V-5

DAFTAR GAMBAR

	Halaman
II-1. Klasifikasi Beberapa Model <i>POS-Tagging</i> yang Berbeda.....	II-5
II-2. Proses Kerja <i>Brill Tagger</i>	II-6
II-3. Skema 10 <i>Fold Cross Validation</i>	II-15
II-3. Arsitektur <i>Rational Unified Process</i>	II-17
III-1. Diagram Tahapan Proses dengan <i>POS-Tagging</i>	III-3
III-2. Diagram Tahapan Proses tanpa <i>POS-Tagging</i>	III-4
IV-1. Contoh Perhitungan Probabilitas Tertinggi.....	IV-25
IV-2. Contoh Perhitungan <i>Precision, Recall, F-Measure</i> dan <i>Accuracy</i>	IV-27
IV-3. Diagram <i>Use Case</i>	IV-28
IV-4. Rancangan Antarmuka Perangkat Lunak	IV-35
IV-5. Diagram Aktivitas <i>Load Data</i>	IV-37
IV-6. Diagram Aktivitas Klasifikasi <i>Multinomial Naïve Bayes</i> tanpa <i>POS-Tagging</i>	IV-38
IV-7. Diagram Aktivitas Klasifikasi <i>Multinomial Naïve Bayes</i> dengan <i>POS-Tagging</i>	IV-39
IV-8. Diagram Alur <i>Load Data</i>	IV-40
IV-9. Diagram Alur <i>Multinomial Naïve Bayes</i> tanpa <i>POS-Tagging</i>	IV-41
IV-10. Diagram Alur <i>Multinomial Naïve Bayes</i> dengan <i>POS-Tagging</i>	IV-42
IV-11. Diagram Alur Pra_Pengolahan_withoutPOS (<i>Subsequence</i>).....	IV-43
IV-12. Diagram Alur Pra_Pengolahan_Testing (<i>Subsequence</i>)	IV-44
IV-13. Diagram Alur Pra_Pengolahan_withPOS1 (<i>Subsequence</i>).....	IV-45
IV-14. Diagram Alur Pra_Pengolahan_withPOS2 (<i>Subsequence</i>).....	IV-46
IV-15. Diagram Kelas Perangkat Lunak.....	IV-48
IV-16. Antarmuka Perangkat Lunak.....	IV-50
V-1. Perbandingan Hasil Evaluasi Klasifikasi	V-6

DAFTAR LAMPIRAN

	Halaman
1. 500 Data Kicauan Twitter Tentang LRT Palembang.....	L-1
2. Kode Program	L-56

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini memberikan penjelasan umum tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah yang menjadi dasar dalam penelitian ini.

1.2 Latar Belakang Masalah

Twitter adalah media sosial yang sering digunakan oleh peneliti sebagai objek riset untuk melakukan analisis sentimen. Twitter juga merupakan indikator yang baik untuk memberi pengaruh dalam penelitian, namun masih belum banyak aplikasi dan metode analisa sentimen yang dikembangkan untuk bahasa Indonesia (Winarko dan Putrianti, 2014). Masalah yang sering muncul dalam penelitian di bidang analisis sentimen adalah banyaknya faktor-faktor seperti penggunaan bahasa yang tidak formal atau bahasa sehari-hari dan faktor lainnya yang dapat mempengaruhi hasil klasifikasi sentimen.

Hal yang bisa dilakukan untuk meningkatkan hasil klasifikasi sentimen adalah dengan melakukan ekstraksi informasi. Salah satu fitur ekstraksi informasi adalah *POS-Tagging*. *POS-Tagging* merupakan salah satu proses umum yang biasanya diperlukan oleh sistem-sistem yang berbasis bahasa alami. *POS-Tagging* bertujuan untuk menentukan kelas kata pada setiap kata yang ada dalam suatu

kalimat. Pemberian label atau kelas kata secara manual akan memakan waktu yang lama dan membutuhkan sumberdaya yang tidak sedikit. Muljono dkk., (2017) menerapkan *POS-Tagging* dalam sebuah sistem yang dapat menentukan kelas kata dari sebuah kalimat bahasa Indonesia secara otomatis. Sistem ini memudahkan dalam mengetahui kelas kata yang ada di dalam suatu kalimat. Selanjutnya Hamzah dan Widyastuti (2015) mengenai pengaruh *POS-Tagging* dalam melakukan identifikasi sebuah dokumen apakah dokumen tersebut mengandung opini atau tidak dengan metode *Hidden Markov Model* Berbagai pendekatan untuk proses *tagging* telah banyak dikembangkan. Beberapa diantaranya menggunakan perhitungan probabilistik, statistika, dan berbasis aturan.

