

**ANALISIS SENTIMEN TWITTER TERHADAP TOKOPEDIA  
MENGUNAKAN ALGORITMA  
SUPPORT VECTOR MACHINE**

**SKRIPSI**



disusun oleh

**Rino Alvian Nur Fauzy**

**16.11.0325**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2020**

**ANALISIS SENTIMEN TWITTER TERHADAP TOKOPEDIA  
MENGUNAKAN ALGORITMA  
SUPPORT VECTOR MACHINE**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



Disusun oleh

**Rino Alvian Nur Fauzy**

**16.11.0325**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2020**

**PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN TWITTER TERHADAP TOKOPEDIA  
MENGUNAKAN ALGORITMA  
SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

**Rino Alvian Nur Fauzy**

**16.11.0325**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 12 April 2020

**Dosen Pembimbing,**

**Anggit Dwi Hartanto, M.Kom**

**NIK. 190302163**

**PENGESAHAN**  
**SKRIPSI**  
**ANALISIS SENTIMEN TWITTER TERHADAP TOKOPEDIA**  
**MENGGUNAKAN ALGORITMA**  
**SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

**Rino Alvian Nur Fauzy**

**16.11.0325**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 16 April 2020

**Susunan Dewan Penguji**

**Nama Penguji**

**Arif Dwi Laksito, M.Kom**  
**NIK. 190302150**

**Mardhiya Hayaty, S.T., M.Kom**  
**NIK. 190302108**

**Anggit Dwi Hartanto, M.Kom.**  
**NIK. 190302163**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 16 April 2020

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom, M.Kom**  
**NIK. 190302096**

