

**IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK
ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN
PROVIDER INDOSAT**

SKRIPSI



disusun oleh

Dexy Arya

17.11.1368

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS
SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN PROVIDER
INDOSAT**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Dexy Arya

17.11.1368

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK
ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
PELAYANAN PROVIDER INDOSAT**

yang dipersiapkan dan disusun oleh

Dexy Arya

17.11.1368

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Januari 2021

Dosen Pembimbing,

Krisnawati, S.Si, MT

NIK. 190302038

PENGESAHAN

SKRIPSI

**IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK
ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
PELAYANAN PROVIDER INDOSAT**

yang dipersiapkan dan disusun oleh

Dexy Arya

17.11.1368

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Februari 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, M.Kom

NIK. 190302185

Anna Baita, M.Kom

NIK. 190302290

Krisnawati, S.Si., M.T

NIK. 190302038

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 23 Februari 2021

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si., M.T

NIK. 190302038

PERNYATAAN

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, 2 Maret 2021



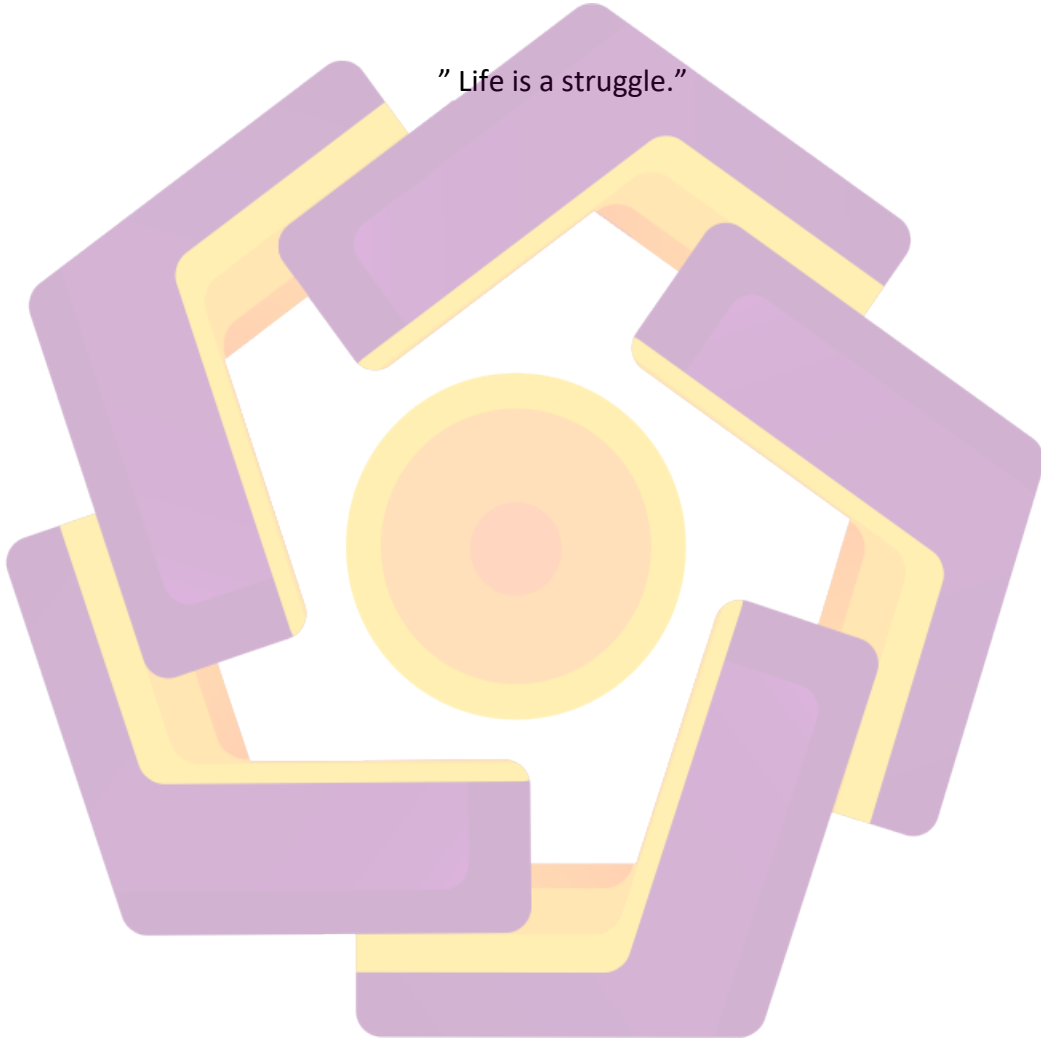
Dexy Arya

NIM 17.11.1368

MOTTO

" your life is a choice."

" Life is a struggle."



PERSEMBAHAN

Puji syukur saya panjatkan kehadiran Allah yang telah memberikan berkat yang luar biasa kepada saya, sehingga saya bisa menyelesaikan skripsi ini dengan baik. Saya juga sangat berterima kasih kepada orang-orang yang telah membantu saya secara langsung maupun tidak langsung dalam menyelesaikan skripsi ini. Skripsi ini saya persembahkan kepada :

