

TESIS

**PREDIKSI PEMILIHAN PRESIDEN AMERIKA
SERIKAT 2016 DENGAN MENGANALISIS
SENTIMEN PADA SOSIAL MEDIA**



ANDY JANUAR WICAKSONO

No. Mhs.: 155302392/PS/MTF

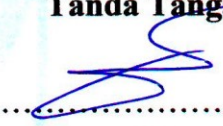
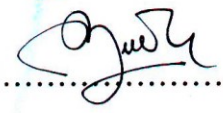
**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCASARJANA
UNIVERSITAS ATMA JAYA YOGYAKARTA
2016**



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : ANDY JANUAR WICAKSONO
Nomor Mahasiswa : 155302392/PS/MTF
Konsentrasi : *Mobile Computing*
Judul Thesis : PREDIKSI PEMILIHAN PRESIDEN AMERIKA
SERIKAT 2016 DENGAN MENGANALISIS
SENTIMEN PADA SOSIAL MEDIA

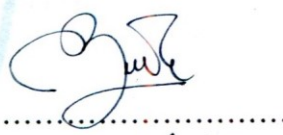
Nama Pembimbing	Tanggal	Tanda Tangan
Prof. Ir. Suyoto, M.Sc., Ph.D.	17-10-2016	
Dr. Pranowo, S.T., M.T.	18-10-2016	

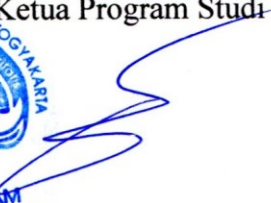


UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : ANDY JANUAR WICAKSONO
Nomor Mahasiswa : 155302392/PS/MTF
Konsentrasi : *Mobile Computing*
Judul Thesis : PREDIKSI PEMILIHAN PRESIDEN AMERIKA
SERIKAT 2016 DENGAN MENGANALISIS
SENTIMEN PADA SOSIAL MEDIA

Nama Pembimbing	Tanggal	Tanda Tangan
Prof. Ir. Suyoto, M.Sc., Ph.D. (Ketua)	25-10-2016	
Dr. Pranowo, S.T., M.T. (Sekretaris)	25-10-2016	
Ir. A. Djoko Budiyanto, M. Eng, Ph. D. (Anggota)	25-10-2016	

Ketua Program Studi

UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
Prof. Ir. Suyoto, M.Sc., Ph.D.



PERNYATAAN KEASLIAN KARYA

Saya yang bertanda-tangan dibawah ini:

Nama	:	ANDY JANUAR WICAKSONO
Nomor Mahasiswa	:	155302392/PS/MTF
Konsentrasi	:	<i>Mobile Computing</i>
Judul Thesis	:	PREDIKSI PEMILIHAN PRESIDEN AMERIKA SERIKAT 2016 DENGAN MENGANALISIS SENTIMEN PADA SOSIAL MEDIA

Menyatakan bahwa penelitian ini adalah hasil pemikiran sendiri dan bukan duplikasi karya tulis yang telah ada sebelumnya. Karya tulis yang telah ada sebelumnya dijadikan acuan oleh penulis guna melengkapi penelitian ini dan dinyatakan secara tertulis dalam penulisan acuan dan daftar pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

