

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN
PENGGUNA PENYEDIA LAYANAN TELEKOMUNIKASI
INDOSAT OREEDOO PADA TWITTER MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI



Disusun oleh

Dimas Rio Pradana Putra Laksana

16.11.0158

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

2020

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN
PENGGUNA PENYEDIA LAYANAN TELEKOMUNIKASI
INDOSAT OREEDOO PADA TWITTER MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



Disusun oleh

Dimas Rio Pradana Putra Laksana

16.11.0158

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

2020

PERSETUJUAN

SKRIPSI

ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA PENYEDIA LAYANAN TELEKOMUNIKASI INDOSAT OREEDOO PADA TWITTER MENGGUNAKAN ALGORITMASUPPORT VECTOR MACHINE (SVM)

yang dipersiapkan dan disusun oleh

Dimas Rio Pradana Putra Laksana

16.11.0158

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Maret 2019

Dosen Pembimbing

Asro Nasiri, Drs, M.Kom
NIK. 190302152

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN
PENGGUNA PENYEDIA LAYANAN TELEKOMUNIKASI
INDOSAT OREEDOO PADA TWITTER MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

yang dipersiapkan dan disusun oleh

Dimas Rio Pradana Putra Laksana
16.11.0158

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Juli 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

M. Rudyanto Arief, S.T, M.T
NIK. 190302098

Wiwi Widayani, M.Kom
NIK. 190302272

Asro Nasiri, Drs, M.Kom
NIK. 190302152

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Juli 2020

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, 10 Agustus 2020



Dimas Rio Pradana Putra Laksana

NIM. 16.11.0158

MOTTO

“A good start is half the battle” –



PERSEMBAHAN

Saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

1. Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayang sampai sejauh ini.
2. Kedua orang tua saya dan keluarga, yang selalu mendoakan, selalu menyemangati dan memenuhi kewajibannya untuk memenuhi kebutuhan anak.
3. Bapak Asro Nasiri, Drs, M.kom. yang telah membimbing saya dari awal sampai akhir pembuatan skripsi.
4. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
5. Teman teman saya yang selalu memberikan ide ide baru membantu saya dalam kesusahan.
6. Teman-teman kelas IF03 2016 yang selalu menemani perkuliahan, mendukung dan memberikan semangat sampai saat ini. Semoga kita dapat bertukar tawa lagi dilain waktu serta tempat yang berbeda.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Analisis Sentimen Terhadap Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Indosat Ooredoo Pada Twitter Menggunakan Algoritma Support Vector Machine (SVM).

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelar Sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Asro Nasiri, Drs, M.kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.
4. Dosen Penguji - serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamanya.
5. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat kepada saya.

6. Orang-orang baik yang senantiasa mendoakan, memberi semangat dan bantuan kepada saya.
7. Teman-teman IF03 angkatan 2016 yang telah menemani selama proses perkuliahan.
8. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung. Semoga Allah memberikan balasan yang lebih kepada semua yang telah ikut membantu saya hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik yang membangun diterima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Yogyakarta, 21 February 2020

