

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN
PENGGUNA LAYANAN PEMESANAN TIKET PESAWAT DAN HOTEL
TRAVELOKA PADA TWITTER MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ARIF GUNAWAN

16.11.0071

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN
PENGGUNA LAYANAN PEMESANAN TIKET PESAWAT DAN
HOTEL TRAVELOKA PADA TWITTER MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ARIF GUNAWAN

16.11.0071

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA
LAYANAN PEMESANAN TIKET PESAWAT DAN HOTEL
TRAVELOKA PADA TWITTER MENGGUNAKAN
ALGORITMA SUPPORT VECTOR
MACHINE (SVM)**

yang dipersiapkan dan disusun oleh

Arif Gunawan

16.11.0071

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 November 2022

Dosen Pembimbing,

Akhmad Dahlan, M.Kom

NIK. 190302174

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA
LAYANAN PEMESANAN TIKET PESAWAT DAN HOTEL
TRAVELOKA PADA TWITTER MENGGUNAKAN
ALGORITMA SUPPORT VECTOR
MACHINE (SVM)**

yang dipersiapkan dan disusun oleh

Arif Gunawan

16.11.0071

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 November 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arifiyanto Hadinegoro, S.Kom, MT
NIK. 190302289

Firman Asharudin, M. Kom
NIK. 190302315

Windha Mega Pradnya D, M.Kom
NIK. 190302185

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 November 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Arif Gunawan
NIM : 16.11.0071

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Terhadap Tingkat Kepuasan Pengguna Layanan Pemesanan Tiket Pesawat dan Hotel Traveloka Pada Twitter Menggunakan Algoritma Support Vector Machine (SVM)

Dosen Pembimbing : Akhmad Dahlan, M.Kom

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, 21 November 2022

Yang Menyatakan,



Arif Gunawan

HALAMAN 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 sayangnya sampai sejauh ini.
2. Kedua orang tua saya dan keluarga, yang selalu mendoakan, selalu menyemangati dan memenuhi kewajibannya untuk memenuhi kebutuhan anak.
3. Bapak Akhmad Dahlan, 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. Pandu, Erwinda, Bambang dan semua teman dekat yang selalu sabar dan hadir saat dibutuhkan.
6. Teman-teman kelas IF02 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 Pemesanan Tiket dan Hotel Traveloka 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 :

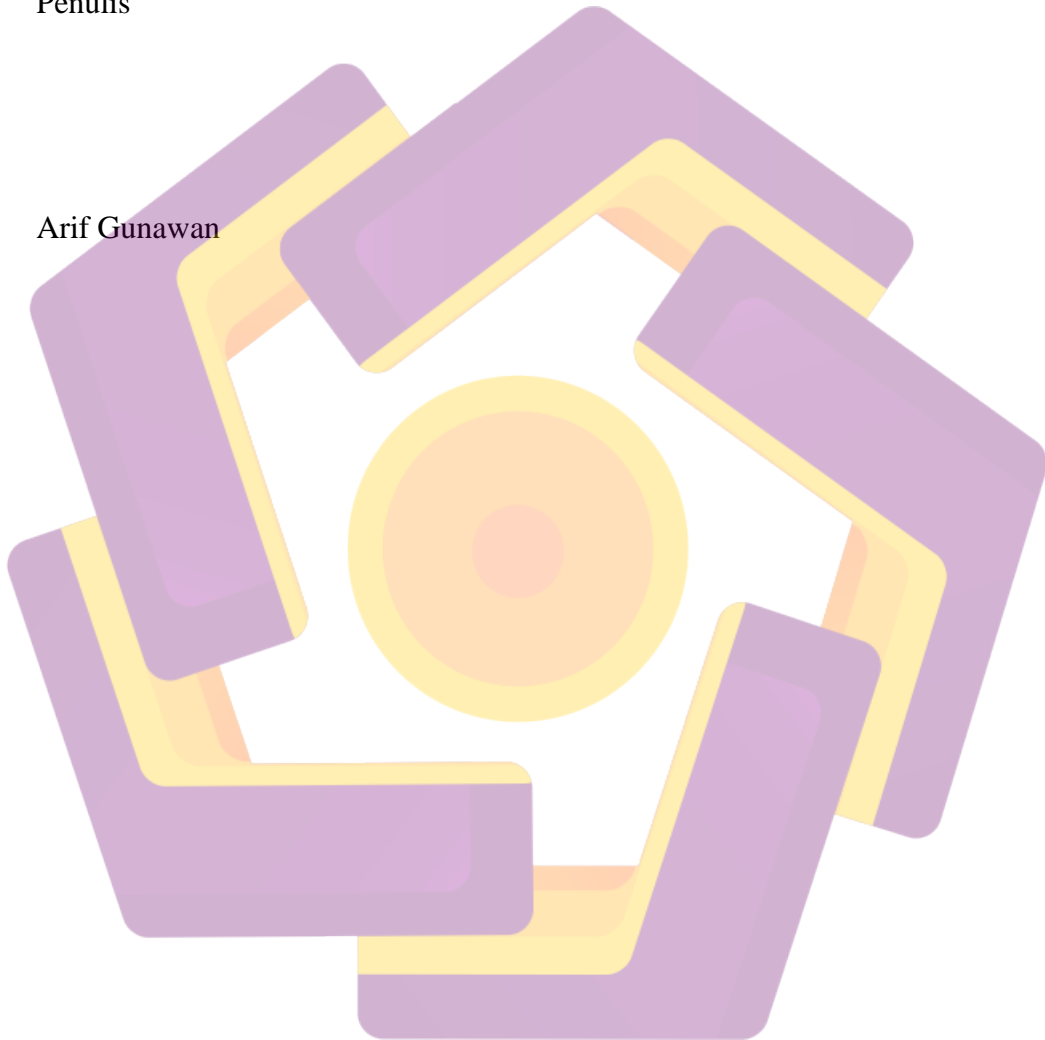
1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Akhmad Dahlan, M.Kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.
4. Dosen Penguji - serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
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 IF02 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 November 2022