Yogyakarta, 16 Oktober 2016
Yang menyatakan,

Andy Januar Wicaksono

KATA PENGANTAR

Puji syukur dan terima kasih saya panjatkan ke hadirat Tuhan Yang Maha Esa atas segala berkat dan rahmat-Nya sehingga saya dapat menyelesaikan penelitian tesis ini tepat waktu. Penelitian tesis ini dapat terselesaikan dengan dukungan banyak pihak, baik secara moril maupun materiil. Melalui kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Ketua Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta, Bapak Prof. Ir. Suyoto, M. Sc., Ph. D. selaku Pembimbing I dan Bapak Dr. Pranowo, S.T., M.T. selaku Pembimbing II yang telah membimbing dan memberikan banyak masukan dalam proses penelitian hingga pembuatan laporan tesis.
2. Bapak Ir. A. Djoko Budiyo, M. Eng., Ph.D. yang telah memberikan masukan saat ujian proposal dan pendadaran.
3. Keluarga yang telah mengizinkan dan membantu penulis untuk melanjutkan studi S2 serta selalu mendukung selama proses perkuliahan hingga penulis menyelesaikan penelitian tesisnya.
4. Teman-teman Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta yang saling memberikan semangat, motivasi, dan dukungan dari awal hingga akhir.
5. Teman-teman Sekolah Tinggi Teknik Surabaya yang telah memberikan bantuan dan selalu menghibur dikala penulis sedang merasa suntuk.
6. Pihak-pihak lain yang telah membantu pembuatan tesis, baik secara langsung maupun tidak langsung.

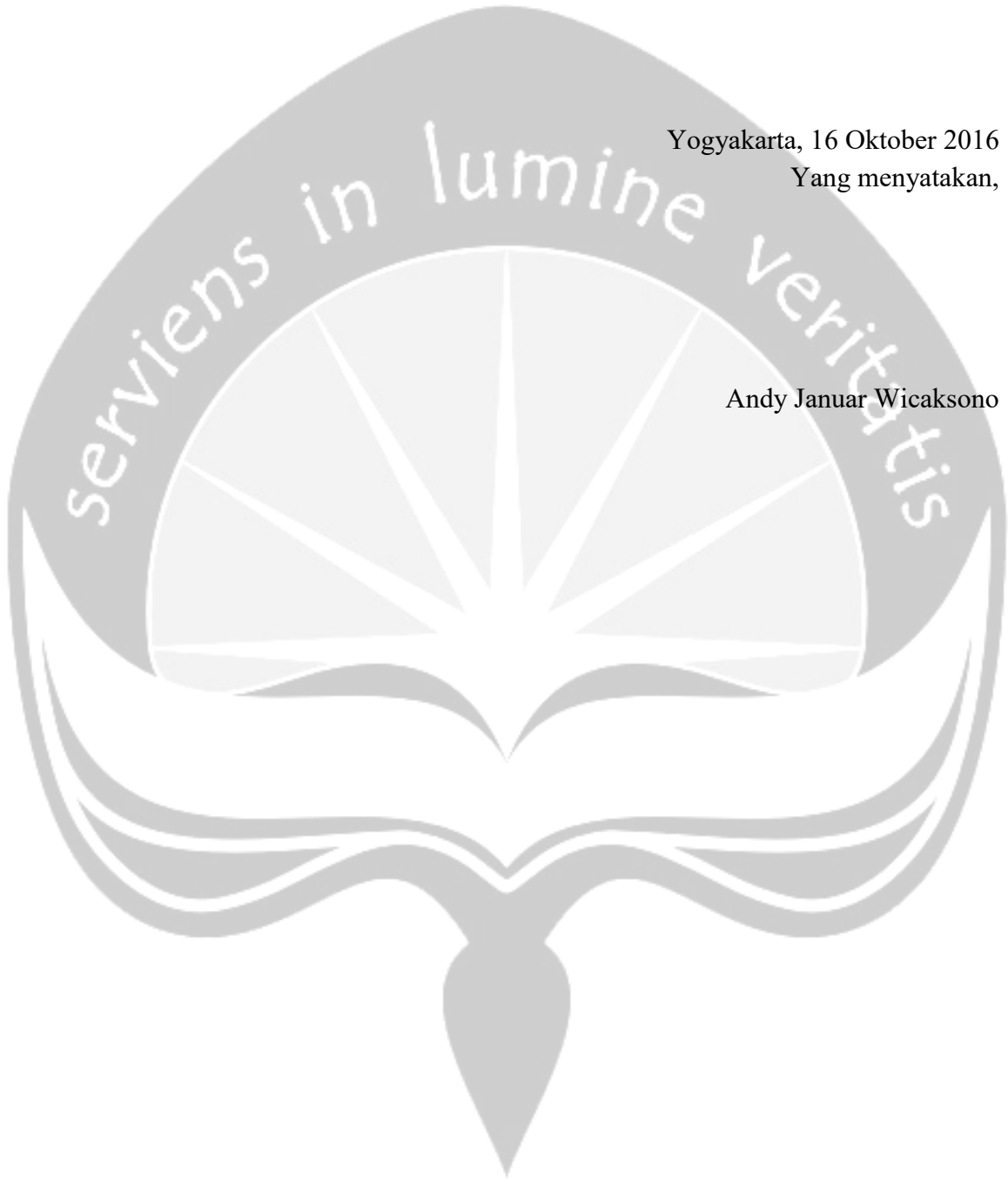
Mungkin tesis ini masih jauh dari kata sempurna, oleh karena itu penulis akan selalu bersikap terbuka untuk menerima kritik dan saran agar dapat mengembangkan penelitian ini maupun penelitian selanjutnya.