Penelitian *POS-Tagging* untuk bahasa Indonesia telah banyak dilakukan seperti penelitian Aprillia (2015) mengenai pengaruh *POS-Tagging* terhadap klasifikasi analisis sentimen. Penelitian tersebut menunjukkan bahwa pengaruh *POS-Tagging* dalam hasil klasifikasi lebih tinggi dibandingkan tanpa menerapkan *POS-Tagging* dengan akurasi 62,4% sedangkan tanpa *POS-Tagging* akurasi yang didapatkan sebesar 52,4%. Penelitian Amrullah dkk., (2017) melakukan perbandingan metode *POS-Tagging* menggunakan *Hidden Markov Model*, dan *Brill Tagger* pada teks berbahasa Indonesia. Penelitian tersebut menunjukkan bahwa akurasi yang didapatkan dengan menggunakan metode *Brill Tagger* lebih baik dibandingkan dengan menggunakan metode HMM dengan menggunakan 30.000 kalimat. Akurasi yang didapatkan *Brill Tagger* adalah 76,78% sedangkan HMM 62,69%. Meskipun demikian, belum ada penelitian sebelumnya yang membahas *POS-Tagging* menggunakan metode *Brill Tagger* untuk kasus

klasifikasi analisis sentimen berbahasa Indonesia. *Brill Tagger* adalah salah satu metode berbasis *rule*. Pendekatan berbasis *rule* dipilih dengan pertimbangan tingginya tingkat akurasi yang berhasil dicapai dengan menggunakan pendekatan berbasis *rule* pada proses *tagging* (Setyaningsih, 2017).

Pada proses analisis sentimen dibutuhkan sebuah metode klasifikasi, salah satu metode klasifikasi yang sering digunakan adalah *Naïve Bayes*. *Naïve Bayes* cenderung bekerja dengan sangat baik sehingga menghasilkan akurasi yang cukup baik dan biasanya waktu pelatihannya lebih cepat dibandingkan dengan model seperti *Support Vector Machine* (Vivek, Arora dan Bhatia 2013). *Naïve Bayes* memiliki dua *event model*, yaitu *Multi-Variate Bernouli* dan *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* dipilih untuk melakukan klasifikasi analisis sentimen karena kinerjanya lebih baik dibandingkan *Multi-Variate Bernouli* pada kasus jumlah kata atau *term* yang banyak. *Multinomial Naïve Bayes* dapat menurunkan kesalahan pada klasifikasi dengan nilai rata-rata 27% bahkan 50% dari percobaan terhadap *Multi-Variate Bernouli* (McCallum dan Nigam, 1998). Pada penelitian ini akan menguji pengaruh *Brill Tagger* terhadap hasil akurasi klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes*.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan diatas, rumusan masalah dalam penelitian ini adalah bagaimana pengaruh *POS-Tagging* menggunakan metode *Brill Tagger* terhadap hasil klasifikasi analisis sentimen menggunakan algoritma *Multinomial Naïve Bayes*. Ada beberapa pertanyaan yang akan dijawab pada usulan penelitian ini, yaitu:

1. Bagaimana mekanisme metode *Brill Tagger* dalam melakukan *POS-Tagging* kata-kata dalam bahasa Indonesia?
2. Bagaimana pengaruh metode *Brill Tagger* terhadap hasil analisis sentimen menggunakan algoritma *Multinomial Naïve Bayes*?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah penelitian di atas maka ada dua tujuan penelitian ini, yaitu :

1. Mengetahui mekanisme metode *Brill Tagger* dalam melakukan *POS-Tagging* kata-kata bahasa Indonesia.
2. Mengetahui pengaruh metode *Brill Tagger* terhadap hasil analisis sentimen menggunakan algoritma *Multinomial Naïve Bayes*.

1.5 Manfaat Penelitian

Adapun manfaat yang di harapkan dari usulan penelitian ini adalah sebagai berikut :

1. Bisa menjadi acuan atau referensi bagi penelitian-penelitian lain di bidang *Brill Tagger*.
2. Dapat digunakan dalam mengetahui polaritas dari sebuah sentimen.

1.6 Batasan Masalah

Untuk membatasi penelitian supaya tidak keluar dari topik bahasan, maka diberikan beberapa batasan masalah penelitian sebagai berikut :

1. Data teks yang diambil berupa teks berbahasa Indonesia, karena kaidah pada pra-pengolahan diatur dalam tata bahasa Indonesia.
2. Data yang digunakan adalah data twitter mengenai LRT Palembang Sumsel.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain:

1.7.1 Bab I. Pendahuluan

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, dan sistematika penulisan.

