

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
ANALISIS SENTIMEN PADA DATA TWITTER TOKOPEDIA**

SKRIPSI



**disusun oleh
Naqiya Zorahima
19.22.2329**

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

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
ANALISIS SENTIMEN PADA DATA TWITTER TOKOPEDIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Naqiya Zorahima
19.22.2329

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

PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA DATA TWITTER TOKOPEDIA

yang dipersiapkan dan disusun oleh

Naqiya Zorahima

19.22.2329

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Februari 2021

Dosen Pembimbing,

Yoga Pristyanto, S.Kom, M.Eng
NIK. 190302412

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA DATA TWITTER TOKOPEDIA

yang dipersiapkan dan disusun oleh

Naqiya Zorahima

19.22.2329

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ainul Yaqin, M.Kom
NIK. 190302255

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

Yoga Prisyanto, S.Kom, M.Eng
NIK. 190302412

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Februari 2021

**DEKAN FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

Hanif Al Fatta, M.Kom
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Naqiya Zorahima
NIM : 19.22.2329

Menyatakan bahwa Skripsi dengan judul berikut:

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA DATA TWITTER TOKOPEDIA

Dosen Pembimbing : Yoga Prisyanto, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Februari 2021

Yang Menyatakan,

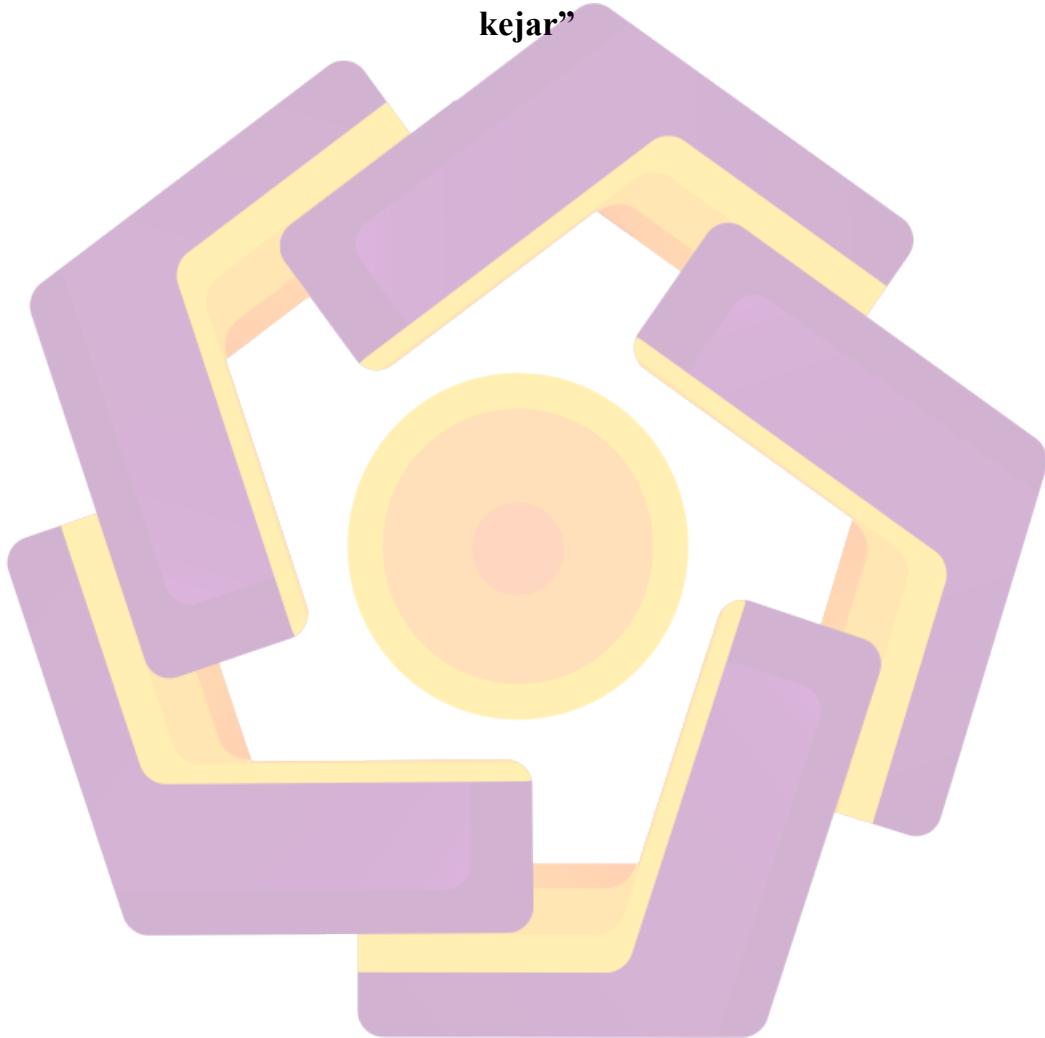


Naqiya Zorahima

MOTTO

“Never give up because great things take time”