Penulis berharap penelitian ini dapat dinikmati dan memberikan manfaat bagi setiap orang yang membaca laporan tesis ini. Sekian dan terima kasih.

Yogyakarta, 16 Oktober 2016

Yang menyatakan,

Andy Januar Wicaksono



INTISARI

Pemilihan presiden Amerika Serikat merupakan sebuah peristiwa yang dinanti-nantikan oleh warga Amerika bahkan masyarakat dunia. Masyarakat dapat mengungkapkan pendapat, pernyataan, maupun tanggapan tentang pemilihan presiden Amerika Serikat melalui internet, khususnya sosial media. Peristiwa tersebut mendorong peneliti untuk mengembangkan suatu penelitian untuk memprediksi hasil pemilihan presiden Amerika Serikat tahun 2016 dengan memanfaatkan data yang dituliskan oleh pengguna dalam sosial media. Sosial media yang digunakan sebagai sumber data penelitian adalah Facebook dan Twitter.

Prediksi hasil pemilihan umum Amerika Serikat tahun 2016 dilakukan dengan menganalisa sentimen yang diungkapkan melalui teks yang dituliskan oleh pengguna di media sosial. Sebelumnya telah dilakukan penelitian-penelitian serupa yang menggunakan teknik analisis sentimen, namun terdapat kekurangan yang ditemukan yaitu sistem pemilihan umum hanya menggunakan *popularity votes*, sedangkan sistem pemilihan umum yang secara nyata diterapkan dalam pemilihan presiden di Amerika Serikat adalah *electoral college*. Peneliti menggunakan penelitian-penelitian terdahulu sebagai acuan untuk mengembangkan penelitian ini, serta memodifikasinya hingga dapat lebih menyerupai situasi aktualnya. Metodologi penelitian yang diusulkan oleh peneliti terdiri dari enam tahap, yaitu: pengumpulan data dan *tool*, pembuatan metode analisis sentimen, menguji dan membandingkan hasil dan performa metode analisis sentimen, analisis sentimen dengan menggunakan data asli, agregasi, dan implementasi sistem *electoral college*. Metode analisis sentimen yang dibandingkan dan diterapkan pada penelitian ini adalah Multinomial Naïve Bayes, Binarized Multinomial Naïve Bayes, SentiWordNet, dan AFINN-111.

Hasil perbandingan metode analisis sentimen dengan menggunakan data uji menunjukkan bahwa Multinomial Naïve Bayes memberikan tingkat akurasi prediksi yang paling baik dibandingkan ketiga metoda analisis sentimen yang lain. AFINN-111 memberikan performa terbaik dalam kecepatan untuk mengklasifikasikan sentimen data uji. Implementasi seluruh tahapan dalam metodologi penelitian diuji dengan menggunakan data asli yang dikumpulkan dari tanggal 7 Agustus 2016 hingga 25 September 2016. Jumlah data asli yang berhasil dikumpulkan adalah sebanyak 22.845 yang dibagi menjadi 36% data untuk kata kunci Democrat, 15% data untuk kata kunci Hillary Clinton, 26% data untuk kata kunci Republican, dan 23% data untuk kata kunci Donald Trump.

