

**IMPLEMENTASI SENTIMEN ANALISIS PADA CALON PRESIDEN
INDONESIA 2019 BERDASARKAN DATA MEDIA SOSIAL
TWITTER MENGGUNAKAN METODE LEXICON
BASED DAN NAIVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Aditya Derry Saputra

15.11.8974

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**



**IMPLEMENTASI SENTIMEN ANALISIS PADA CALON PRESIDEN
INDONESIA 2019 BERDASARKAN DATA MEDIA SOSIAL
TWITTER MENGGUNAKAN METODE LEXICON
BASED DAN NAIVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Aditya Derry Saputra

15.11.8974

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI SENTIMEN ANALISIS PADA CALON PRESIDEN
INDONESIA 2019 BERDASARKAN DATA MEDIA SOSIAL
TWITTER MENGGUNAKAN METODE LEXICON
BASED DAN NAIVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Aditya Derry Saputra

15.11.8974

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 April 2019

Dosen Pembimbing,



Heri Sismoro, M.Kom

NIK. 190302057

PENGESAHAN

SKRIPSI

IMPLEMENTASI SENTIMEN ANALISIS PADA CALON PRESIDEN INDONESIA 2019 BERDASARKAN DATA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE LEXICON BASED DAN NAIVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Aditya Derry Saputra

15.11.8974

telah dipertahankan di depan Dewan Penguji
pada tanggal 11 April 2019

Susunan Dewan Penguji

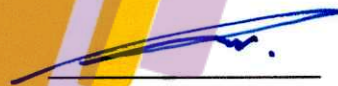
Nama Penguji

Tanda Tangan

Tonny Hidayat, M.Kom
NIK. 190302182



Bambang Sudaryatno, Drs., M.M
NIK. 190302029



Bety Wulan Sari, M.Kom
NIK. 190302254



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 April 2019

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 18 April 2019



Aditya Derry Saputra

NIM. 15.11.8974

MOTTO

”Hidup Mulia atau Mati Syahid”

“Barangsiapa bertakwa kepada Allah niscaya Dia akan mengadakan baginya jalan keluar. Dan memberinya rezki dari arah yang tiada disangka – sangkanya.”

(QS. Ath Tholaq: 2-3)



PERSEMBAHAN

Alhamdulillah, segala puji bagi Allah, Rabb semesta alam yang telah memberikan segala kenikmatan yang tidak bisa hitung jumlahnya. Shalawat serta salam semoga senantiasa ditujukan bagi Rasulullah, keluarga, para sahabat, dan siapa saja yang meneladani mereka dengan baik hingga hari kiamat.

Terimakasih kepada orang tua saya tercinta, Mawardi dan Tuti yang telah mendidik saya, mendukung saya dalam melangkah, menemani saya setiap saat, dan setiap hari mendo'akan saya hingga saya mampu untuk menyelesaikan segala kewajiban saya di bangku pendidikan. Kepada dua adik saya, Yoga Adiwijaya dan Arif Wahyu Widodo, terimakasih atas segala dukungan dan motivasinya, semoga keluarga kita kelak dikumpulkan didalam surga – Nya.

Terimakasih kepada dosen – dosen yang telah sabar dan ikhlas dalam mendidik saya hingga mampu melewati seluruh ujian mata kuliah, terutama kepada Bapak Heri Sisworo, M.Kom semoga ilmu yang beliau berikan bermanfaat bagi seluruh mahasiswa khususnya Informatika angkatan 2015 Universitas Amikom Yogyakarta.

Terimakasih kepada keluargaku di Yogyakarta, semoga kelak di akhirat kita termasuk golongan orang yang mendapat naungan Allah dimana tidak ada naungan kecuali naungan – Nya.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Segala puji bagi Allah, Rabb semesta alam. Shalawat dan salam semoga senantiasa ditujukan bagi Rasulullah SAW, keluarga, para sahabat, dan siapa saja yang meneladani mereka dengan baik hingga hari kiamat.