## 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, 1 Mei 2020



Rino Alvian Nur Fauzy

NIM. 16.11.0325

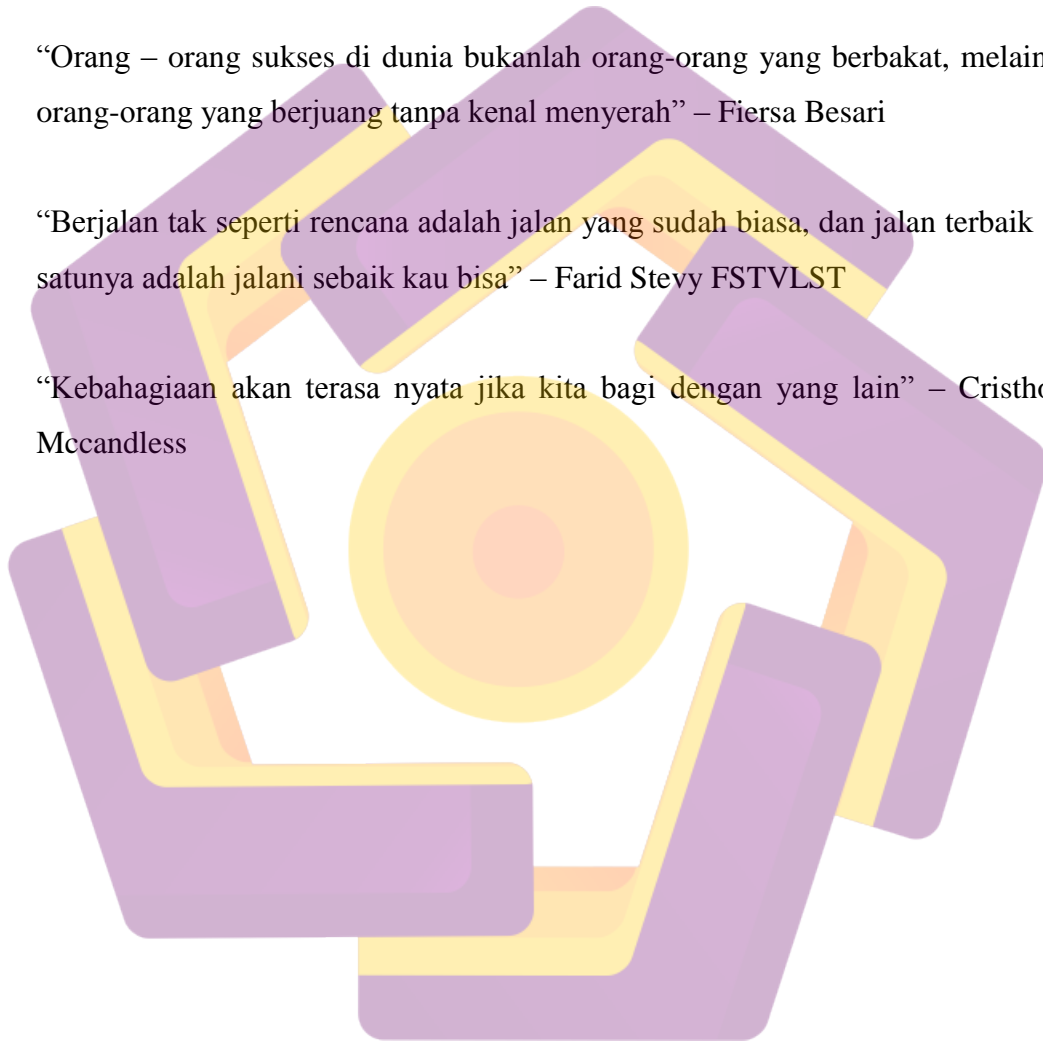
## MOTTO

“Manusia adalah proses belajar tiada henti dari lahir hingga ke liang lahat. Jadi, tidak perlu lagi mencari jati diri. Karena jati diri tidak untuk dicari, melainkan dibentuk.” – Fiersa Besari

“Orang – orang sukses di dunia bukanlah orang-orang yang berbakat, melainkan orang-orang yang berjuang tanpa kenal menyerah” – Fiersa Besari

“Berjalan tak seperti rencana adalah jalan yang sudah biasa, dan jalan terbaik satu satunya adalah jalani sebaik kau bisa” – Farid Stevy FSTVLST

“Kebahagiaan akan terasa nyata jika kita bagi dengan yang lain” – Cristhoper Mccandless



## 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 telah memberikan rahmat dan membimbing saya selama menempuh jalan-Nya.
2. Orang tua saya yang selalu menjadi tempat saya berpulang, selalu mampu menjadi tumpuan, memberikan semangat dan doa yang tiada henti agar saya berhasil mewujudkan apa yang saya citakan.
3. Bapak Anggit Dwi Hartanto, 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 kelas IF 05 2016, yang menemani saya selama menempuh pendidikan di kampus ungu.
6. Kawan CMWW dan Omeng yang menjadi keluarga kedua saya selama berada di Jogja. Memberikan kenangan canda tawa dan susah bersama. Semoga kalian berbahagia dengan jalan yang kalian tempuh masing-masing.
7. Manusia – manusia baik yang tidak mampu saya sebutkan satu per satu. Sungguh semangat, keyakinan dan harapan yang kalian tujukan kepada diri saya sangat berarti.

## KATA PENGANTAR

Puji dan syukur panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Analisis Sentimen Terhadap Tokopedia Menggunakan Algoritma Support Vector Machine.

Skripsi ini saya buat guna menyelesaikan studi jenjang Starta 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 Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing yang selalu bijaksana dalam memberikan bimbingan, nasehat serta waktunya selama pembuatan skripsi ini.
4. Dosen penguji Ibu Mardhiya Hayaty, S.T., M.Kom, Bapak Arif Dwi Laksito, M.Kom dan segenap Dosen serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
5. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat kepada saya.
6. Manusia - manusia baik yang senantiasa mendoakan, memberi semangat dan bantuan kepada saya.
7. Teman – teman CMWW dan OMENG yang selalu memberi semangat dan bantuan kepada saya.