1. Bapak dan Ibu yang tidak pernah lelah mendoakan, selalu support baik finansial maupun dalam dukungan lainnya. Terimakasih telah menjadi orang tua yang baik.
2. Ibu Krisnawati, S.Si, M.T selaku dosen pembimbing yang telah membimbing peneliti didalam pembuatan skripsi
3. Serta semua pihak yang telah membantu serta mendukung saya yang tidak bisa saya sebutkan satu persatu.

KATA PENGANTAR

Assalamua'laikum Warohmatullahi Wabarokatuhu

Alhamdulillah, puji syukur penulis panjatkan atas kehadiran Allah Yang Maha Esa yang telah memberikan segala rahmat dan hidayah-Nya sehingga penelitian dengan judul “Implementasi Support Vector Machine (SVM) Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Provider Indosat”.

Dalam Penyusunannya, skripsi yang dibuat merupakan salah satu dari syarat kelulusan bagi setiap mahasiswa Universitas AMIKOM Yogyakarta dan menjadi sebuah bukti bahwa mahasiswa telah mampu menyelesaikan perkuliahan pada jenjang program Strata 1 dan untuk mendapat gelar Sarjana Komputer. Pada kesempatan kali ini penulis ingin mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas AMIKOM Yogyakarta .
2. Ibu Krisnawati, S.Si, M.T selaku dosen pembimbing yang telah membimbing peneliti didalam pembuatan skripsi
3. Bapak dan Ibu Dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmu pengetahuan bagi peneliti dalam proses perkuliahan
4. Dan seluruh pihak yang penulis tidak dapat disebutkan satu-persatu yang telah memberikan seluruh dukungan baik dalam bentuk moril dan material.

Peneliti sendiri memiliki kesadaran bahwa pembuatan skripsi yang dilakukan masih sangat banyak kekurangan dan kelemahan. Oleh karena itu, peneliti berharap kepada semua pihak untuk dapat memberikan kritik dan masukan yang membangun guna menambah kesempurnaan skripsi ini. Dan peneliti tetap memiliki harapan bahwa skripsi yang dibuat memiliki manfaat bagi seluruh pihak yang membacanya.