Dalam menyelesaikan skripsi ini, banyak pihak yang telah memberikan bantuan baik secara moril, nasihat, dan semangat maupun materil. Atas segala bantuan yang telah diberikan, penulis ingin menyampaikan doa dan ucapan terimakasih yang sedalam – dalamnya kepada :

1. Bapak Prof. Suyanto selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Sudarmawan, S.Si, MT selaku dekan Fakultas Sains dan Teknologi dan ketua program studi S1 Informatika Universitas Amikom Yogyakarta beserta seluruh staf.
3. Bapak Heri Sismoro, M.Kom selaku dosen pembimbing penulis yang telah meluangkan waktu untuk membimbing, mengarahkan dan memberi masukan kepada penulis dalam pengerjaan skripsi ini hingga akhir.
4. Seluruh Dosen, Laboran, dan Staff Administrasi Jurusan Informatika Universitas Amikom Yogyakarta, terima kasih atas segala ilmu dan bimbingannya.

5. Segenap dosen teknik informatika yang telah memberikan bimbingan keilmuan kepada penulis selama masa studi Seluruh rekan-rekan studi yang tidak dapat disebutkan satu persatu, terima kasih atas segala kebaikan yang diberikan kepada penulis.

Berbagai kekurangan dan kesalahan mungkin pembaca temukan dalam penulisan skripsi ini, untuk itu penulis menerima segala kritik dan saran yang membangun dari pembaca sekalian. Semoga apa yang menjadi kekurangan bisa disempurnakan oleh peneliti selanjutnya dan semoga karya ini senantiasa dapat memberi manfaat.

Wassalamualikum Wr.Wb.

Yogyakarta, 18 April 2019

Penulis,

Aditya Derry Saputra

NIM. 15.11.8974

DAFTAR ISI

JUDUL.....	i
PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN.....	iv
MOTTO.....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian.....	4
1.7 Sistematika Laporan.....	6
BAB II LANDASAN TEORI.....	7
2.1 Tinjauan Pustaka.....	7
2.2 Analisa Sentimen.....	9
2.3 <i>Text Mining</i>	10
2.3.1 <i>Text Preprocessing</i>	11
2.3.2 <i>Feature Selection</i>	11
2.4 Twitter.....	13
2.5 Morfologi.....	15
2.5.1 Proses Morfologi.....	16

2.5.1.1	Afiksasi.....	16
2.5.1.2	Awalan (Prefiks).....	17
2.5.1.3	Sisipan (Infiks).....	23
2.5.1.4	Akhiran (Sufiks).....	23
2.5.1.5	Konfiks.....	24
2.6	<i>Lexicon Based</i>	25
2.7	<i>Naive Bayes Classifier</i>	25
2.8	<i>Z-Score</i>	28
BAB III ANALISA DAN PERANCANGAN SISTEM		29
3.1	Deskripsi Umum.....	29
3.2	Data Tweet.....	30
3.2.1	Tabel <i>Tweet Training</i>	30
3.2.2	Tabel <i>Tweet Testing</i>	31
3.2.3	Data <i>Stopword</i>	32
3.3	<i>Preprocessing</i>	32
3.3.1	<i>Case Folding</i>	33
3.3.2	<i>Cleansing</i>	35
3.3.3	<i>Normalization</i>	36
3.3.4	<i>Convert Negation</i>	38
3.3.5	<i>Tokenizing</i>	39
3.3.6	<i>Stopword Removal</i>	41
3.3.7	<i>Stemming</i>	42
3.3.8	<i>Part of Speech Tagging</i>	45
3.3.9	Metode <i>Lexicon Based</i>	47
3.4	<i>Naive Bayes Classifier</i>	48
3.4.1	Data <i>Training</i>	50
3.4.2	Data <i>Testing</i>	53
3.5	Perancangan Desain Sistem.....	56
3.6	Perancangan Antarmuka.....	57
3.6.1	Antarmuka Halaman Utama.....	57
3.6.2	Antarmuka Proses <i>Crawling</i>	58