Penulis

Arif Gunawan



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
INTISARI	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	7
2.2.1 Analisis Sentimen	7
2.2.2 Text Mining	7
2.2.3 Crawling.....	8
2.2.4 Preprocessing	8
2.2.5 Term Frecquency-Inverse Document Frequency (TF-IDF).....	9
2.2.6 Support Vector Machine	10
2.2.7 Cosine Similarity	12

2.2.8	Confusion Matrix	12
2.2.9	K-Fold Cross Validation	14
2.2.10	Bahasa Pemrograman Python	14
BAB III METODE PENELITIAN		15
3.1	Objek Penelitian.....	15
3.2	Alur Penelitian	15
3.2.1	Pengumpulan Data	15
3.2.2	Preprocessing Data.....	19
3.2.3	Pembobotan TF-IDF	31
3.2.4	Cosine Similarity	46
3.2.5	Support Vector Machine	55
3.2.6	Evaluasi Klasifikasi	63
3.3	Alat dan Bahan.....	65
3.3.1	Data Penelitian	65
3.3.2	Alat Penelitian.....	67
BAB IV HASIL DAN PEMBAHASAN		68
4.1	Deskripsi Sistem	68
4.2	Implementasi Crawling Data Twitter	68
4.3	Preprocessing Data	69
4.4	Implementasi Proses Training Data	70
4.5	Implementasi Proses Prediksi dan Klasifikasi	72
4.6	Evaluasi Program.....	74
BAB V PENUTUP		79
5.1	Kesimpulan	79
5.2	Saran	79
REFERENSI		80

DAFTAR TABEL

Tabel 2.1. Perbandingan metode	6
Tabel 2.2. Confusion Matrix	13
Tabel 3.1 Data ulasan sebelum cleaning data/ remove punctuation	21
Tabel 3.2 Data ulasan setelah cleaning data/ remove punctuation	22
Tabel 3.3 Data ulasan setelah proses Case Folding	24
Tabel 3.4 Hasil tokenisasi	25
Tabel 3.5 Hasil stopword removal	27
Tabel 3.6 Hasil stemming	29
Tabel 3.7 Term positif dan negative	30
Tabel 3.8 Hasil perhitungan term frequency positif	32
Tabel 3.9 Hasil perhitungan IDF seluruh term positif	33
Tabel 3.10 Hasil perhitungan IDF Term Frequency Positif	36
Tabel 3.11 Hasil perhitungan term frequency negative	39
Tabel 3.12 Hasil perhitungan IDF seluruh term negative	41
Tabel 3.13 Hasil perhitungan IDF Term Frequency Negatif	43
Tabel 3.14 Cosine Similiarity komentar positif	48
Tabel 3.15 Cosine Similiarity Komentar Negatif	52
Tabel 3.16 Hasil cosine similiarity	54
Tabel 3.17 Hasil perhitungan persamaan positif	59
Tabel 3.18 Hasil perhitungan persamaan negative	63
Tabel 3.19 Tabel Klasifikasi Data	63
Tabel 3.20 Tabel data Confussion matrix	63
Tabel 4.1 Tabel Confusion Matrix Data training	75
Tabel 4.2 Tabel cross validation	76