Wassalamu'alaikum Warahmatullahi Wabarakatuhu

Yogyakarta, _____ 2021

Dexy Arya

17.11.1368

DAFTAR ISI

PERNYATAAN	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
INTISARI	xv
ABSTRAK	xvi
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	3
1.4. Maksud dan Tujuan Penelitian	3
1.5. Metode Penelitian	4
1.5.1 Metode Pengumpulan Data	4
1.5.2 Metode Analisis	5
1.5.3 Metode Perancangan	6
1.6. Sistematika Penulisan	7
BAB II	9
LANDASAN TEORI	9
2.1. Tinjauan Pustaka	9
2.2. Sentiment Analysis	12
2.3. Data Mining	12
2.4. Text Mining	13
2.5. Algoritma Support Vector Machine (SVM)	15
2.6. Twitter	19

2.6.1	Twitter API.....	20
2.7.	Python	21
2.7.1	Pandas	22
2.7.2	NLTK	22
2.8.	Term Frequency	22
2.9.	Confusion Matrix	23
BAB III		24
METODE PENELITIAN		24
3.1.	Alat dan Bahan Penelitian.....	24
3.1.1.	Hardware	24
3.1.2.	Software	24
3.2.	Perancangan Sistem.....	25
3.3.	Metode Pengumpulan Data.....	26
3.4.	Diagram Alir Preprocessing.....	27
3.4.1.	Diagram Alir Case Folding	28
3.4.2.	Tokenizing	29
3.4.3.	Stopword Removing.....	31
3.4.4.	Normalisasi	32
3.4.5.	Stemming	33
3.5.	Pembobotan.....	34
3.5.1.	Term Frequency (TF)	35
3.5.2.	Invers Document Frequency (IDF).....	35
3.5.3.	TF-IDF	36
3.6.	Proses Klasifikasi SVM	37
3.7.	Pembagian Dataset.....	38
BAB IV		40
Hasil Dan Pembahasan		40
4.1.	Proses Crawling	40
4.1.1.	Preprocessing.....	41
4.1.2.	Hasil Cleaning	41

4.1.3.	Hasil Case Folding	43
4.1.4.	Hasil Tokenizing.....	44
4.1.5.	Hasil Stopword Removal	45
4.1.5.	Hasil Proses Normalisasi	46
4.1.5.	Hasil Stemming	48
4.1.6.	Hasil Proses TF-IDF	49
4.2.	Pengujian SVM	52
4.2.1.	Pengujian Pertama.....	52
4.2.2.	Pengujian Kedua.....	54
4.2.3.	Percobaan Ketiga	55
4.2.4.	Percobaan Keempat.....	56
4.2.5.	Percobaan Kelima	57
4.2.6.	Percobaan Ke enam	58
4.2.7.	Percobaan Ketujuh	59
4.3.	Pengujian Kernel	60
BAB V	62
KESIMPULAN DAN SARAN	62
5.1.	Kesimpulan.....	62
5.2.	Saran	62
DAFTAR PUSTAKA	64