Kata-kata kunci: pemilihan presiden Amerika Serikat, analisis sentimen, sosial media

ABSTRACT

United States (US) presidential elections is an event anticipated by US citizens and people all around the world. Any person can share his/her opinion, statement, or thought about US presidential elections via internet, especially social media. Researchers moved to make a research to predict the result of US presidential elections 2016 by utilizing data provided by social media. This research uses Facebook and Twitter as data resources.

This research uses sentiment analysis to predict the result of US presidential elections 2016. Sentiment analysis is used to classify the sentiment class of text written by users in social media. Previously there are many similar researches, but those researches only uses popularity votes. In fact, US applying electoral college as its election system or process. Previous researches provides any knowledge required for predicting US presidential election, this research modify the previous and adding some features in order to match the actual situation. This research proposes six stages of research methodology, which is: data and tool collection, developing the sentiment analysis models, test and compare the result and performance of sentiment analysis methods, process the real data using sentiment analysis model, aggregation, and predict the winning candidate or political party using electoral college. There are four sentiment analysis methods used in this research, Multinomial Naïve Bayes, Binarized Multinomial Naïve Bayes, SentiWordNet, dan AFINN-111.

The comparison of sentiment analysis methods using data test explains that Multinomial Naïve Bayes has the highest accuracy and AFINN-11 has the fastest processing time. This research tried to test the implementation of research methodology by using data real collected from August 7th, 2016 to September 25th, 2016. There are four keywords used to collect the data real, which is Democrat, Hillary Clinton, Republican, and Donald Trump. In summary, 22.845 data were collected, divided into 36% for Democrat, 15% data for Hillary Clinton, 26% data for Republican, and 23% for Donald Trump.

Keywords: *US presidential elections, sentiment analysis, social media*

DAFTAR ISI

PENGESAHAN TESIS	Error! Bookmark not defined.
PERNYATAAN KEASLIAN KARYA	iii
KATA PENGANTAR	iv
INTISARI.....	vi
<i>ABSTRACT</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiii
BAB I.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Keaslian Penelitian	4
1.5. Manfaat.....	4
1.6. Tujuan.....	5
1.7. Sistematika Penulisan.....	6
BAB II.....	8
BAB III	15
3.1. <i>Electoral College</i>	15
3.2. Analisis Sentimen.....	16
BAB IV	17
4.1. Pengumpulan Data dan <i>Tool</i>	17
4.1.1. Kamus Singkatan	18
4.1.2. Daftar Negara Bagian.....	18
4.1.3. Data Latih dan Data Uji	18
4.1.4. Data Asli.....	19
4.1.5. Daftar <i>Stop Words</i>	20
4.1.6. Leksikon SentiWordNet.....	21
4.1.7. Leksikon AFINN-111	21

4.2.	Pembuatan Model Analisis sentimen	21
4.2.1.	Multinomial Naïve Bayes dan Binarized Multinomial Naïve Bayes 22	
4.2.2.	SentiWordNet.....	33
4.2.3.	AFINN-111	37
4.3.	Menguji dan Membandingkan Hasil Analisis sentimen.....	38
4.4.	Analisis Sentimen dengan Menggunakan Data Asli	40
4.5.	Agregasi.....	41
4.6.	Implementasi <i>Electoral College</i>	41
BAB V		43
5.1.	Perbandingan Metode Analisis sentimen	43
5.2.	Metodologi Penelitian	46
BAB VI		60
6.1.	Kesimpulan.....	60
6.2.	Saran	61
DAFTAR PUSTAKA		62
Lampiran 1.	Diagram Alir Metodologi Penelitian	65
Lampiran 2.	Kamus Singkatan.....	66
Lampiran 3.	Tabel <i>Elector</i> Negara Bagian	78
Lampiran 4.	Hasil Klasifikasi Sentiment Menggunakan Data Uji.....	79
Lampiran 5.	Bukti Publikasi	108