“Jika tidak dikejar-kejar hari ini, pasti besok juga akan tetap dikejar-kejar”



PERSEMBAHAN

Skripsi ini saya persembahkan untuk orang tua saya dan diri saya sendiri. Untuk dukungan dari orang tua saya disetiap proses yang saya lalui untuk mencapai titik ini. Kasih sayang, semangat dan dukungan selama saya menyelesaikan kewajiban saya sebagai mahasiswa. Terima kasih untuk diri saya sendiri karena sudah kuat dan telah bekerja keras bahwa apa pun itu pasti semua akan terlewati. Skripsi ini sebagai tanda bahwa dukungan dari orang tua saya dan diri saya tidak akan sia-sia. Semoga Allah selalu memberikan kesehatan dan kemudahan dalam segala urusan



KATA PENGANTAR

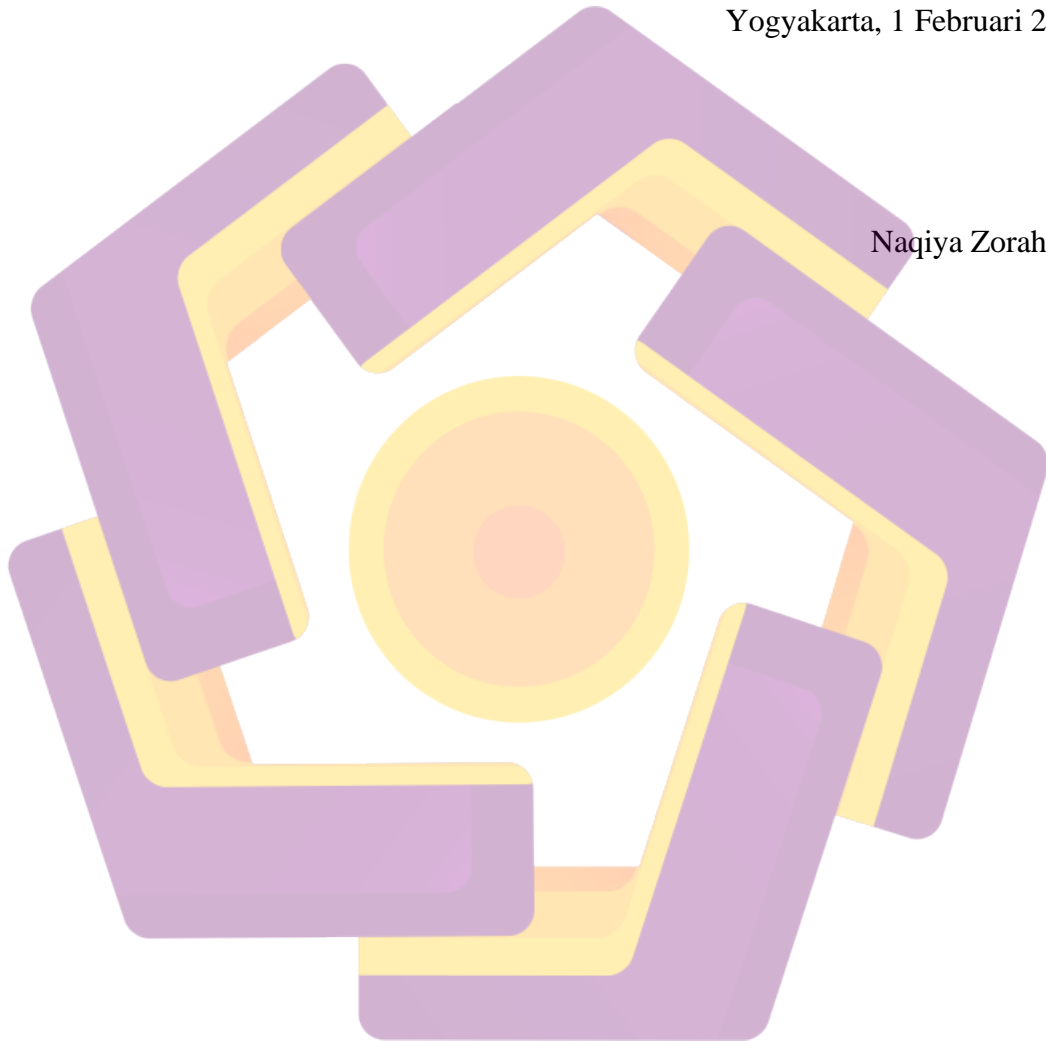
Segala puji bagi Allah SWT karena atas nikmat dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Dalam proses penyelesaian skripsi ini banyak pihak yang telah memberikan bantuan, dukungan, saran dan kritik yang telah penulis terima, oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Allah SWT, yang telah memberikan pertolongan dan kekuatan dalam proses pembuatan skripsi ini.
2. Kedua orang tua penulis, yang selalu mendoakan dan selalu memberikan dukungan kepada penulis dalam menyelesaikan skripsi ini.
3. Bapak Yoga Pristyanto, S.Kom., M.Eng selaku dosen pembimbing skripsi, atas waktu yang telah diberikan untuk membimbing, serta memberikan masukan kepada penulis dalam pengerjaan skripsi ini hingga akhir.
4. Seluruh dosen Fakultas Ilmu Komputer Universitas Amikom Yogyakarta atas segala ilmu pengetahuan yang diberikan kepada penulis.
5. Reka Alamsyah atas dukungan selama penulis menyelesaikan skripsi.