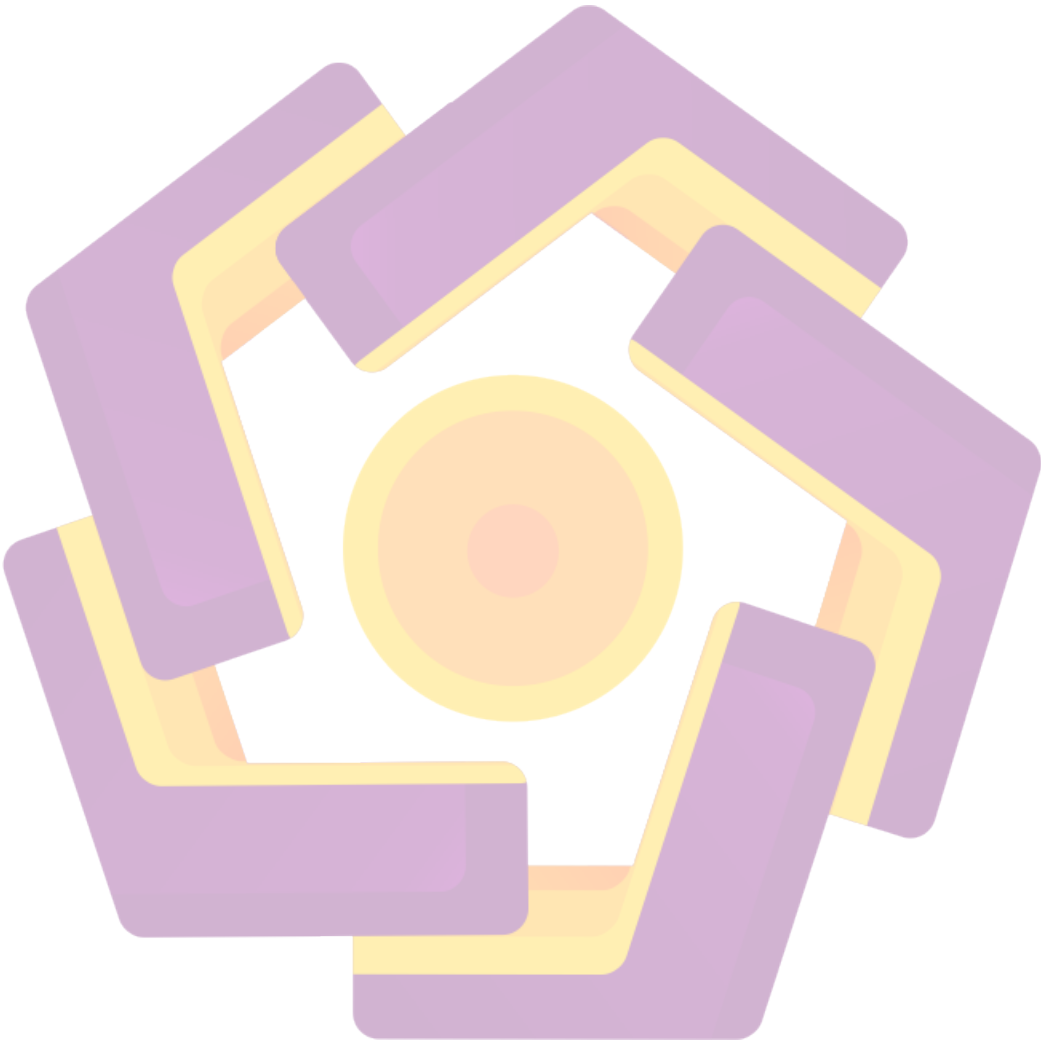
DAFTAR GAMBAR

Gambar 3. 1 Perancangan Sistem	25
Gambar 3. 2 Diagram Alir Preprocessing	27
Gambar 3. 3 Diagram alir Proses Case Folding	29
Gambar 3. 4 Diagram alir Proses Tokenizing	30
Gambar 3. 5 Diagram Alir Stopword Removing	32
Gambar 3. 6 Diagram Alir Normalisasi	33
Gambar 3. 7 Diagram Alir Stemming	34
Gambar 3. 8 Diagram Alir Proses SVM	37
Gambar 4. 1 Implementasi cleaning	42
Gambar 4. 2 implementasi case folding	43
Gambar 4. 3 Implementasi Tokenizing	44
Gambar 4. 4 Implementasi Stopword Removal	45
Gambar 4. 5 Implementasi normalisasi.	47
Gambar 4. 6 Implementasi Stemming	48
Gambar 4. 7 Hasil Stemming	49
Gambar 4. 8 Implementasi Program visualisai kernel	61
Gambar 4. 9 Hasil Visualisasi Kernel	61