DAFTAR TABEL

Tabel 2.1. Perbandingan dengan Penelitian Sebelumnya	12
Tabel 4.1. Contoh Data pada Situs NLP Stanford	28
Tabel 4.2. Bentuk Biner Contoh Data pada Situs NLP Stanford	31
Tabel 4.3. Tabel Konversi POS <i>Tagger</i> ke POS SentiWordNet.....	34
Tabel 4.4. Ilustrasi <i>Confusion Matrix</i>	39
Tabel 5.1. <i>Confusion Matrix</i> Metode Analisis sentimen.....	44
Tabel 5.2. Nilai <i>precision</i> , <i>recall</i> , <i>accuracy</i> , dan <i>F1-score</i>	45
Tabel 5.3. Waktu Pemrosesan Metode Analisis sentimen	45
Tabel 5.4. Tabel Contoh Hasil Pra-proses dan Analisa Sentimen	53
Tabel 5.5. Hasil Prediksi Pemilihan Presiden AS 2016 dengan Menggunakan Data Asli pada Tanggal 25 September 2016	56

DAFTAR GAMBAR

Gambar 4.1. Fungsi Hapus Elemen HTML dan <i>Twitter Mention</i>	23
Gambar 4.2. Fungsi Tokenisasi.....	24
Gambar 4.3. Fungsi Mengartikan Singkatan	25
Gambar 4.4. Fungsi Filtrasi.....	26
Gambar 4.5. Fungsi Lematisasi atau <i>Stemming</i>	27
Gambar 4.6. Fungsi Pembuatan Unigram dan Bigram	28
Gambar 4.7. Fungsi Multinomial Naïve Bayes.....	31
Gambar 4.8. Fungsi POS <i>Tagging</i>	33
Gambar 4.9. Kelas Implementasi SentiWordNet.....	35
Gambar 4.10. Kelas AFINN-111	37
Gambar 5.1. Persentase Jumlah Data Asli Menurut Kata Kuncinya yang Diperoleh dari Tanggal 7 Agustus hingga 25 September	47
Gambar 5.2. Contoh <i>Post</i> dan Komentar pada Halaman Facebook Donald Trump	48
Gambar 5.3. Tampilan Aplikasi untuk Menyimpan Data dari Facebook	48
Gambar 5.4. Tampilan Aplikasi untuk Menampilkan dan Menambahkan Data dari Twitter secara Manual.....	50
Gambar 5.5. Tampilan Data Twitter yang Tersimpan secara Otomatis.....	51
Gambar 5.6. Contoh Hasil Pra-proses dan Analisis Sentimen dengan Binarized Multinomial Naïve Bayes	53
Gambar 5.7. Tampilan Aplikasi untuk Prediksi Pemilihan Presiden AS 2016 dengan Menggunakan Data Asli pada Tanggal 25 September 2016.....	55

Gambar 5.8 Hasil Prediksi <i>Electoral Votes</i> Mingguan dengan Menggunakan Multinomial Naïve Bayes dari Tanggal 7 Agustus 2016 hingga 25 September 2016.....	57
Gambar 5.9 Hasil Prediksi <i>Electoral Votes</i> Mingguan dengan Menggunakan Binarized Multinomial Naïve Bayes dari Tanggal 7 Agustus 2016 hingga 25 September 2016	58
Gambar 5.10 Hasil Prediksi <i>Electoral Votes</i> Mingguan dengan Menggunakan SentiWordNet dari Tanggal 7 Agustus 2016 hingga 25 September 2016	58
Gambar 5.11 Hasil Prediksi <i>Electoral Votes</i> Mingguan dengan Menggunakan AFINN-111 dari Tanggal 7 Agustus 2016 hingga 25 September 2016.....	59

DAFTAR LAMPIRAN

Lampiran 1. Diagram Alir Metodologi Penelitian	65
Lampiran 2. Kamus Singkatan.....	66
Lampiran 3. Tabel <i>Elector</i> Negara Bagian	78
Lampiran 4. Hasil Klasifikasi Sentiment Menggunakan Data Uji.....	79
Lampiran 5. Bukti Publikasi	108