Berbagai kekurangan dan kesalahan mungkin pembaca temukan dalam penulisan skripsi ini, untuk itu kritik dan saran sangat diharapkan untuk perbaikan yang akan datang. Semoga apa yang menjadi kekurangan bisa disempurnakan dan semoga skripsi ini dapat bermanfaat dan menambah wawasan serta pengetahuan bagi para pembaca.

Yogyakarta, 1 Februari 2021

Naqiya Zorahima



DAFTAR ISI

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA DATA TWITTER TOKOPEDIA	ii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	v
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR TABEL	xv
DAFTAR GAMBAR.....	xvii
DAFTAR SOURCE CODE.....	xix
INTISARI.....	xx
ABSTRACT	xxi
1. BAB I 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	3
1.5.1 Metode Pengumpulan Data	3
1.5.2 Metode Analisis dan Perancangan	4
1.5.3 Metode Testing.....	5
1.5.4 Metode Implementasi.....	5

1.6	Sistematika Penulisan	6
2.	BAB II LANDASAN TEORI	8
1.1	Tinjauan Pustaka	8
2.1.1	Penelitian Terdahulu	9
2.2	Landasan Teori.....	11
2.2.1	<i>Data Mining</i>	11
2.2.2	<i>Text Mining</i>	11
2.2.3	Sentimen Analisis.....	12
2.2.4	<i>Preprocessing</i>	13
2.2.5	Pembobotan Kata	14
2.2.6	<i>Support Vector Machine</i>	15
2.2.7	<i>Confusion Matrix</i>	17
2.2.8	Twitter	18
2.2.9	Tokopedia.....	20
2.2.10	Twint	20
2.2.11	Flask	21
3.	BAB III METODE PENELITIAN.....	22
2.1	Deskripsi Umum	22
2.2	Kebutuhan Sistem	22
3.2.1	Perangkat Keras.....	22
3.2.2	Perangkat Lunak.....	22
3.3	Perancangan Desain Sistem	23
3.4	Alur Penelitian	23
3.4.1	<i>Scrapping Data Twitter</i>	25