DAFTAR TABEL

Tabel 3. 1 Spesifikasi Hardware	24
Tabel 3. 2 Data Tweet	26
Tabel 4. 1 Data Hasil Crawling	40
Tabel 4. 2 Hasil Cleaning Dokumen	42
Tabel 4. 3 case folding	43
Tabel 4. 4 Hasil Tokenizing	44
Tabel 4. 5 Hasil Proses Stopword Removal	46
Tabel 4. 6 Hasil Proses Normalisasi	47
Tabel 4. 7 Nilai TF dan IDF Kata	49
Tabel 4. 8 Hasil TF-IDF	51
Tabel 4. 9 Perbandingan Dataset	52
Tabel 4. 10 Pengujian Pertama	53
Tabel 4. 11 Hasil confusion matrix percobaan pertama.	53
Tabel 4. 12 Pengujian Kedua	54
Tabel 4. 13 Hasil Confusion Matrix Percobaan Kedua	54
Tabel 4. 14 Pengujian Ketiga	55
Tabel 4. 15 Hasil confusion matrix percobaan ketiga	55
Tabel 4. 16 Pengujian Keempat	56
Tabel 4. 17 Hasil confusion matrix percobaan keempat	56
Tabel 4. 18 Pengujian Kelima	57
Tabel 4. 19 Hasil confusion matrix percobaan kelima	57
Tabel 4. 20 Pengujian Ke enam	58
Tabel 4. 21 Hasil confusion matrix percobaan ke enam	58
Tabel 4. 22 Pengujian Ketujuh	59
Tabel 4. 23 Hasil Confusion Matrix dari percobaan Ketujuh	59

Tabel 4. 24 Hasil Pengujian Kernel..... 60



INTISARI

Sosial media merupakan salah satu media yang di gunakan untuk beropin dan mengekspresikan diri bagi setiap penggunanya. Twitter adalah salah satu social media yang sangat populer terutama di Indonesia, pada tahun 2019 pengguna twitter meningkat sebanyak 17 persen yaitu 145 juta pengguna dan Indonesia menjadi salah satu negara yang pertumbuhan pengguna aktif harian paling besar,

Media twitter merupakan media berekspresi setiap pengguna tanpa adanya batasan. tweet yang berupa ekspresi dan apresiasi dari pengguna dapat di gunakan untuk ulasan dari sebuah layanan atau produk. Dalam tugas akhir ini di lakukan analisis sentimen terhadap layanan provider Indosat menggunakan metode *Support Vector Mechine*, pengumpulan data dilakukan dengan menggunakan twitter API, Data set terdiri dari dua data yaitu data training dan data testing, data tersebut akan di gunakan dalam proses pembuatan model *Support Vector Mechine* untuk menguji performa model menggunakan metode *K-Fold Cross Validation* dan *Confusion Matrix*.

Berdasarkan 7 pengujian dengan jumlah data training dan data testing yang berbeda didapatkan akurasi tertinggi pada pengujian ketujuh dengan nilai akurasi 0.81944 atau 82.9%, dari 288 data yang di ujikan terdapat 82 data negatif, 70 data netral dan 84 data positif yang di prediksi secara tepat dan benar, hasil pengujian kernel didapat nilai tertinggi pada kernel polynomial dengan tingkat akurasi sebesar 0.82 Precision 0.83 dan recall 0.82.

Kata Kunci: *Sentimen analysis, SVM, K-Fold Cross Validation, Confusion Matrix, indosat, data maining.*

ABSTRAK

Social media is one of the media that is used to think about and express itself for every user. Twitter is a very popular social media, especially in Indonesia, in 2019 Twitter users increased by 17 percent, namely 145 million users and Indonesia became one of the countries With the biggest daily active user growth,

Twitter is a medium of expression for each user without any restrictions. Tweets in the form of expressions and appreciation from users can be used for reviews of a service or product. In this final project, a sentiment analysis is carried out on Indosat service providers. using the Support Vector Machine method, data collection is done using the twitter API, The data set consists of two data, namely training data and testing data, these data will be used in the process of making a Support Vector Machine model to test the model performance using the K-Fold Cross Validation method. and Confusion Matrix.

Based on 7 tests with different amounts of training data and testing data obtained the highest accuracy in the seventh test with an accuracy value of 0.81944 or 82.9%, of the 288 data tested there were 82 negative data, 70 neutral data and 84 positive data predicted correctly and correctly, kernel test results obtained the highest value in the polynomial kernel with an accuracy rate of 0.82 Precision 0.83 and recall 0.82.

Keyword: *Sentiment analysis, SVM, K-Fold Cross Validation, Confusion Matrix, indosat, data maining.*