Dimas Rio Pradana Putra Laksana

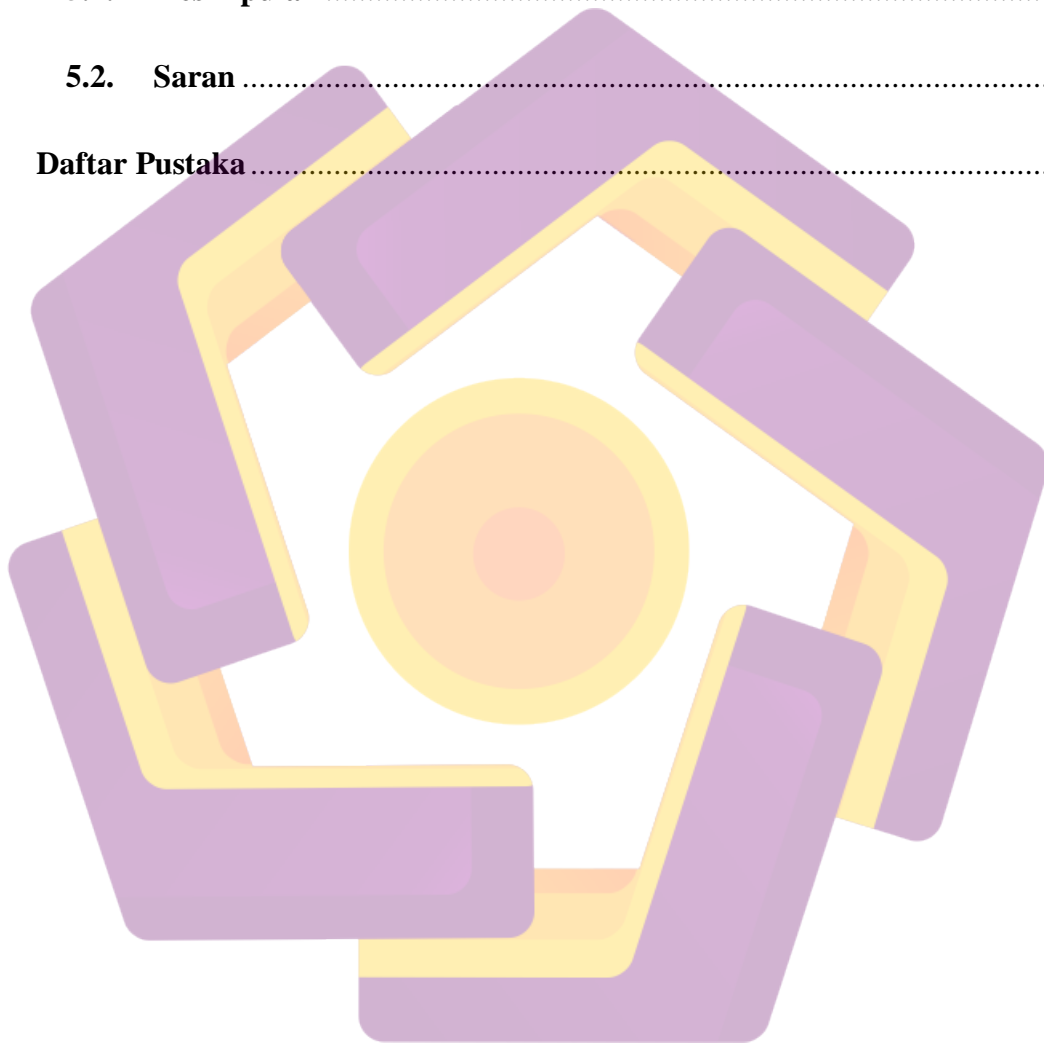
DAFTAR ISI

JUDUL	i
PERSETUJUAN	ii
PERNYATAAN	iv
MOTTO	v
PERSEMBAHAN	vii
KATA PENGHANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
INTISARI	xvi
ABSTRACT	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Metode Penelitian	5
1.6 Sistematika Penulisan	8

BAB II LANDASAN TEORI	10
2.1. Kajian Pustaka	10
2.2. Dasar Teori	12
2.2.1. Pengertian Data Mining	12
2.2.2. Pengertian Analisis Sentimen	14
2.2.3. Pengertian Text Mining	14
2.2.4. Pengertian Pre-Processing	16
2.2.5. Term Frequency-Inverse Document Frequency	17
2.2.6. Pengertian Support Vector Machine	18
2.2.7. Consine Similarity	20
2.2.8. K-Fold Cross Validation dan Confusion matrix	21
2.2.9. Flowchart	23
2.2.10. Basaha Pemograman Python	25
2.2.11. Langkah Pengembangan Sistem	26
BAB III METODE PENELITIAN	30
3.1. Analisis Masalah	30
3.2. Alat dan Bahan Penelitian	31
3.3. Alur Penelitian	32
3.3.1. Pengumpulan Data	32