3.4.2	<i>Dataset</i>	25
3.4.3	<i>Labeling Dataset</i>	26
3.4.4	<i>Preprocessing</i>	26
3.4.4.1	<i>Case Folding</i>	26
3.4.4.2	<i>Hashtag, Mention, URL Removal</i>	27
3.4.4.3	<i>Punctuation removal</i>	27
3.4.4.4	<i>Whitespace removal</i>	27
3.4.4.5	<i>Stopword Removal</i>	28
3.4.4.6	<i>Stemming</i>	28
3.4.6	<i>Split Data</i>	29
3.4.7	<i>Algoritma Support Vector Machine</i>	29
3.5	<i>Perancangan</i>	30
3.5.1	<i>Tahap – Preprocessing</i>	30
3.5.2	<i>Tahap – Klasifikasi Sentimen</i>	30
3.6	<i>Implementasi</i>	31
3.7	<i>Pengujian</i>	31
4.	BAB IV HASIL DAN PEMBAHASAN	32
4.1	<i>Implementasi Pengumpulan Data</i>	32
4.2	<i>Implementasi Pelabelan Data</i>	34
4.3	<i>Preprocessing</i>	35
4.3.1	<i>Case Folding</i>	35
4.3.2	<i>Hashtag, Mention,dan URL Removal</i>	37
4.3.3	<i>Punctuation Removal</i>	38
4.3.4	<i>Whitespace Removal</i>	39

4.3.5	<i>Stopword Removal</i>	41
4.3.6	<i>Stemming</i>	42
4.4	Ekstraksi Fitur TF-IDF	43
4.5	Pembagian Data Latih dan Data Uji	44
4.6	Implementasi Algoritma Klasifikasi <i>Support Vector Machine</i>	44
4.6.1	Uji Coba Model – Jumlah Data	45
4.6.2	Uji Coba Model – Kernel SVM	45
4.7	Pengujian Model dengan <i>Confusion Matrix</i>	46
4.7.1	Percobaan Pertama – 500 Data.....	46
4.7.2	Percobaan Kedua – 1000 Data	48
4.7.3	Percobaan Ketiga – 1500 Data	49
4.7.4	Percobaan Keempat – 2000 Data	51
4.7.5	Percobaan Kelima – 2500 Data	52
4.7.6	Percobaan Keenam – 3000 Data	54
4.7.7	Percobaan Ketujuh – 3500 Data.....	55
4.7.8	Percobaan Kedelapan – 4000 Data.....	57
4.7.9	Percobaan Kesembilan – 4500 Data.....	58
4.7.10	Percobaan Kesepuluh – 5000 Data.....	60
4.8	Desain dan Implementasi GUI.....	65
4.8.1	Halaman Utama.....	66
4.8.2	Halaman Proses	66
4.8.3	Halaman Uji	68
5	BAB V PENUTUP.....	72
6.1	Kesimpulan	72