1.7.2 Bab II. Kajian Literatur

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, antara lain penelitian terkait, definisi-definisi Analisis Sentimen, *Preprocessing*, *POS-Tagging*, *Brill Tagger*, *Algoritma Multinomial Naïve Bayes*, *K-Fold Cross Validation*, *Confusion Matrix*, dan *Rational Unified Process*.

1.7.3 Bab III. Metodologi Penelitian

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

1.7.4 Bab IV. Pengembangan Perangkat Lunak

Pada bab ini akan membahas tahapan-tahapan yang dilakukan dalam proses pengembangan perangkat lunak berdasarkan metode *Rational Unified Process (RUP)*.

1.7.5 Bab V. Hasil dan Analisis Penelitian

Pada bab ini akan membahas hasil pengujian berdasarkan rencana pengujian yang telah ditentukan. Analisis hasil penelitian akan diberikan sebagai dasar dari kesimpulan yang diambil pada penelitian ini.

1.7.6 Bab VI. Kesimpulan dan Saran

Pada bab ini berisikan kesimpulan dan saran dari penelitian ini. Kesimpulan dan saran diambil dari semua pembahasan yang dilakukan pada bab sebelumnya..

1.8 Kesimpulan

Penelitian ini berfokus untuk melihat pengaruh *POS-Tagging* terhadap hasil klasifikasi analisis sentimen pada twitter menggunakan algoritma *Multinomial Naïve Bayes*. Metode *POS-Tagging* yang digunakan adalah *Brill Tagger*.

DAFTAR PUSTAKA

- Amrullah, A. Z., Hartanto, R., & Mustika, I. W. (2017). A comparison of different part-of-speech tagging technique for text in Bahasa Indonesia. *Proceedings - 2017 7th International Annual Engineering Seminar, InAES 2017*.
- Aprillia (2018). “Pengaruh Part Of Speech Tagging dalam Analisis Sentimen menggunakan Algoritma Multinomial Naïve Bayes.” Skripsi Jurusan Teknik Informatika, Fak. Ilmu Komputer, UNSRI – Tidak diterbitkan.
- Blum, A., & Mitchell, T. (1998). *Combining Labeled and Unlabeled Data with Co-Training*. 92–100.
- Bobby, N., Mirna, A., Jelita, A., & Tahaghoghi, S. M. M. (2007). Stemming Indonesian: A Confix-Stripping Approach. *Conferences in Research and Practice in Information Technology Series*, 38(September 2018), 307–314.
- Dinakaramani, A., Rashel, F., Luthfi, A., & Manurung, R. (2014). Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus. *Proceedings of the International Conference on Asian Language Processing 2014, IALP 2014*, 66–69. <https://doi.org/10.1109/IALP.2014.6973519>
- Hamzah, A., & Widyastuti, N. (2015). Document subjectivity and target detection in opinion mining using HMM POS-Tagger. *Proceedings of 2015 International Conference on Information and Communication Technology and Systems, ICTS 2015, September 2015*, 83–87.

- Handayani, F., & Pribadi, F. S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, 7(1), 19–24.
- Jurafsky, D., & Martin, J. H. (2000). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Daniel. *Journal of Orthodontics*, 39(2), 139.
- Kruchten, P., & Kroll, P. (2003). The Rational Unified Process Made Easy. In *Rational Unified Process Made Easy: A Practitioner's Guide to the RUP*.
- Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing, Second Edition*, 2, 568.
- McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. 58(8), 582–591. <https://doi.org/10.1002/em.22125>
- Megyesi, B. (1992). *Brill's rule-based PoS tagger. section 3*, 1–10.
- Muljono, Afini, U., Supriyanto, C., & Nugroho, R. A. (2017). The development of Indonesian *POS-Tagging* system for computer-aided independent language learning. *International Journal of Emerging Technologies in Learning*,
- Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8206 LNCS, 194–201.

- Prasetyo, H., Adiwijaya, & Astuti, W. (2019). *Klasifikasi Multi -Label pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia Menggunakan Mutual Information dan Backpropagation Neural Network*. 6(2), 9086–9098.
- Setyaningsih, E. R. (2017). Part of Speech Tagger Untuk Bahasa Indonesia Dengan Menggunakan Modifikasi Brill. *Dinamika Teknologi*, 9(1), 37–42.
- Song, J., Kim, K. T., Lee, B., Kim, S., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. *KSI Transactions on Internet and Information Systems*, 11(6), 2996–3011.
- Winarko, E., & Putrianti, N. D. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91–
- Wu, Wei., Zhang, Bin., Ostendorf, Mari (2010). Automatic generation of personalized annotation tags for Twitter users. *ACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*
- Xu, R. (2014). POS weighted TF-IDF algorithm and its application for an MOOC search engine. *ICALIP 2014 - 2014 International Conference on Audio, Language and Image Processing, Proceedings*, 868–873.