

**ANALISIS SENTIMEN UNTUK KLASIFIKASI SIFAT POLITIK
TERHADAP PARTAI POLITIK INDONESIA MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE**

SKRIPSI



disusun oleh

Satrio Yudho Pangestu

15.11.9189

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

**ANALISIS SENTIMEN UNTUK KLASIFIKASI SIFAT POLITIK
TERHADAP PARTAI POLITIK INDONESIA MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE**

SKRIPSI

Untuk memenuhi sebagian persyaratan
mencapai gelar sarjana
pada program studi informatika



disusun oleh

Satrio Yudho Pangestu

15.11.9189

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

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN UNTUK KLASIFIKASI SIFAT POLITIK
TERHADAP PARTAI POLITIK INDONESIA MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

Satrio Yudho Pangestu

15.11.9189

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 mei 2018

Dosen Pembimbing,



Yuli Astuti, M.Kom

NIK. 190302146

PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN UNTUK KLASIFIKASI SIFAT POLITIK TERHADAP PARTAI POLITIK INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

yang disusun oleh

Satrio Yudho Pangestu

15.11.9189

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 April 2019

Susunan Dewan Penguji

Nama Penguji

Arifiyanto Hadinegoro, S.Kom, MT
NIK. 190302289

Erni Seniwati, S.Kom, M.Cs
NIK. 190302231

Yuli Astuti, M.Kom
NIK. 190302146

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
tanggal 30 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, 30 April 2019



Satrio Yudho Pangestu

NIM. 15.11.9189

MOTO

“Nikmat terbaik adalah rasa takut kepada Allah SWT.”

(anonim)

“Ingatlah hanya dengan mengingat Allah lah hati menjadi tenang”.

(Q.S Ar-Ra’d: 28)

“Hai orang-orang yang beriman, mintalah pertolongan kepada Allah dengan sabar dan shalat. Sesungguhnya Allah beserta orang-orang yang sabar”.

(Q.S Al-Baqarah: 153)

“Jangan biarkan kesulitanmu membuatmu gelisah, karena bagaimanapun juga hanya di malam yang paling gelap lah bintang – bintang tampak bersinar lebih terang.”

(Ali bin Abi Thalib, RA)

PERSEMBAHAN

Puji syukur atas barokah, rahmat dan karunia Allah SWT yang telah membari kemudahan, kelancaran, nikmat sehat, nikmat dalam kesusahan dan nikmat kemampuan yang telah diberikan kepada saya, sehingga dapat menyelesaikan skripsi ini dengan sebaik – baiknya. Sholawat serta salam tetap tercurahkan kepada baginda Nabi Muhammad SAW yang selalu setia mendoakan dan mencintai umatnya.

Karya ini menjadi persembahan bagi orang – orang yang peneliti kasihi dan sayangi, semoga Allah senantiasa membalas kebaikan mereka dan memberikan kemudahan dalam segala urusan mereka

1. Kepada Ibu Yuli Astuti, M.Kom atas bimbingannya selama penelitian berlangsung.
2. Kepada kedua orang tua peneliti, Ibunda tercinta Rahmawati Sribulan dan Ayahanda tersayang Badarudin Antono. Terima kasih telah memberikan motivasi, dukungan dan tentunya doa yang tidak pernah putus, terima kasih selalu memberi nasehat dan selalu percaya dengan ananda yudho. Untuk kakaku satu - satunya Yuni Permatasari yang selalu menyemangati, menghibur dan selalu siap membantu ketika ada kesulitan.
3. Kepada Rindana Zulfasari Afifatul Sulha, terima kasih telah menjadi tempat untuk berkeluh kesah. Terima kasih selalu ada, menyemangati membantu peneliti ketika *overthinking*, stress dan *anxiety*. Terima kasih tidak menyerah dan membiarkan peneliti menyerah pada diri sendiri.
4. Kepada kak Fregy Damara, Octavia dan Adheza Putri yang telah bersedia untuk membagikan ilmu yang sangat berguna bagi peneliti.
5. Untuk para penghuni kosan Bloop, terima kasih kalian selalu membantu, terima kasih telah menjadi penghibur dikala peneliti *stuck*, semoga kalian cepat menyelesaikan studi kalian.

6. Keluarga Jogjaku, para rakyat Atlantis, para atlet pitik *epsort*. Terima kasih kalian selalu memberi pengertian, terima kasih selalu menerima kehadiran peneliti, meskipun peneliti sibuk. Cepat selesaikan skripsi kalian.
7. Sahabat Supernova, Amel, Desi, Linda, Kifran dan Tira. Terima kasih telah menjadi teman yang baik, perjuangan kita bersama dari SMA akan selalu dikenang.
8. Teman seangkatan dari Ambon, Nabila, Zidan, Jimly, Dayat, Fandi, Risman dan yang lainnya. Terima kasih atas hiburan yang kalian berikan. Semoga cepat selesai skripsi kalian.
9. Sahabat dan juga suadaraku Achmad Immanudin, terima kasih *brother* atas segala hiburan dan bantuan yang telah kau berikan. Sukses selalu.
10. Serta semua pihak yang tidak dapat disebutkan satu per-satu. Terima kasih jasa kalian akan selalu peneliti kenang.