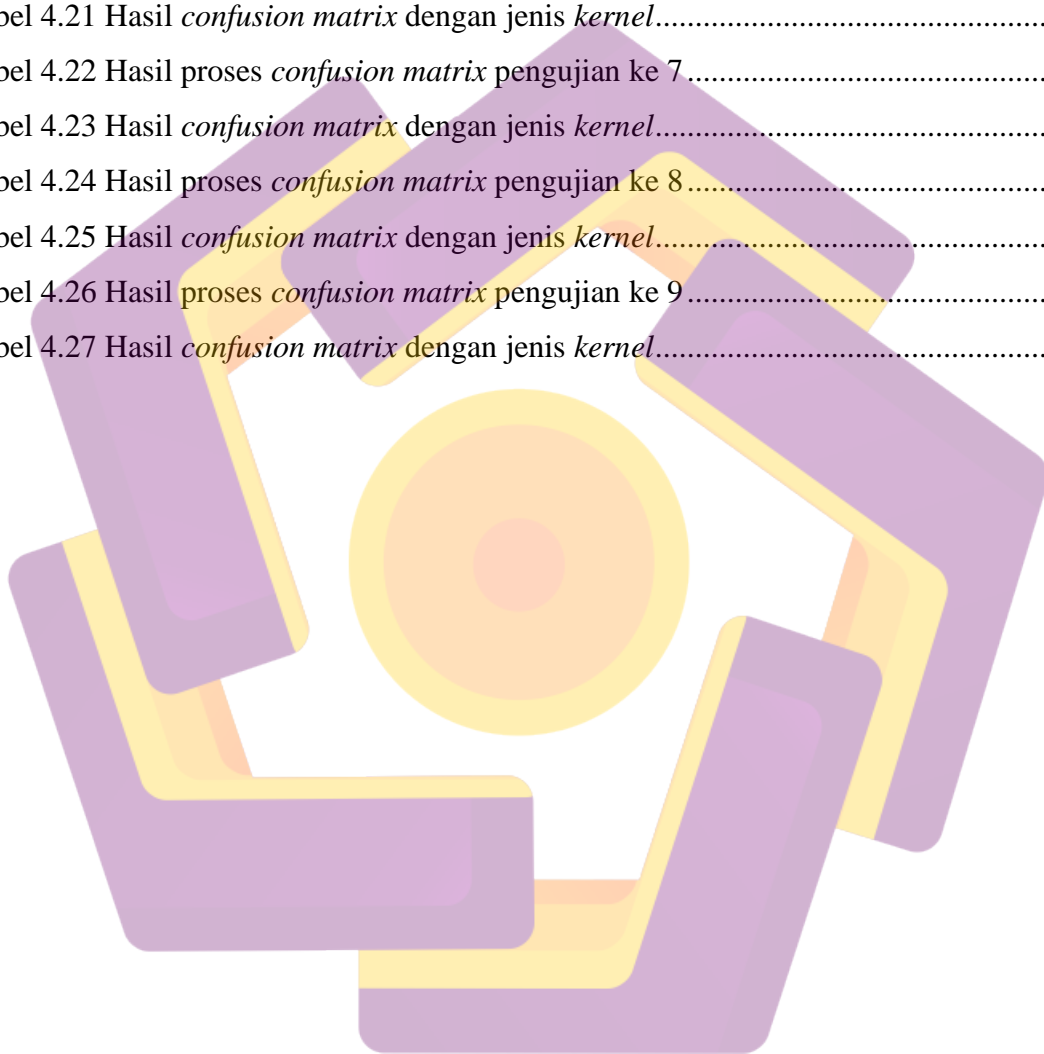
6.2	Saran	73
	DAFTAR PUSTAKA.....	74



DAFTAR TABEL

Tabel 2.1 Studi Literatur Sejenis	9
Tabel 2.2 Contoh <i>Case Folding</i>	13
Tabel 2.3 Contoh Proses Tokenisasi	13
Tabel 2.4 Proses <i>Stemming</i>	14
Tabel 2.5 Tabel <i>Confusion Matrix</i>	18
Tabel 3.1 Proses <i>case folding</i>	26
Tabel 3.2 Proses <i>Hashtag, Mention, URL Removal</i>	27
Tabel 3.3 Proses <i>punctuation removal</i>	27
Tabel 3.4 Proses <i>whitespace removal</i>	28
Tabel 3.5 Proses <i>stopword removal</i>	28
Tabel 3.6 Proses <i>stemming</i>	28
Tabel 4.1 Cuplikan hasil <i>scrapping</i> data <i>twitter</i>	33
Tabel 4.2 <i>Dataset</i> setelah dilakukan pelabelan	34
Tabel 4.3 Hasil proses <i>preprocessing - case folding</i>	36
Tabel 4.4 Hasil proses <i>preprocessing - Hashtag, mention, url removal</i>	38
Tabel 4.5 Hasil proses <i>preprocessing - punctuation removal</i>	39
Tabel 4.6 Hasil proses <i>preprocessing - whitespace removal</i>	40
Tabel 4.7 Hasil setelah <i>preprocessing - stopwords removal</i>	41
Tabel 4.8 Hasil proses <i>stemming</i>	43
Tabel 4.9 Uji coba jumlah data	45
Tabel 4.10 Hasil percobaan <i>confusion matrix</i> uji coba 1	47
Tabel 4.11 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	47
Tabel 4.12 Hasil proses <i>confusion matrix</i> pengujian ke 2	48
Tabel 4.13 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	48
Tabel 4.14 Hasil proses <i>confusion matrix</i> pengujian ke 3	49
Tabel 4.15 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	50

