



**IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK  
KLASIFIKASI KALIMAT TERHADAP ULASAN APLIKASI TRAVELIO PADA  
GOOGLE PLAY STORE**

**SKRIPSI**

**IRFAN MUHAMMAD GUVIAN**

**NIM. 2010511084**

**PROGRAM STUDI S1 INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2024**



**IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK  
KLASIFIKASI KALIMAT TERHADAP ULASAN APLIKASI TRAVELIO PADA  
GOOGLE PLAY STORE**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Komputer**

**IRFAN MUHAMMAD GUVIAN**

**NIM. 2010511084**

**PROGRAM STUDI S1 INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
2024**

## PERNYATAAN ORISINALITAS

### PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri serta semua sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Irfan Muhammad Guvian  
NIM : 2010511084  
Tanggal : 20 Maret 2024

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 20 Maret 2024

Yang Menyatakan,

  
  
Irfan Muhammad Guvian

## PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

### PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Irfan Muhammad Guvian  
NIM : 2010511084  
Fakultas : Ilmu Komputer  
Program Studi : S1 Informatika

Demi pembangunan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

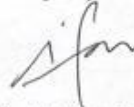
#### **Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Kalimat Terhadap Ulasan Aplikasi Travelio Pada Google Play Store**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta  
Pada Tanggal : 20 Maret 2024

Yang menyatakan,



Irfan Muhammad Guvian

## LEMBAR PENGESAHAN

### LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Irfan Muhammad Guvian  
NIM : 2010511084  
Program Studi : S1 Informatika  
Judul : Implementasi Algoritma *Support Vector Machine* Untuk Klasifikasi  
Kalimat Terhadap Ulasan Aplikasi Travelio Pada Google Play Store

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Nur Hafifah Matondang, S.Kom., M.M.

Penguji I



Theresia Wati, S.Kom., MTI

Penguji II



Neny Rosmawarni, M.Kom.

Pembimbing I



Muhammad Panji Muslim, S.Pd., M.Kom

Pembimbing II



Prof. Dr. Supriyanto, ST., M.Sc., IPM

Dekan



Dr. Widya Cholil, M.I.T

Ketua Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 27 Mei 2024

# IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI KALIMAT TERHADAP ULASAN APLIKASI TRAVELIO PADA GOOGLE PLAY STORE

Irfan Muhammad Guvian

## ABSTRAK

Penelitian ini memiliki latar belakang yaitu adanya perbedaan antara ulasan yang diberikan oleh pengguna aplikasi Andorid dengan rating yang diberikan. Maka dari itu tujuan dari penelitian ini adalah membangun model klasifikasi kalimat yang dapat mengklasifikasikan sentimen ulasan pengguna aplikasi Travelio pada laman Google Play Store menjadi 3 kelas sentimen, positif, negatif, dan netral menggunakan algoritma *Support Vector Machine*. Selain itu penelitian ini dilakukan untuk mengetahui hasil evaluasi klasifikasi model yang sudah dibangun dengan *confusion matrix* dan melakukan perbandingan hasil klasifikasi dengan rating. Tahapan yang dilakukan untuk membangun model meliputi pengumpulan data, praproses data, pelabelan otomatis menggunakan algoritma VADER, pembobotan kata menggunakan TF-IDF serta melakukan *hyperparameter* menggunakan *Grid Search* dengan parameter *Cost* (C) 1,10,100 dan *gamma* 0.1 sampai dengan 1. Pada penelitian ini dilakukan perbandingan tiga komposisi yaitu 70:30, 80:20, dan 90:10 dengan tujuan untuk mengetahui komposisi terbaik berdasarkan keseimbangan model dalam melakukan klasifikasi yang ditunjukkan pada hasil *confusion matrix* yang dilambangkan pada besarnya *F1-Score*. Hasil klasifikasi berdasarkan *confusion matrix* menunjukkan komposisi terbaik didapat menggunakan komposisi 70:30 dengan nilai *accuracy* sebesar 89,8%, *precision* 82%, *recall* 70% dan *F1-Score* 74%. Penelitian ini juga menemukan berdasarkan hasil klasifikasi bahwa rating belum dapat dijadikan parameter kepuasan pelanggan karena masih memuat sentimen yang tidak sesuai.