Jazzakumullahu Khair, semoga kalian semua diberi kesehatan, keberkahan, kemudahan, kesuksesan serta selalu dalam lindungan-Nya.

Yogyakarta, 30 April 2018



Satrio Yudho Pangestu

KATA PENGANTAR

Alhamdulillah rabbil'aalamin, puji syukur kehadiran Allah SWT yang telah memberi petunjuk, kekuatan, dan rahmatnya kepada peneliti hingga berhasil menyelesaikan skripsi dengan baik. Shalawat serta salam tercurahkan kepada Nabi Muhammad SAW. Skripsi ini disusun sebagai syarat mendapatkan gelar Sarjana Komputer dari Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Pada kesempatan ini, peneliti mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, baik dalam proses penelitian, penyusunan, dan dukungan moral, kepada:

1. Allah SWT atas rahmat, hidayah, serta karunia-Nya yang telah diberikan kepada penulis sehingga Tugas Akhir ini dapat terselesaikan.
2. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas AMIKOM Yogyakarta.
3. Ibu Krisnawati, S.Si.,MT. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Ibu Yuli Astuti, M.Kom. Selaku dosen pembimbing yang telah banyak memberikan pengarahan bagi peneliti dalam pembuatan tugas akhir.
5. Kedua orang tua penulis dan keluarga yang telah mendoakan dan mendukung selama ini.
6. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah berperan memberikan ilmu pengetahuan, pengalaman selama masa perkuliahan.

7. Seluruh Teman Kelas 15 S1 Informatika 10 dan Semua pihak yang telah membantu, mohon maaf tidak dapat disebutkan satu persatu.
8. Serta semua pihak yang telah membantu baik dukungan moril maupun materil, pikiran, dan tenaga dalam penyelesaian tugas akhir ini.

Penulis berharap semoga kebaikan mereka diterima Allah SWT, dan dicatat sebagai amal shaleh. Penulis berharap semoga skripsi ini dapat berguna dan bermanfaat bagi semua pihak yang membutuhkannya. Atas segala kekurangan dari skripsi ini penulis meminta maaf dan mengharapkan kritik serta saran demi sempurnanya skripsi ini.

Yogyakarta, 30 April 2019

Peneliti,

Satrio Yudho Pangestu

15.11.9189

DAFTAR ISI

PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN.....	iv
MOTO.....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
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 Penulisan.....	4
1.6 Metode Penelitian	5
1.6.1 Metode Pengumpulan Data	5
1.6.2 Metode analisis.....	5
1.6.3 Perancangan Sistem.....	6
1.6.4 Implementasi	7
1.6.5 Pengujian	7
1.7 Sistematika Penulisan	7
BAB II LANDASAN TEORI.....	9
2.1 Tinjauan Pustaka.....	9
2.2 Data Mining	12
2.3 Sentimen Analisis	12
2.4 Text Mining	13
2.5 Preprocessing	14

2.6	Term Frequency-Inverse Document Frequency(TF-IDF)	15
2.7	Support Vector Machine	16
2.8	K-Fold Cross Validation	20
2.9	Confusion Matrix	21
2.10	Flowchart	23
2.11	Perangkat Lunak Yang Digunakan	24
	2.11.1 Bahasa Pemrograman Python	24
	2.11.2 Integrated Development Environment PyCharm	26
	2.11.3 API Twitter	26
BAB III ANALISIS DAN PERANCANGAN		28
3.1	Analisis Masalah	28
3.2	Solusi Masalah	29
3.3	Analisis Kebutuhan Sistem	30
	3.3.1 Analisis Kebutuhan Fungsional	30
	3.3.2 Analisis Kebutuhan Non-Fungsional	30
3.4	Perancangan Sistem Aplikasi	31
	3.4.1 Gambaran Umum Sistem	31
	3.4.2 Perancangan File Pengumpulan Data Twitter	32
	3.4.3 Pelabelan Data Twitter	33
	3.4.4 Diagram Alir Preprocessing	34
	3.4.5 Perancangan Proses Pembobotan Data Twitter	35
	3.4.6 Diagram Alir klasifikasi Support Vector Machine	36
3.5	Analisis Algoritma	37
	3.5.1 Pengumpulan Data	37
	3.5.2 Pelabelan Tweet	38
	3.5.3 Cleaning Data	40
	3.5.4 Case Folding	42
	3.5.5 Tokenization	42
	3.5.6 Normalisasi Kata	44
	3.5.7 Penghapusan Stopword	45
	3.5.8 Stemming	46
	3.5.9 Vector Feature	47