Tabel 4.16 Hasil proses <i>confusion matrix</i> pengujian ke 4.....	51
Tabel 4.17 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	51
Tabel 4.18 Hasil proses <i>confusion matrix</i> pengujian ke 5.....	52
Tabel 4.19 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	53
Tabel 4.20 Hasil proses <i>confusion matrix</i> pengujian ke 6.....	54
Tabel 4.21 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	54
Tabel 4.22 Hasil proses <i>confusion matrix</i> pengujian ke 7.....	55
Tabel 4.23 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	56
Tabel 4.24 Hasil proses <i>confusion matrix</i> pengujian ke 8.....	57
Tabel 4.25 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	57
Tabel 4.26 Hasil proses <i>confusion matrix</i> pengujian ke 9.....	58
Tabel 4.27 Hasil <i>confusion matrix</i> dengan jenis <i>kernel</i>	59



DAFTAR GAMBAR

Gambar 2.1 Tahapan dalam Penyusunan <i>Knowledge Discovery in Database</i> (KDD) ..	11
Gambar 2.2 <i>Hyperplane</i> Memisahkan Dua Kelas	16
Gambar 3.1 Alur Kerja Penelitian	24
Gambar 3.2 Alur kerja <i>scrapping</i> data	25
Gambar 3.3 Rancangan Tampilan Halaman <i>Preprocessing</i>	30
Gambar 3.4 Rancangan Tampilan Halaman Uji.....	31
Gambar 4.1 Persebaran Data	35
Gambar 4.2 Percobaan <i>kernel</i> SVM dengan 500 data.....	47
Gambar 4.3 Percobaan <i>kernel</i> SVM dengan 1000 data.....	49
Gambar 4.4 Percobaan <i>kernel</i> SVM dengan 1500 data.....	50
Gambar 4.5 Percobaan <i>kernel</i> SVM dengan 2000 data.....	52
Gambar 4.6 Percobaan <i>kernel</i> SVM dengan 2500 data.....	53
Gambar 4.7 Percobaan <i>kernel</i> SVM dengan 3000 data.....	55
Gambar 4.8 Percobaan <i>kernel</i> SVM dengan 3500 data.....	56
Gambar 4.9 Percobaan <i>kernel</i> SVM dengan 4000 data.....	58
Gambar 4.10 Percobaan <i>kernel</i> SVM dengan 4500 data.....	59
Gambar 4.11 Percobaan <i>kernel</i> SVM dengan 5000 data.....	61
Gambar 4.12 Grafik percobaan <i>kernel</i> SVM Linear terhadap jumlah data.....	61
Gambar 4.13 Grafik percobaan <i>kernel</i> SVM RBF terhadap jumlah data.....	62
Gambar 4.14 Grafik percobaan <i>kernel</i> SVM Poly terhadap jumlah data.....	62
Gambar 4.15 Grafik perbandingan nilai <i>f1 score</i> pada setiap percobaan	63
Gambar 4.16 Grafik perbandingan nilai akurasi skor pada setiap percobaan	64
Gambar 4.17 Grafik perbandingan nilai presisi pada setiap percobaan	64
Gambar 4.18 Grafik perbandingan nilai <i>recall</i> pada setiap percobaan	65
Gambar 4.19 Tampilan Halaman Utama.....	66
Gambar 4.20 Tampilan Halaman Proses	67
Gambar 4.21 Tampilan Halaman Proses - <i>Preprocessing</i>	67