**Kata Kunci:** *Support Vector Machine*, Travelio, VADER, *confusion matrix*

**IMPLEMENTATION OF THE SUPPORT VECTOR MACHINE ALGORITHM FOR  
SENTENCE CLASSIFICATION ON TRAVELIO APPLICATION REVIEWS ON GOOGLE  
PLAY STORE**

**Irfan Muhammad Guvian**

**ABSTRACT**

*This research has a background, namely that there is a difference between the reviews given by Android application users and the ratings given. Therefore, this research aims to build a sentence classification model that can classify the sentiment of user reviews of the Travelio application on the Google Play Store page into 3 sentiment classes, positive, negative, and neutral using the Support Vector Machine algorithm. Apart from that, this research was conducted to find out the results of evaluating the classification of models that have been built using the confusion matrix and to compare the classification results with the ratings. The steps taken to build the model include data collection, data preprocessing, automatic labeling using the VADER algorithm, word weighting using TF-IDF, and hyperparameterization using Grid Search with parameters Cost (C) 1,10,100 and gamma 0.1 to 1. In this research, A comparison of three compositions was carried out, namely 70:30, 80:20, and 90:10 to find out the best composition based on the balance of the model in carrying out classification which is shown in the confusion matrix results which are symbolized by the magnitude of the F1-Score. The classification results based on the confusion matrix show that the best composition was obtained using a 70:30 composition with an accuracy value of 89.8%, precision of 82%, recall of 70%, and F1-Score 74%. This research also found based on the classification results that the rating cannot be used as a parameter for customer satisfaction because it still contains inappropriate sentiments.*

**Keywords:** *Support Vector Machine, Travelio, VADER, confusion matrix*

## KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi/tugas akhir dengan judul “Implementasi Algoritma *Support Vector Machine* Untuk Klasifikasi Kalimat Terhadap Ulasan Aplikasi Travelio Pada Google Play Store” sebagai syarat kelulusan dari Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Skripsi ini tidak akan selesai tanpa bantuan dan dukungan dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta
2. Dr. Widya Cholil, M.I.T. selaku Ketua Program Studi S1 Informatika
3. Neny Rosmawarni, M. Kom. selaku Dosen Pembimbing I
4. Muhammad Panji Muslim, S.Pd., M. Kom. selaku Dosen Pembimbing II
5. Almarhumah Yuni Widiastiwi, S. Kom. M.Si selaku dosen pembimbing selama penulisan bab 1 hingga bab 3
6. Bambang Chandra Gunawan dan Evi Arjusnita selaku kedua orang tua yang selalu memberikan kasih sayang, doa, dukungan, dan nasihat
7. Semua pihak yang telah membantu dan tidak dapat disebutkan satu per satu

Dalam penulisan skripsi ini masih banyak kekurangan dan kesalahan, karena itu segala kritik dan saran yang membangun akan menyempurnakan penulisan skripsi ini serta bermanfaat bagi penulis dan para pembaca.

Jakarta, 27 Mei 2024

Penulis,

Irfan Muhammad Guvian



## DAFTAR ISI

PERNYATAAN ORISINALITAS.....	i
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	ii
LEMBAR PENGESAHAN.....	iii
ABSTRAK.....	iv
ABSTRACT .....	v
KATA PENGANTAR.....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	ix
DAFTAR GAMBAR.....	x
DAFTAR PERSAMAAN.....	xi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian .....	3
1.4.1 Manfaat Penulis .....	3
1.4.2 Manfaat Pembaca.....	3
1.4.3 Manfaat Bagi Instansi.....	4
1.5 Batasan Masalah.....	4
1.6 Luaran Penelitian.....	4
1.7 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA .....	6
2.1 Klasifikasi Kalimat.....	6
2.2 Travelio.....	6
2.3 Google Play Store .....	6
2.4 <i>Natural Language Processing</i> .....	7
2.5 Praproses Data.....	7
2.6 Pelabelan <i>Lexicon Based</i> .....	8
2.7 Pembobotan TF-IDF.....	9
2.8 <i>Support Vector Machine</i> .....	10
2.8.1 <i>Hyperplane</i> .....	11
2.8.2 <i>Kernel</i> .....	11

