

TESIS

**SENTIMENT ANALYSIS PEMILU PRESIDEN INDONESIA
2019 PADA TWITTER DENGAN LEXICON-BASED DAN
SUPPORT VECTOR MACHINE (SVM)**



Fredericus Dwi Nugraha Putra

No. Mhs.: 175302672/PS/MTF

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PROGRAM PASCASARJANA

UNIVERSITAS ATMA JAYA YOGYAKARTA

2019

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
ABSTRAK	iii
KATA PENGANTAR	iv
DAFTAR TABEL.....	v
DAFTAR LAMPIRAN.....	x
BAB I PENDAHULUAN	1
A. Latar Belakang	1
1.1 Rumusan Permasalahan	4
1.2 Batasan Masalah.....	4
1.3 Keaslian Penelitian.....	4
1.4 Manfaat Penelitian	5
1.5 Tujuan Penelitian	5
1.5 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	8
B Penelitian Terkait	8
BAB III TINJAUAN PUSTAKA.....	8
Landasan Teori	12
3.1. Media Social Twitter	12
3.2. Big Data	13
3.3. Text Mining	16
3.4. Neural Network Processing (NNP).....	17
3.5. Mechine Learning	18
3.6. Knowledge-Based	21
3.6. Scraping Data	22
3.6. Lexicon Based	23

3.6. Pengaruh Media Social Pada Pemilihan Umum	24
BAB IV METODOLOGI PENELITIAN	26
4.1 Alat dan Bahan	26
4.1.1 Dataset Tweet Twitter	26
4.1.2 Perangkat Lunak	27
4.1.3 Perangkat Keras	28
4.2. Langkah Penelitian	28
4.2.1 Pengumpulan Data	29
4.2.2 Crawling Atau Mengambil data Mentah Twitter	30
4.2.3 Replace Kata/Pengantian Kata	35
4.2.4 Klasifikasi	37
4.2.5 Support Vector Mechine (SVM)	40
BAB V HASIL DAN PEMBAHASAN	41
5.1. Proses Akuisisi Data	43
5.2. Preprocessing	43
5.2.1 Pembersihan Dokumen	43
5.2.2 Case Folding	44
5.2.3 Parsing	45
5.2.4 Filtering	46
5.2.5 Stemming	47
5.2.6 Proses Pembobotan	47
5.3. Klasifikasi Dengan Support Mechine Vector (SVM)	49
5.34. Percobaan Dengan Data Pelatihan 90% dan Pengujian 20%...	51
Bab VI KESIMPULAN DAN SARAN.....	53
Kesimpulan dan Saran	53
DAFTAR PUSTAKA	55
LAMPIRAN-LAMPIRAN	56



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : FREDERICUS DWI NUGRAHA PUTRA
Nomor Mahasiswa : 175302672/PS/MTF
Konsentrasi : Enterprise Information System
Judul Tesis : **SENTIMENT ANALYSIS PEMILU PRESIDEN
INDONESIA 2019 PADA TWITTER DENGAN
LEXICON-BASED DAN SUPPORT VECTOR
MACHINE (SVM)**

Nama Pembimbing

Tanggal

Tanda Tangan

Dr. Pranowo, S.T., M.T.

(KETUA)

25-7-2019

Ir. A. Djoko Budiyanto, M.Eng., Ph.D

(SEKRETARIS)

25-7-2019

Prof. Ir. Suyoto, M.Sc., Ph.D.

(ANGGOTA)

25-7-2019

Ketua Program Studi

Ir. A. Djoko Budiyanto, M.Eng., Ph.D.
PASCASARJANA

HALAMAN PERNYATAAN TESIS

Saya yang bertanda – tangan dibawah ini:

NAMA : Fredericus Dwi Nugraha Putra
NIM : 175302672/PS/MTF
JUDUL TESIS : SENTIMENT ANALYSIS PEMILU PRESIDEN INDONESIA
2019 PADA TWITTER DENGAN LEXICON-BASED DAN
SUPPORT VECTOR MACHINE (SVM)

Menyatakan dengan sesungguhnya bahwa Tesis ini merupakan hasil penelitian, pemikiran dan pemaparan asli saya sendiri. Saya tidak mencantumkan tanpa pengakuan bahan - bahan yang telah dipublikasikan sebelumnya atau ditulis oleh orang lain, atau sebagai bahan yang pernah diajukan untuk gelar atau ijazah pada Universitas Atma Jaya Yogyakarta atau perguruan tinggi lainnya.

Apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan peraturan yang berlaku di Universitas Atma Jaya Yogyakarta.

Demikian pernyataan ini saya buat.

Yogyakarta, 18 / 07 / 2019

Yang membuat pernyataan,

(FREDERICUS DWI NUGRAHA PUTRA)

ITISARI

Perkembangan media social saat ini sangat pesat, dan twitter merupakan media social yang banyak digunakan orang dalam menyebarkan informasi, bahkan dalam proses pemilihan umum presiden media social twitter memiliki peran penting dalam penyebaran informasi yang bertujuan baik ataupun jahat dengan tujuan menjatuhkan citra salah satu calon presiden. Penelitian ini bertujuan menganalisa seluruh informasi yang di dapatkan dari twitter untuk memperoleh nilai positif dan negatif hingga didapatkan prediksi elektabilitas calon presiden. Pada proses klasifikasi penelitian ini menggunakan metode Lexicon-Base dan Support Vector Machine (SVM). Tahapan preprocessing data menggunakan Part-Of-Speech Tagging, chi-square test, dan aggregating opinion pada entity (NN & NNP) yang bertujuan supaya tidak terjadi low recall dari metode lexicon-based. Data yang digunakan mengambil tweet baik berbahasa indonesia maupun inggris dengan jumlah dataset 2000 tweet dengan pemisahan dua opinion dari masing-masing calon presiden. Dari Akurasi polaritas di dapatkan nilai dengan keyword jokowi memperoleh nilai positif 24.10%, negatif 38.30% dan netral 38.30% sedangkan untuk keyword prabowo memperoleh nilai positif 0.20%, negatif 0.10% dan netral 0.70%.

Keyword: Sentiment Analysis, Twitter, Lexicon-Base, SVM

ABSTRACT

The development of social media is currently very rapid, and Twitter is a social media that is widely used by people in disseminating information, even in the presidential election process of social media twitter has an important role in the dissemination of good or evil information to drop the image of one presidential candidate. This study aims to analyze all information obtained from Twitter to obtain positive and negative values to obtain the electability predictions of presidential candidates. In the process of classification, this research uses the Lexicon-Base method and Support Vector Machine (SVM). The preprocessing data stage uses Part-Of-Speech Tagging, chi-square test, and aggregating opinion on the entity (NN & NNP) which aims to avoid low recall from the lexicon-based method. The data used takes tweets both in Indonesian and English with several datasets of 2000 tweets with the separation of two opinions from each presidential candidate. From the polarity accuracy, get the value with the jokowi keyword to get a positive value of 24.10%, negative 38.30% and neutral 38.30% while for the keyword prabowo it gets a positive value of 0.20%, negative 0.10% and neutral 0.70%.

Keyword: Sentiment Analysis, Twitter, Lexicon-Base, SVM

KATA PENGANTAR

Dengan segala puji syukur penulis kepada Tuhan Yesus Kristus, atas segala perlindungannya dan dan berkatya, sehingga tesis dengan judul “SENTIMENT ANALYSIS PEMILU PRESIDEN INDONESIA 2019 PADA TWITTER DENGAN LEXICON-BASED DAN SUPPORT VECTOR MACHINE (SVM)” ini dapat diselesaikan.

Tesis ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Magister Teknik Informatika (M.T.) Universitas Atma Jaya Yogyakarta.

Pada kesempatan ini penulis menyampaikan rasa hormat dan menghaturkan terima kasih yang sebesar-besarnya, kepada :

1. Bapak Dr. Pranowo, S.T., M.T. atas bimbingan, arahan dan waktu yang telah diluangkan kepada penulis untuk berdiskusi selama menjadi dosen pembimbing, dan perkuliahan .
2. Bapak Ir. A. Djoko Budiyanto, M.Eng., Ph.D. yang telah memberikan masukan dan saran selamat bimbingan tesis penulis berlangsung.
3. Ketua program studi Pascasarjana Magister Teknik Informatika Prof. Ir. Suyoto, M.Sc., Ph.D.
4. Seluruh Dosen program Pascasarja Magister Teknik Informatika yang telah memberikan arahan dan bimbingan untuk mendalami ilmu Geoteknik.