Gambar 4.22 Tampilan Halaman Proses - Nilai *Confusion Matrix* 68

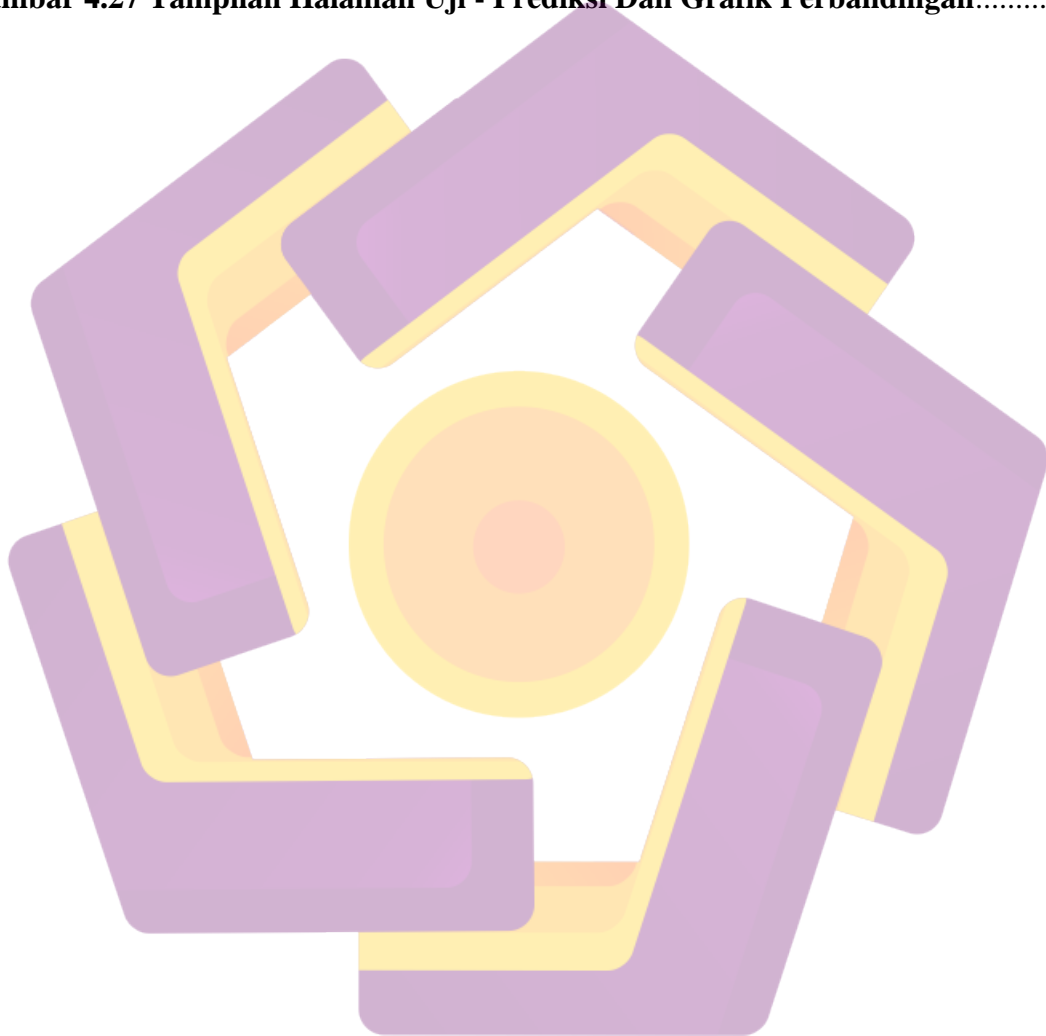
Gambar 4.23 Tampilan Halaman Uji 69

Gambar 4.24 Tampilan Halaman Uji - Pengujian dan Prediksi Kalimat Positif 69

Gambar 4.25 Tampilan Halaman Uji - Pengujian dan Prediksi Kalimat Negatif 70

Gambar 4.26 Tampilan halaman uji - pengujian dan prediksi dengan dataset 70

Gambar 4.27 Tampilan Halaman Uji - Prediksi Dan Grafik Perbandingan..... 71



DAFTAR SOURCE CODE

Source Code 4.1 Menambahkan pustaka <i>twint</i>	32
Source Code 4.2 CLI <i>twint</i> untuk <i>scrapping</i> data Tokopedia	32
Source Code 4.3 Proses <i>case folding</i>	36
Source Code 4.4 Proses <i>Hashtag, Mention, URL Removal</i>	37
Source Code 4.5 Proses <i>punctuation removal</i>	38
Source Code 4.6 Proses <i>whitespace removal</i>	40
Source Code 4.7 Proses <i>stopword</i> removal	41
Source Code 4.8 <i>Stemming</i>	42
Source Code 4.9 Ekstraksi fitur TF-IDF	43
Source Code 4.10 Pembagian data latih dan data uji	44
Source Code 4.11 Klasifikasi <i>Support Vector Machine</i>	44
Source Code 4.12 <i>Confusion Matrix</i>	46

INTISARI

Twitter adalah sebuah layanan jejaring sosial dan mikroblog. *Twitter* memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks dengan 280 karakter. Berdasarkan data yang dilansir oleh Hootsuite pengguna *twitter* di Indonesia mencapai 10 juta pengguna aktif. *Twitter* menjadi salah satu jejaring sosial media untuk berbagi opini dalam bentuk cuitan.

Analisis sentimen merupakan perpaduan dari *text mining* dan *data mining*, atau sebuah metode yang digunakan untuk memahami, mengekstrak pendapat dan mengolah pendapat yang diberikan konsumen atau publik melalui banyak media mengenai sebuah produk, jasa atau sebuah instansi. *Tweet* atau cuitan dibagi menjadi dua kategori tweet dengan sentimen positif dan tweet dengan sentimen negatif sehingga dapat dilihat bagaimana pandangan publik dari cuitan yang diberikan. Metode yang digunakan pada penelitian ini adalah *Support Vector Machine* dengan sumber data yang digunakan adalah data *Twitter* Tokopedia. Metode *Support Vector Machine* ini dilakukan dengan pengumpulan data, processing, pembobotan data dan pembuatan model klasifikasi.