3.6.3	Antarmuka Proses <i>preprocessing</i>	58
3.6.4	Antarmuka Proses Seleksi Fitur <i>Lexicon Based</i>	59
3.6.5	Antarmuka Proses Bobot <i>Naive Bayes</i>	59
BAB IV HASIL DAN PEMBAHASAN		60
4.1	Implementasi.....	60
4.1.1	Pengumpulan Data <i>Training</i>	60
4.1.2	Pengumpulan Data <i>Testing</i>	62
4.2	Proses <i>Preprocessing</i>	64
4.2.1	<i>Case Folding</i>	64
4.2.2	<i>Cleansing</i>	64
4.2.3	<i>Normalization</i>	65
4.2.4	<i>Convert Negation</i>	66
4.2.5	<i>Tokenizing</i>	67
4.2.6	<i>Stopword Removal</i>	68
4.2.7	<i>Part of Speech Tagging</i>	68
4.2.8	<i>Stemming</i>	69
4.2.9	Seleksi Fitur <i>Lexicon Based</i>	70
4.3	<i>Naive Bayes Classifier</i>	71
4.3.1	Pembobotan <i>Naive Bayes</i>	72
4.3.2	Klasifikasi <i>Naive Bayes</i>	72
4.4	Desain dan Implementasi Antarmuka.....	73
4.4.1	Halaman Utama.....	73
4.4.2	Halaman <i>Crawling</i>	74
4.4.3	Halaman Seleksi Fitur <i>Lexicon Based</i>	75
4.4.4	Halaman Pembobotan <i>Naive Bayes</i>	76
4.5	Hasil Pengujian	76
BAB V PENUTUP		
5.1	Kesimpulan.....	78
5.2	Saran.....	78
DAFTAR PUSTAKA		80