2.9	<i>Confusion Matrix</i> .....	12
2.10	Penelitian Terdahulu.....	14
BAB III METODOLOGI PENELITIAN .....		18
3.1	Alur Penelitian .....	18
3.2	Tahapan Penelitian .....	19
3.2.1	Identifikasi Masalah.....	19
3.2.2	Studi Pustaka .....	19
3.2.3	Pembangunan Model.....	19
3.3	Alat dan Bahan Penelitian .....	28
3.3.1	Hardware .....	28
3.3.2	Software.....	28
3.4	Jadwal Pelaksanaan .....	29
BAB IV HASIL DAN PEMBAHASAN.....		30
4.1	Pengumpulan Data .....	30
4.2	Praproses Data.....	30
4.2.1	Persiapan data.....	31
4.2.2	Normalisasi.....	32
4.2.3	<i>Case Folding</i> .....	34
4.2.4	<i>Stopword Removal</i> dan Perubahan Kata <i>Slang</i> .....	35
4.2.5	<i>Stemming</i> .....	35
4.2.6	Tokenisasi.....	36
4.3	Pelabelan Otomatis Menggunakan <i>Lexicon Based</i> .....	38
4.4	Pembobotan Kata.....	41
4.5	Klasifikasi.....	48
4.6	Evaluasi.....	50
4.6.1	Evaluasi Model dengan <i>Confusion Matrix</i> .....	50
4.6.2	Hasil Klasifikasi Ulasan Travelio.....	52
BAB V KESIMPULAN DAN SARAN .....		54
5.1	Kesimpulan.....	54
5.2	Saran .....	55
DAFTAR PUSTAKA .....		56
DAFTAR RIWAYAT HIDUP .....		59
LAMPIRAN .....		60

## DAFTAR TABEL

Tabel 2. 1 Fungsi kernel .....	11
Tabel 2. 2 Review penelitian terdahulu .....	14
Tabel 2. 3 Perbandingan penelitian terdahulu .....	15
Tabel 3. 1 Parameter fungsi google-play-scrapper.....	20
Tabel 3. 2 Parameter kedua .....	21
Tabel 3. 3 Kandidat hyperparameter .....	25
Tabel 3. 4 Kombinasi kemungkinan hyperparameter .....	26
Tabel 3. 5 Parameter GridsearchCV .....	26
Tabel 4. 1 Contoh data sebelum normalisasi.....	32
Tabel 4. 2 Contoh data hasil normalisasi.....	33
Tabel 4. 3 Contoh data hasil case folding.....	34
Tabel 4. 4 Contoh data hasil stopword removal dan perubahan kata slang.....	35
Tabel 4. 5 Contoh data hasil stemming. ....	35
Tabel 4. 6 Contoh data hasil tokenisasi .....	36
Tabel 4. 7 Contoh kalimat untuk pehitungan kelas sentimen .....	39
Tabel 4. 8 Bobot nilai dari setiap kata .....	39
Tabel 4. 9 Contoh hasil perhitungan menggunakan VADER .....	40
Tabel 4. 10 Persebaran data training dan testing (sumber pribadi) .....	42
Tabel 4. 11 Contoh hasil praproses dan tokenisasi.....	43
Tabel 4. 12 Contoh hasil perhitungan TF-IDF .....	46
Tabel 4. 13 Hasil Hyperparamater GridSearchCV .....	49
Tabel 4. 14 Hasil dari confusion matrix komposisi 80:20.....	50
Tabel 4. 15 Hasil perhitungan setiap kelas pada komposisi 80:20 .....	51
Tabel 4. 16 Hasil confusion matrix berdasarkan komposisi data .....	52

## DAFTAR GAMBAR

Gambar 2. 1 Contoh grafik Implentasi Support Vector Machine (Wu & Gao, 2022) .....	10
Gambar 2. 2 Contoh <i>confusion matrix</i> (Normawati et al, 2021) .....	12
Gambar 3. 1 Alur penelitian .....	18
Gambar 3. 2 Alur pengambilan data menggunakan Google-Play-Scraper .....	20
Gambar 3. 3 Alur praproses data .....	21
Gambar 3. 4 Alur pelabelan data menggunakan algoritma VADER .....	23
Gambar 3. 5 Jadwal Pelaksanaan .....	29

## DAFTAR PERSAMAAN

Persamaan 2. 1 Persamaan Pencarian <i>Compound VADER</i> .....	9
Persamaan 2. 2 Persamaan Mencari <i>Term Frequent (TF)</i> .....	9
Persamaan 2. 3 Persamaan Mencari <i>Inverse Document Frequency (IDF)</i> .....	10
Persamaan 2. 4 Persamaan Mencari <i>TF-IDF</i> .....	10
Persamaan 2. 5 Persamaan Mencari <i>Accuracy</i> .....	13
Persamaan 2. 6 Persamaan Mencari <i>Precisio</i> .....	13
Persamaan 2. 7 Persamaan Mencari <i>Recall</i> .....	13
Persamaan 2. 8 Persamaan Mencari <i>f1-score</i> .....	14