3.3.2.	Cleaning Data	32
3.3.3.	Case Folding	34
3.3.4.	Tokenization	35
3.3.5.	Stemming	37
3.3.6.	Pembobotan TF-IDF	39
3.3.7.	Consine Similarity	45
3.3.8.	Support Vector Machine	49
3.3.9.	K-Fold Cross Validation	52
3.4.	Perancangan Sistem	53
3.4.1.	Diagram Alur Preprocessing	53
3.4.2.	Proses Klasifikasi SVM	54
3.5.	Perancangan Data Training	56
BAB IV Hasil DAN PEMBAHASAN		57
4.1	Deskripsi Sistem	57
4.2	Implementasi Download Data Twitter	57
4.3	Implementasi Pembuatan Data Training	59
4.4	Implementasi Proses Training Data	60
4.5	Implementasi Proses Prediksi dan Klasifikasi	62
4.6	Pembuatan Interface	64
4.6.1.	Halaman Analisis Sentimen	64

4.6.2. Halaman Hasil Analisa Sentimen.....	65
4.7 Evaluasi Program.....	66
BAB V PENUTUP	74
5.1. Kesimpulan	74
5.2. Saran	75
Daftar Pustaka	76



DAFTAR TABEL

Tabel 2. 1 Penelitian Sebelumnya.....	11
Tabel 2. 2 Confusion Matrix.....	22
Tabel 3. 1 Data Tweet	33
Tabel 3. 2 Data Tweet Setelah Cleaning.....	34
Tabel 3. 3 Data Tweet Setelah Case Folding.....	35
Tabel 3. 4 Data Tweet Setelah Tokenization.....	36
Tabel 3. 5 Data Tweet Setelah Stemming.....	37
Tabel 3. 6 Term Positif dan Negatif	38
Tabel 3. 7 Term frequency Positive.....	40
Tabel 3. 8 TF-IDF Positif.....	43
Tabel 3. 9 TF-IDF Negatif	44
Tabel 3. 10 Consine Similarity Negatif.....	47
Tabel 3. 11 Consine Similarity Positif	48
Tabel 3. 12 Consine Similarity.....	49
Tabel 3. 13 Perhitungan Persamaan Positif Terhadap Data Training	50
Tabel 3. 14 Perhitungan Persamaan Negatif Terhadap Data Training.....	51
Tabel 3. 15 Hasil Klasifikasi Data	52
Tabel 3. 16 Fold Pertama.....	52
Tabel 3. 17 Fold Kedua.....	52
Tabel 3. 18 Fold Ketiga.....	53
Tabel 4. 1 Confusion Matrix.....	66
Tabel 4. 2 Perbandingan Hasil Evaluasi Terhadap Rasio Data	68

Tabel 4. 3 Hasil Cross Validation SVM 68



DAFTAR GAMBAR

Gambar 2. 1 Hyperplane pada Support Vector Machine	19
Gambar 2. 2 Alur Analisa Sentimen	26
Gambar 2. 3 Langkah Preprocessing	28
Gambar 3. 1 Diagram Alur Preprocessing	53
Gambar 3. 2 Diagram Alur Klasifikasi SVM	54
Gambar 3. 3 Data Training	56
Gambar 4. 1 Perintah Crawling Data Twitter	58
Gambar 4. 2 Data File JSON Twitter	59
Gambar 4. 3 Data Training	60
Gambar 4. 4 Akses Data Training	60
Gambar 4. 5 Proses Preprocessing	61
Gambar 4. 6 Proses Pembobotan dan Klasifikasi	61
Gambar 4. 7 Proses Download Data Tweet	62
Gambar 4. 8 Proses Cleaning dan Lower Casing Data Twitter	63
Gambar 4. 9 Proses Stemming	63
Gambar 4. 10 Proses Klasifikasi	64
Gambar 4. 11 Halaman Utama	65
Gambar 4. 12 Halaman Hasil Analisa Sentimen	65