8. Teman – teman IF 05 2016 yang telah menemani selama proses perkuliahan.
9. Teman – teman kontrakan yang selalu memberikan semangat dan masukan dalam pembuatan skripsi.
10. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Allah SWT memberikan balasan yang lebih kepada semua yang telah ikut membantu saya hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik membangun diterima dengan senang hati dan terima kasih. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 1 Mei 2020

Rino Alvian Nur Fauzy

## DAFTAR ISI

JUDUL .....	i
PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN.....	iv
MOTTO .....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL .....	xii
DAFTAR GAMBAR.....	xiii
INTISARI .....	xiv
ABTRACT .....	xv
PERNYATAAN.....	ERROR! BOOKMARK NOT DEFINED.
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
<b>1.1 LATAR BELAKANG.....</b>	<b>1</b>
<b>1.2 RUMUSAN MASALAH .....</b>	<b>2</b>
<b>1.3 BATASAN MASALAH .....</b>	<b>2</b>
<b>1.4 MAKSUD DAN TUJUAN PENELITIAN .....</b>	<b>3</b>
<b>1.5 MANFAAT PENELITIAN.....</b>	<b>3</b>
<b>1.6 METODE PENELITIAN.....</b>	<b>3</b>
<b>1.7 SISTEMATIKA PENULISAN .....</b>	<b>6</b>
<b>BAB II LANDASAN TEORI .....</b>	<b>8</b>
<b>2.1 KAJIAN PUSTAKA .....</b>	<b>8</b>

2.2	<b>DATA MINING</b> .....	11
2.3	<b>TEXT MINING</b> .....	12
2.4	<b>ANALISIS SENTIMEN</b> .....	13
2.5	<b>TEXT PREPROCESSING</b> .....	13
2.5.1	<b>CLEANSING</b> .....	14
2.5.2	<b>CASE FOLDING</b> .....	14
2.5.3	<b>TOKENIZING</b> .....	14
2.5.4	<b>STEMMING</b> .....	14
2.6	<b>PEMBOBOTAN KATA</b> .....	15
2.7	<b>SUPPORT VECTOR MACHINE</b> .....	16
2.8	<b>COSINE SIMILARITY</b> .....	19
2.9	<b>K-FOLD CROSS VALIDATION</b> .....	20
2.10	<b>EVALUASI</b> .....	20
<b>BAB III ANALISIS DAN PERANCANGAN</b> .....		23
3.1.	<b>ANALISIS MASALAH</b> .....	23
3.2.	<b>SOLUSI YANG DIPILIH</b> .....	24
3.3.	<b>ANALISIS KEBUTUHAN</b> .....	24
3.3.1	<b>ANALISIS KEBUTUHAN FUNGSIONAL</b> .....	24
3.3.2	<b>ANALISIS KEBUTUHAN NON-FUNGSIONAL</b> .....	25
3.4	<b>ANALISIS SENTIMEN DENGAN KLASIFIKASI ALGORITMA SVM</b> .....	25
3.4.1	<b>PENGUMPULAN DATASET</b> .....	25
3.4.2	<b>CLEANING DATA</b> .....	26
3.4.3	<b>CASE FOLDING</b> .....	27
3.4.4	<b>TOKENIZATION</b> .....	28
3.4.5	<b>STEMMING</b> .....	28
3.4.6	<b>TF-IDF</b> .....	30
3.4.6	<b>CONSINE SIMILARITY</b> .....	36
3.4.7	<b>SUPPORT VECTOR MACHINE</b> .....	42
3.4.8	<b>K-FOLD CROSS VALIDATION</b> .....	45
3.5.1	<b>DIAGRAM PENGEMBANGAN APLIKASI</b> .....	46