5. Istri Saya Tercinta Theresia Erika Kristiantini yang tiada henti dan lelah selalu mendukung, mendoakan dan memberikan semangat selama proses perkuliahan dan pengerjaan Tesis hingga selesai.
6. Orang tua dan Kakak yang selalu mendukung dan mensupport penulis dari awal kuliah hingga tesis ini selesai di susun.
7. Rekan Mahasiswa Magister Teknik Informatika angkatan 2017 yang selalu kompak dalam menjalani proses perkuliahan dan tesis.
8. Kepada semua pihak yang telah membantu yang tidak dapat penulis sebutkan satu persatu.

Dengan keterbatasan pengalaman, ilmu maupun pustaka yang ditinjau, penulis menyadari bahwa tesis ini masih banyak kekurangan dan pengembangan lanjut agar benar benar bermanfaat. Oleh sebab itu, penulis sangat mengharapkan kritik dan saran agar tesis ini lebih sempurna serta sebagai masukan bagi penulis untuk penelitian dan penulisan karya ilmiah di masa yang akan datang.

Akhir kata, penulis berharap tesis ini memberikan manfaat bagi kita semua terutama untuk pengembangan ilmu pengetahuan yang ramah lingkungan.

Yogyakarta, Juli 2019

Fredericus Dwi Nugraha Putra

DAFTAR PUSTAKA

- Giuseppe Castellucci, Danilo Croce, Roberto Basili, 2016 “*Context-aware Convolutional Neural Networks for Twitter Sentiment Analysis in Italian*”, Fifth International Workshop EVALITA.
- Nan Li, Desheng Dash Wu, 2010 “Using text mining and sentiment analysis for online forums hotspot detection and forecast”, *Decision Support Systems* 48 354–368
- Prem Melville, Wojciech Gryc, Richard D. Lawrence , 2009 “*Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification*” , Melville.
- Alexander Pak, Patrick Paroubek, 2014 “*Twitter as a Corpus for Sentiment Analysis and Opinion Mining*”,
- Sara Rosenthal, Noura Farra, Preslav Nakov, 2017 “*SemEval-2017 Task 4: Sentiment Analysis in Twitter*”.
- Bing Liu, Lei Zhang, 2009 "A SURVEY OF OPINION MINING AND SENTIMENT ANALYSIS"
- Prem Melville, Wojciech Gryc, Richard D. Lawrence, 2016 "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification"
- Jaka Aulia Pratama, Yadi Suprijadi, dan Zulhanif, 2017 "Analisis Sentimen Sosial Media Twitter Dengan Algoritma Machine Learning Menggunakan Software R" *JURNAL FOURIER* | Oktober, Vol. 6, No. 2, 85-89
- Patrice Y. Simard, Dave Steinkraus, John C. Platt, 2003 “*Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*”, Conference, Document Analysis and Recognition,
- Xiang Zhang, Junbo Zhao, Yann LeCun, 2015 "Character-level Convolutional Networks for Text Classification", *Proceeding*.
- Han, J., Kamber, M., & Pei, J. 2012. *Data Mining Concepts And Techniques* (3rd ed.). USA: Elsevier.
- Antonius Rachman, Yuan Lukito, 2016 "SENTIPOL: DATASET SENTIMEN KOMENTAR PADA KAMPANYE PEMILU PRESIDEN INDONESIA 2014 DARI FACEBOOK PAGE" Universitas Kristen Dutawacana.
- Alvandi Fadhil Sabily, Putra Pandu A., Muh. Ali Fauzi, mei 2019 "Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy" Inoversitas Brawijaya.
- Petrix Nomleni, 2015 "SENTIMENT ANALYSIS MENGGUNAKAN SUPPORT VECTOR MACHINE(SVM)" Institut Teknologi Sepuluh Nopember Surabaya.