INTISARI

Teknologi telekomunikasi merupakan sarana yang penting dalam membantu dan memudahkan aktivitas masyarakat pada saat masa pandemi COVID-19 seperti sekarang, dimana jaringan internet merupakan hal yang penting dalam menunjang aktivitas sehari-hari seperti bekerja dari rumah dan belajar dari rumah. Untuk mengukur kualitas layanan yang diberikan salah satunya adalah dengan menilai komentar-komentar yang diberikan pengguna di media sosial twitter menggunakan algoritma *Support Vector Machine*.

Algoritma Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan metode pembelajaran mesin (*supervised learning*) yang memprediksi kelas berdasarkan pola dari hasil proses *training* yang diciptakan. Klasifikasi dilakukan dengan membuat garis pembatas (*hyperlane*) yang memisahkan antara kelas yang ditentukan yaitu positif dan negatif. Tingkat akurasi dan keberhasilan SVM ditentukan oleh kualitas dataset yang digunakan pada saat *training* dan kernel yang digunakan. Kernel linear pada SVM baik digunakan untuk memisahkan data yang bersifat linear seperti sentimen positif dan negatif.

Proses Analisa sentiment dimulai dengan mencari dataset. Dataset ini digunakan untuk proses pembelajaran sistem. Data twitter yang masih dalam bentuk teks kemudian diubah menjadi data vector menggunakan metode pembobotan TF-IDF. Setelah mendapatkan nilai vector dari masing-masing data twitter Langkah selanjutnya adalah melakukan proses klasifikasi menggunakan algoritma *Support Vector Machine*. Hasil pada proses klasifikasi ini nantinya akan menghasilkan sebuah model persamaan yang digunakan untuk menentukan sebuah data twitter digolongkan negatif atau positif. Setelah dilakukan evaluasi, akurasi pada SVM dalam penelitian ini cukup baik yaitu mencapai 95% dan berhasil menggolongkan sentimen dari pengguna layanan indosat dengan rincian 69% sentiment negatif dan 31% sentimen positif sehingga dapat disimpulkan bahwa sentiment pengguna layanan Indosat cenderung negatif.

Kata kunci : Analisis Sentimen, data mining, Support Vector Machine, Indosat, Hyperplane.

ABSTRACT

Telecommunications technology is an important tool in assisting and facilitating community activities during the COVID-19 pandemic such as now, where the internet network is important in supporting daily activities such as working from home and learning from home. To measure the quality of services provided, one of them is to assess the comments given by users on Twitter social media using the Support Vector Machine algorithm.

Support Vector Machine (SVM) algorithm is a classification method using a machine learning method (supervised learning) that predicts classes based on the pattern of the results of the training process created. Classification is done by creating a dividing line (hyperplane) that separates between the classes specified, namely positive and negative. The level of accuracy and success of SVM is determined by the quality of the dataset used at the time of training and the kernel used. Linear kernels in SVM are good for separating linear data such as positive and negative sentiments.

The sentiment analysis process starts with searching for a dataset. This dataset is used for the system learning process. Twitter data that is still in the form of text is then converted into vector data using the TF-IDF weighting method. After getting the vector value from each Twitter data, the next step is to carry out the classification process using the Support Vector Machine algorithm. The results of this classification process will later produce an equation model used to determine a Twitter data classified as negative or positive. After evaluating, the accuracy of the SVM in this study is quite good, reaching 95% and successfully classifying the sentiments of Indosat service users with 69 details % negative sentiment and 31% positive sentiment so it can be concluded that the sentiment of Indosat service users tends to be negative.

Keywords: *Sentiment Analysis, data mining, Support Vector Machine, Indosat, hyperplane.*