Pada penelitian yang dilakukan yaitu membuat sebuah sistem yang mampu mengklasifikasikan cuitan kedalam kategori sentimen positif atau negatif. Penelitian ini terbukti bahwa algoritma *Support Vector Machine* ini dapat melakukan klasifikasi pada data *Twitter* yang diberikan dan dapat menganalisis secara otomatis. Hasil akurasi yang didapatkan melalui tiga pengujian dengan tiga percobaan jumlah data yang berbeda yaitu 500 data, 1000 data, dan 5000 data. Akurasi tertinggi berada pada pengujian ketiga dengan tingkat akurasi sebesar 90%.

Kata Kunci: Analisis Sentimen, *Twitter*, *Support Vector Machine*, Tokopedia

ABSTRACT

Twitter is a social networking service and microblogging. Twitter allows users to send and read text-based messages with up to 280 characters. Based on data reported by Hootsuite users, Twitter reached 10 million active users in Indonesia. Twitter has become one of the social media networks to share opinions in the form of tweets.

Sentiment analysis is a blend of text mining and data mining, or a method used to understand, extract opinions, and process opinions from consumers or the public through many media about a product, service, or institution. Tweets are divided into categories of tweets with positive and negative sentiments so that you can see how the public views the tweets. The method used in this research is Support Vector Machine, with the data source used being data Twitter Tokopedia. Support Vector Machine process by collecting data, processing, weighing data, and creating a classification model.

Researchers create a system that can classify positive or negative sentiment tweets. This research proves that the algorithm Support Vector Machine can classify Twitter data given and analyse it automatically. Accuracy results were obtained through three tests with three trials of different amounts of data, namely 500 data, 1000 data, and 5000 data. The highest accuracy is in the third test, with an accuracy rate of 90%.

Keyword: *Sentiment Analyst, Twitter, Support Vector Machine, Tokopedia*