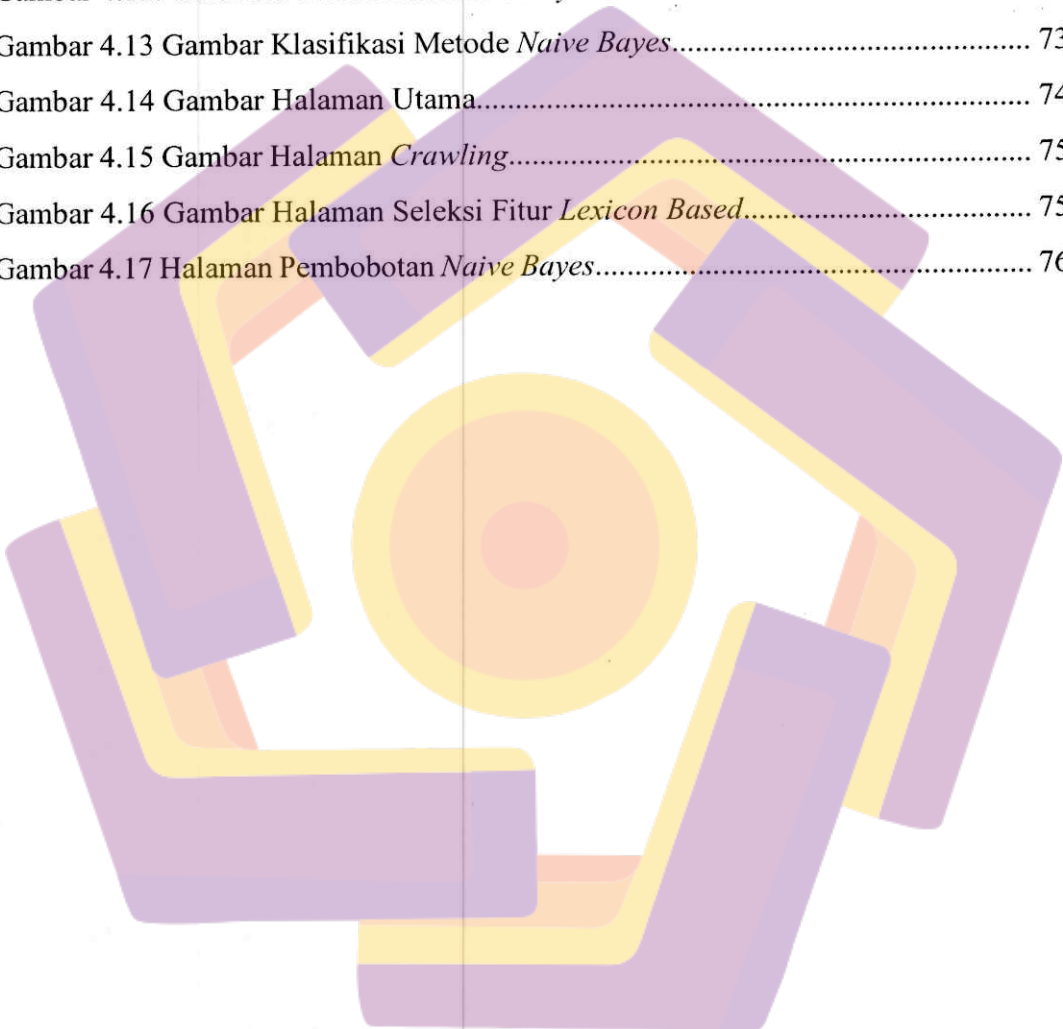
DAFTAR TABEL

Tabel 3.1 Tabel <i>Tweet Training</i>	31
Tabel 3.2 Tabel <i>Tweet Testing</i>	31
Tabel 3.3 Tabel Data <i>Stopword</i>	32
Tabel 3.4 Tabel <i>Case Folding</i>	34
Tabel 3.5 Tabel Proses <i>Cleansing</i>	36
Tabel 3.6 Tabel <i>Normalization</i>	37
Tabel 3.7 Tabel <i>Convert Negation</i>	39
Tabel 3.8 Tabel <i>Tokenizing</i>	40
Tabel 3.9 Tabel <i>Stopword Removal</i>	42
Tabel 3.10 Tabel <i>Stemming</i>	45
Tabel 3.11 Tabel <i>Part of Speech Tagging</i>	46
Tabel 3.12 Tabel Metode <i>Lexicon Based</i>	47
Tabel 3.13 Contoh Kasus Data <i>Training</i>	51
Tabel 3.14 Perhitungan Probabilitas Data <i>Training</i>	51
Tabel 3.15 Contoh Kasus Data <i>Testing</i>	53
Tabel 3.16 Nilai Probabilitas Data <i>Testing</i>	55
Tabel 4.1 Tabel Hasil Pengujian.....	77

DAFTAR GAMBAR

Gambar 3.1 Gambaran Umum Sistem.....	29
Gambar 3.2 Diagram Alir <i>Preprocessing</i>	33
Gambar 3.3 Diagram Alir <i>Case Folding</i>	34
Gambar 3.4 Diagram Alir <i>Cleansing</i>	35
Gambar 3.6 Diagram Alir <i>Normalization</i>	37
Gambar 3.7 Diagram Alir <i>Convert Negation</i>	38
Gambar 3.8 Diagram Alir <i>Tokenizing</i>	40
Gambar 3.9 Diagram Alir <i>Stopword Removal</i>	41
Gambar 3.10 Diagram alir <i>Stemming</i>	44
Gambar 3.11 Diagram Alir <i>Part of Speech Tagging</i>	46
Gambar 3.12 Diagram Alir Metode <i>Naive Bayes</i>	50
Gambar 3.13 Gambar Desain Sistem.....	56
Gambar 3.14 Gambar Antarmuka Halaman Utama.....	57
Gambar 3.15 Antarmuka Proses <i>Crawling</i>	58
Gambar 3.16 Antar Muka Proses <i>Preprocessing</i>	58
Gambar 3.17 Antarmuka Proses Seleksi Fitur <i>Lexicon Based</i>	59
Gambar 3.18 Antar Muka Proses Bobot <i>Naive Bayes</i>	59
Gambar 4.1 Gambar halaman Utama.....	61
Gambar 4.2 Gambar Pengumpulan Data <i>Testing</i>	63
Gambar 4.3 Gambar <i>Case Folding</i>	64
Gambar 4.4 Gambar Proses <i>Cleansing</i>	65
Gambar 4.5 Gambar Proses <i>Normalization</i>	66
Gambar 4.6 Gambar Proses <i>Convert Negation</i>	67
Gambar 4.7 Gambar Proses <i>Tokenize</i>	68
Gambar 4.8 Gambar Proses <i>Stopword</i>	68