DAFTAR GAMBAR

Gambar 2.1. Hyperplane pada Support Vector Machine	11
Gambar 3.1. Alur Penelitian	15
Gambar 3.2. Halaman Twitter Developer	16
Gambar 3.3. Halaman pengajuan API	17
Gambar 3.4. Halaman <i>Terms</i> pengajuan API	18
Gambar 3.5. Tampilan <i>Dashboard Twitter Developer</i>	18
Gambar 3.6. Halaman kode akses API Twitter	19
Gambar 3.7. Tahap preprocessing data	20
Gambar 3.8 Diagram alir cleaning data atau remove punctuation	21
Gambar 3.9 Diagram alir case folding	23
Gambar 3.10 Diagram alir tokenisasi	25
Gambar 3.11 Alur proses stopword removal	27
Gambar 3.12 Alur proses stemming	29
Gambar 3.13 Alur support vector machine	55
Gambar 3.14 Data training mentah	66
Gambar 3.15 Data Training Positif	66
Gambar 3.16 Data Training Negatif	66
Gambar 4.1 Perintah Crawling data twitter	68
Gambar 4.2 Data training hasil crawling dari twitter	70
Gambar 4.3 Data training yang telah diberi label	70
Gambar 4.4 Kode perintah untuk mengakses file data training	71
Gambar 4.5 Kode program untuk melakukan proses preprocessing	71
Gambar 4.6 Perintah pembobotan dan klasifikasi data	72
Gambar 4.7 Perintah untuk melakukan download data twitter	72
Gambar 4.8 Proses Cleaning dan Lower Casing	73
Gambar 4.9 Perintah untuk menjalankan proses stemming	73
Gambar 4.10 Proses klasifikasi secara system	74
Gambar 4.11 Tampilan pengujian Confussion Matrix	75

INTISARI

Dalam era digital seperti yang terjadi saat ini, teknologi memiliki peran penting dalam perancangan produk dan layanan yang berbasis digital. Teknologi berbasis digital dapat ditemui pada setiap aspek kehidupan masyarakat. Saat ini banyak media digital yang memanfaatkan teknologi digital untuk memudahkan manusia mencari informasi dan berkomunikasi. Untuk mengukur kualitas layanan yang diberikan salah satunya adalah dengan menilai komentar-komentar yang diberikan pengguna di media sosial twitter.

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 dan kuantitas 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.

Hasil dari proses klasifikasi ini nantinya akan menghasilkan sebuah model persamaan yang digunakan untuk menentukan apakah sebuah data twitter digolongkan negatif atau positif. Setelah dilakukan evaluasi terhadap data uji, akurasi SVM dalam penelitian ini cukup baik yaitu mencapai 90%. Selanjutnya dilakukan pengujian dengan metode K-Fold Cross Validation dan diperoleh rata-rata tingkat akurasi sebesar 95%.

Kata Kunci : Analisis Sentimen, Data mining, Support Vector Machine, Traveloka, Hyperplane.

ABSTRACT

In the digital era as it is happening today, technology has an important role in designing digital-based products and services. Digital-based technology can be found in every aspect of people's lives. Currently, there are many digital media that utilize digital and information technology to make it easier for humans to find information and communicate. To measure the quality of services provided, one of them is to assess the comments given by users on Twitter social media.

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 (hyperlane) 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 results of this calcification process will later produce an equation model used to determine a Twitter data classified as negative or positive. After evaluating the test data, the accuracy of SVM in this study was quite good, reaching 90%. Furthermore, testing was carried out using the K-Fold Cross Validation method and obtained an average accuracy rate of 95%.

Keyword: *Sentiment Analysis, data mining, Support Vector Machine, Traveloka, Hyperplane.*