3.5.2	DIAGRAM ALUR PREPROCESSING.....	48
3.5.3	PROSES KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) .....	49
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>51</b>
4.1	DESKRIPSI IMPLEMENTASI.....	51
4.2	IMPLEMENTASI PENGUMPULAN DATA TRAINING .....	51
4.3	IMPLEMENTASI PELABELAN DATA TRAINING.....	53
4.4	IMPLEMENTASI PERANCANGAN FUNGSI DAN SISTEM .....	54
4.4.1	MENGAKSES DATASET TRAINING .....	54
4.4.2	IMPLEMENTASI FUNGSI <i>PREPROCESSING</i> DAN <i>STEMMING</i> DATA TRAINING .....	54
4.4.3	IMPLEMENTASI TF-IDF .....	54
4.4.4	IMPLEMENTASI FUNGSI KLASIFIKASI SUPPORT VECTOR MACHINE..	55
4.4.5	MENGAKSES DATA TESTING .....	55
4.4.6	IMPLEMENTASI FUNGSI <i>CLEANING</i> DATA TESTING .....	56
4.4.7	IMPLEMENTASI FUNGSI TOKENIZING .....	57
4.4.8	IMPLEMENTASI FUNGSI <i>STEMMING</i> DATA TESTING.....	57
4.4.9	IMPLEMENTASI FUNGSI PREDIKSI SENTIMEN.....	58
4.5	PEMBUATAN ANTARMUKA .....	59
4.5.1	HALAMAN UTAMA .....	59
4.5.2	HALAMAN TESTING .....	60
4.5.3	HALAMAN HASIL KLASIFIKASI.....	61
4.6	EVALUASI PROGRAM.....	61
<b>BAB V PENUTUP .....</b>		<b>67</b>
5.1	KESIMPULAN.....	67
5.2	SARAN .....	67
<b>DAFTAR PUSTAKA .....</b>		<b>68</b>

## Daftar Tabel

Tabel 2.1 Tabel Perbanding Penelitian Terdahulu.....	10
Tabel 2.2 Tabel Confusion Matrix.....	20
Tabel 3.1 Data Tweet Hasil Crawling.....	26
Tabel 3.2 Data Tweet Setelah Proses Cleansing.....	26
Tabel 3.3 Data Tweet Setelah Proses Case Folding.....	27
Tabel 3.4 Data Tweet Setelah Proses Tokenization.....	28
Tabel 3.5 Data Tweet Setelah Proses Stemming.....	29
Tabel 3.6 Tabel Feature List.....	29
Tabel 3.7 TF-IDF Positif.....	31
Tabel 3.8 TF-IDF Negatif.....	34
Tabel 3.9 Hasil Consine Similarity Positif.....	37
Tabel 3.10 Hasil Consine Similarity Negatif.....	39
Tabel 3.11 Hasil Consine Similarity.....	41
Tabel 3.12 Perhitungan Persamaan Positif Terhadap Training Data Negatif.....	44
Tabel 3.13 Persamaan $w_1$ , $w_2$ dan $b$ Pada Setiap Label.....	44
Tabel 3.14 Perhitungan Persamaan Negatif Terhadap Data Training Positif.....	44
Tabel 3.15 Prediksi Pada Data Training.....	45
Tabel 3.16 Fold Pertama.....	45
Tabel 3.17 Fold Kedua.....	45
Tabel 3.18 Fold Ketiga.....	45
Tabel 4.1 Confusion Matrix SVM.....	62
Tabel 4.2 Hasil K-Fold Cross Validation.....	63