3.5.10	TF-IDF.....	48
3.5.12	Implementasi Support Vector Machine.....	58
3.6	Rancangan Antarmuka Pengguna (User Interface).....	63
BAB IV IMPLEMENTASI DAN PEMBAHASAN		67
4.1	Deskripsi Implementasi	67
4.2	Implementasi Pengumpulan Data Tweet	67
4.3	Implementasi Pelabelan Data Training.....	70
4.4	Implementasi Perancangan Fungsi dan Sistem.....	70
4.4.1	Mengakses Dataset Training	71
4.4.2	Implementasi Processing Data	71
4.5	Vektor Tweet	75
4.5.1	Implementasi TF IDF	75
4.5.2	Implementasi Pembuatan Model SVM	76
4.6	Implementasi Rancangan Antar Muka Program.....	77
4.7	Evaluasi Model Klasifikasi.....	82
BAB V PENUTUP.....		89
5.1	Kesimpulan	89
5.2	Saran	89
DAFTAR PUSTAKA		91
LAMPIRAN.....		94

DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian Terkait Dengan Sistem Yang Dibangun	10
Tabel 2. 2 Confusion Matrix	21
Tabel 2. 3 Daftar Simbol Dokumen Flowchart	23
Tabel 3. 1 Tweet Yang Telah Di Ambil	38
Tabel 3. 2 Tweet Berlabel	39
Tabel 3. 3 Hasil Cleaning Data	40
Tabel 3. 4 Hasil Case Folding	42
Tabel 3. 5 Hasil Tokenization	43
Tabel 3. 6 Normalisasi Kata	44
Tabel 3. 7 Tabel Penghapusan Stopword	45
Tabel 3. 8 Proses Stemming	46
Tabel 3. 9 Vector Feture	47
Tabel 3. 10 Tf Tweet Kelas Positif	48
Tabel 3. 11 Tf Tweet Kelas Negatif	49
Tabel 3. 12 Df Tweet T1-T6	50
Tabel 3. 13 Idf Tweet T1-T6	51
Tabel 3. 14 Tf-Idf Tweet Positif	54
Tabel 3. 15 Tf-Idf Tweet Negatif	55
Tabel 3. 16 Hasil Akhir Tf-Idf	58
Tabel 3. 17 Persamaan W1, W2 Dan B Pada Setiap Kelas	60
Tabel 3. 18 Klasifikasi Svm	61
Tabel 3. 19 Implementasi Confussion Matrix	62
Tabel 3. 20 Accuracy, Precision, Recall, Dan F-Score	62
Tabel 4. 1 Confussion Matrix Skenario 1	84
Tabel 4. 2 Hasil Pengujian Model Dengan 10-Fold Cross Validation Dan Confusion Matrix	86

DAFTAR GAMBAR

Gambar 2. 1 Contoh Pencarian Hyperlane Dua Dimensi	18
Gambar 2. 2 Hyperlane Terbaik.....	18
Gambar 2. 3 Ilustrasi Pencarian Hyperlane Terbaik	19
Gambar 3. 1 Gambaran Umum Sistem	32
Gambar 3. 2 Alur Pengumpulan Data	33
Gambar 3. 3 Hasil Pengumpulan Data.....	33
Gambar 3. 4 Data Yang Telah Dilabeli.....	34
Gambar 3. 5 Diagram Alir Preprocessing	34
Gambar 3. 6 Diagram Alir Klasifikasi Support Vector Machine.....	36
Gambar 3. 7 Ui Main Menu	63
Gambar 3. 8 Halaman Klasifikasi	64
Gambar 3. 9 Halaman Hasil Klasifikasi.....	65
Gambar 3. 10 Ui Hasil Klasifikasi Dalam Bentuk Grafik	65
Gambar 3. 11 Ui Hasil Klasifikasi Dalam Bentuk Pie Chart.....	66
Gambar 4. 1 Script Crawling Data Twitter	68
Gambar 4. 2 Hasil Pengumpulan Data.....	69
Gambar 4. 3 Hasil Pelabelan Data	70
Gambar 4. 4 Script Akses Datatraining.Txt.....	71
Gambar 4. 5 Script Case Fold Dan Tweet Cleaning	71
Gambar 4. 6 Script Proses Tokenize	72
Gambar 4. 7 Contoh Kamus Normalisasi	72
Gambar 4. 8 Script Normalisasi	72
Gambar 4. 9 Contoh Daftar Stopword	73
Gambar 4. 10 Script Fungsi Stopword.....	73
Gambar 4. 11 Script Pemanggilan Fungsi Stopword.....	73
Gambar 4. 12 Script Fungsi Stem	74
Gambar 4. 13 Script Pembuatan Feature Vektor	75
Gambar 4. 14 Script Tf-Idf Dengan Sklearn.....	75
Gambar 4. 15 Script Implementasi Svm	76