Gambar 4.9 Gambar Proses <i>Part of Speech Tagging</i>	69
Gambar 4.10 Gambar Proses <i>Stemming</i>	70
Gambar 4.11 Gambar Proses <i>Fitur Lexicon Based</i>	71
Gambar 4.12. Gambar Pembobotan <i>Naive Bayes</i>	72
Gambar 4.13 Gambar Klasifikasi Metode <i>Naive Bayes</i>	73
Gambar 4.14 Gambar Halaman Utama.....	74
Gambar 4.15 Gambar Halaman <i>Crawling</i>	75
Gambar 4.16 Gambar Halaman Seleksi <i>Fitur Lexicon Based</i>	75
Gambar 4.17 Halaman Pembobotan <i>Naive Bayes</i>	76



INTISARI

Pemilihan Presiden Indonesia akan dilangsungkan pada 17 April 2019 yang keramaiannya sudah mulai bisa dirasakan, para calon Presiden sering memanfaatkan media sosial untuk berkampanye dan meningkatkan popularitas serta mengukur opini publik untuk maju sebagai calon Presiden Indonesia 2019.

Salah satu media jejaring sosial yang telah dimanfaatkan dalam pemilihan umum adalah Twitter yang sekarang ini menjadi salah satu tempat promosi atau kampanye yang efektif dan efisien. Oleh karena itu, penelitian ini mencoba memanfaatkan Twitter dengan menganalisis *tweet* berbahasa Indonesia yang membicarakan calon Presiden. Analisis dilakukan dengan melakukan klasifikasi *tweet* yang berisi sentimen masyarakat tentang calon Presiden.

Metode yang digunakan dalam penelitian ini, untuk *preprocessing* data menggunakan *tokenisasi*, *cleansing* dan *filtering*, untuk menentukan kelas sentimen dengan metode *Lexicon Based*. Untuk proses klasifikasinya menggunakan metode *Naïve Bayes Classifier* (NBC). Hasil penelitian ini menghasilkan sebuah sistem yang dapat mengklasifikasi sentimen secara otomatis dengan hasil pengujian 100 *tweet* mencapai 91 % dengan 1000 data *training*.

Kata Kunci : Sentimen Analisis, *Lexicon Based Method*, *Naive bayes Classifier*

ABSTRACT

The Indonesian Presidential Election will be held on April 17, 2019, where the crowds have begun to be felt, the presidential candidates often use social media to campaign and increase popularity and measure public opinion to advance as candidates for the Indonesian President 2019.

One of the social networking media that has been utilized in the elections general is that Twitter is now one of the most effective and efficient promotion or campaign venues. Therefore, this study tried to use Twitter by analyzing Indonesian-language tweets that talked about presidential candidates. The analysis was carried out by classifying tweets containing public sentiments about presidential candidates.

The method used in this study, to preprocessing data using tokenisation, cleansing and filtering, to determine the class sentiment using the Lexicon Based method. For the classification process using the Naïve Bayes Classifier (NBC) method. The results of this study produce a system that can classify sentiments automatically with the results of testing 100 tweets reaching 91% with 1000 training data.

Keywords: *Sentiment Analysis, Lexicon Based Method, Naive bayes Classifier*