## Daftar Gambar

Gambar 2.1 (a) Mencari Hyperplane Paling Optimal (b) Hyperplane Optimal.....	17
Gambar 2.2 Memaksimalkan Margin Hyperplane.....	17
Gambar 3.1 Diagram Pengembangan Aplikasi.....	46
Gambar 3.2 Diagram Alir Preprocessing .....	48
Gambar 3.3 Diagram Proses SVM.....	49
Gambar 4.1 Script Scraping Twitterscraper.....	51
Gambar 4.2 File Data Training Sebelum Diolah.....	52
Gambar 4.3 File Data Training Setelah Diolah.....	53
Gambar 4.4 Script Akses Data Training.....	54
Gambar 4.5 Script Preprocessing dan Stemming Data Testing.....	54
Gambar 4.6 Script Implentasi TF-IDF.....	54
Gambar 4.7 Script Klasifikasi Support Vector Machine.....	55
Gambar 4.8 Script Crawling Data Testing.....	56
Gambar 4.9 Script Preprocessing Data Testing.....	57
Gambar 4.10 Script Tokenizing.....	57
Gambar 4.11 Script Stemming Data Testing.....	58
Gambar 4.12 Script Preprocessing dan Stemming Data Testing.....	58
Gambar 4.13 Tampilan Halaman Utama.....	59
Gambar 4.14 Tampilan Hasil Testing.....	60
Gambar 4.15 Tampilan Halaman Hasil Klasifikasi.....	61
Gambar 4.16 Tampilan Halaman Hasil Klasifikasi.....	63

## INTISARI

Situs jual beli online semakin diminati oleh banyak orang dari berbagai usia. Kemudahan dalam berbelanja menjadikan banyak situs jual beli berlomba-lomba dalam meningkatkan kualitas demi memberikan pengalaman yang dirasa cocok oleh pengguna. Tokopedia masih menempati peringkat pertama pada kuartal ketiga tahun ini sebagai e-commerce yang paling banyak diakses sejak kuartal ketiga tahun 2018. Demi menjaga performa dan meningkatkan pelayanan, dibutuhkan masukan atau opini dari pengguna untuk mengetahui kelemahan yang dapat dijadikan bahan evaluasi dan peningkatan pelayanan. Calon pengguna e-commerce juga butuh bahan pertimbangan untuk memilih mana aplikasi yang dirasa paling tepat mendukung keberlangsungan usahanya.

Oleh karena itu, penelitian ini mencoba melakukan analisis sentimen mengenai situs jual beli online khususnya pada Tokopedia. Pengumpulan dataset akan dilakukan menggunakan package dari python yang berjumlah 600 data. Selanjutnya data tersebut memasuki tahap preprocessing untuk meningkatkan akurasi data dan mengurangi noise. Setelah itu data akan dihitung tingkat kemiripan dokumennya menggunakan pendekatan TF-IDF dan Algoritma Support Vector Machine untuk mengetahui klasifikasi dari setiap data.

Berdasarkan pengujian yang sudah dilakukan, tingkat akurasi yang dihasilkan mencapai 72% dengan komposisi data latih berjumlah 50% dan komposisi data uji berjumlah 50%.

**Keywords** – SVM, Analisis Sentimen, Cosine Similarity, TF-IDF

## **ABSTRACT**

*Marketplace are increasingly in demand by many people of various ages. Ease of shopping makes some marketplace competing in improving quality in order to provide an experience that feels suitable by the user. Tokopedia still ranked first in the third quarter of this year as the most widely accessed e-commerce since the third quarter of 2018.*

*In order to maintain performance and improve service, input or opinions from users are needed to determine weaknesses that can be used as material for evaluation and service improvement. Prospective e-commerce users also need consideration to choose which application is most appropriate to support the sustainability of their business.*

*Therefore, this research tries to do sentiment analysis about online buying and selling sites especially on Tokopedia. The collection of datasets will be done using a package of python, amounting to 600 data. Furthermore, the data enters the preprocessing stage to improve data accuracy and reduce noise. After that the data will be calculated the level of similarity of the document using the TF-IDF approach and Support Vector Machine Algorithm to determine the classification of each data. Based on the tests that have been done, the resulting level of accuracy reaches 72% with the composition of training data at 50% and the composition of the test data at 50%.*

**Keywords** – SVM, Analisis Sentimen, Cosine Similarity, TF-IDF