Gambar 4. 16 Script Save Model	77
Gambar 4. 17 Menu Utama.....	77
Gambar 4. 18 Menu Klasifikasi Dengan Kata Kunci “Jokowi”	78
Gambar 4. 19 Menu Klasifikasi Dengan Kata Kunci “Prabowo”.....	78
Gambar 4. 20 Halaman Hasil Klasifikasi "Jokowi"	79
Gambar 4. 21 Halaman Hasil Klasifikasi "Prabowo"	79
Gambar 4. 22 Hasil Klasifikasi "Jokowi" Dalam Bentuk Grafik.....	80
Gambar 4. 23 Hasil Klasifikasi "Prabowo" Dalam Bentuk Grafik	80
Gambar 4. 24 Hasil Klasifikasi “Jokowi” Dalam Bentuk Pie.....	81
Gambar 4. 25 Hasil Klasifikasi "Prabowo" Dalam Bentuk Pie	82
Gambar 4. 26 Pembuatan Data Uji Evaluasi.....	83
Gambar 4. 27 Proses Evaluasi.....	83
Gambar 4. 28 Script Pembuatan Data Uji K-Fold Cross Validation Dan Confussion Matrix	85
Gambar 4. 29 Script Implementasi K-Fold Cross Validation Dan Confussion Matrix.....	86
Gambar 4. 30 Grafik Hasil Evaluasi K-Fold Dan Confussion Matrix.....	88

INTISARI

Ketersediaan informasi yang banyak dan cepat dari Twitter membuka peluang baru untuk melaksanakan sebuah penelitian yang bertujuan untuk menganalisis nilai sentiment pada *tweet* dari seorang pengguna Twitter. Analisis sentimen atau *opinion mining* merupakan proses memahami dan mengolah data tekstual untuk mendapatkan informasi sentiment yang terkandung dalam suatu kalimat. Peran opini-opini masyarakat yang berasal dari berbagai latar belakang dapat berguna sebagai *feedback* untuk partai politik yang mengusung pasangan calon presiden, wakil presiden dan tim sukses. Dengan demikian, pendapat-pendapat tersebut dapat memberikan nilai potensial yang dapat digunakan untuk tujuan yang lebih baik.

Untuk mencapai tujuan yang telah disebutkan dapat dibuat sistem untuk analisis sentimen dengan cara memilah data dari twitter yang merupakan opini terhadap partai politik dan calon eksekutif yang diusungnya. Data tersebut dipilah menjadi 2 kategori yaitu kategori positif dan negatif. Metode – metode yang akan digunakan untuk analisis sentimen diantaranya *preprocessing*, pembototan kata dengan TF-IDF, dan membuat model klasifikasi dengan pendekatan Support Vector Machine dan K-Fold Cross Validation untuk menguji tingkat akurasi model.

Hasil dari pembuatan model klasifikasi *Support Vector Machine* dengan data latih sejumlah 900 mendapatkan *accuracy* sebesar 86% dan pengujian dengan menggunakan *10-Fold Cross Validation* mendapatkan tingkat rata – rata *accuracy* sebesar 71% dengan *error rate* sebesar 29%.

Kata kunci : Analisis sentimen, Twitter, *opinion mining*, *Support Vector Machine*, politik

ABSTRACT

The availability of a lot of information and fast from twitter opening new opportunities to carry out a research that aims to analyze the value of sentiment on tweet from a user twitter. Sentiment analysis or opinion mining is the process of understanding, extracting and processing textual data automatically to get sentiment information contained in an opinion sentence. The role of public opinion originating from various backgrounds can be useful as feedback for political parties that carry the presidential and vice-presidential candidate pairs, successful teams, and also parties that are in coalition with the party. Thus, these opinions can provide potential values that can be used for better purposes.

From the review data or opinions, it can be implemented for sentiment analysis for various needs to improve the performance of political parties, presidential candidate candidates and vice presidential candidates, or even as a campaign strategy. to achieve the objectives mentioned previously, namely by creating a system for sentiment analysis or commonly called opinion mining by sorting data from twitter which is an opinion on political parties and the executive candidates it carries. The data is divided into 2 categories, positive and negative categories. The methods that will be used for sentiment analysis include preprocessing, word staining with TF-IDF, and making a classification model with the Support Vector Machine and K-Fold Cross Validation approach to test the accuracy of the model.

The result of making a classification model is Support Vector Machine with training data of 900 get 86% accuracy and testing using 10-Fold Cross Validation get an average rate of 71% with an error rate of 29%.

Keywords : *Sentiment analysis, Twitter, opinion mining, Support Vector Machine, politics*