

**ANALISIS SENTIMEN APLIKASI WHOOSH PADA PLAY STORE DAN
APP STORE MENGGUNAKAN SUPPORT VECTOR MACHINE**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Christofer Alexander Kumentas
00000034802

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

**ANALISIS SENTIMEN APLIKASI WHOOSH PADA PLAY STORE DAN
APP STORE MENGGUNAKAN SUPPORT VECTOR MACHINE**



Christofer Alexander Kumentas

00000034802

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Christofer Alexander Kumentas
Nomor Induk Mahasiswa : 00000034802
Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Aplikasi Whoosh pada Play Store dan App Store Menggunakan Support Vector Machine

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 5. November 2024

[met



,-]

(Christofer Alexander Kumentas)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN APLIKASI WHOOSH PADA PLAY STORE DAN APP STORE MENGGUNAKAN SUPPORT VECTOR MACHINE

oleh

Nama : Christofer Alexander Kumentas
NIM : 00000034802
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 22 Oktober 2024. Sidang
Pukul 09.00 s/s 11.00 dan dinyatakan

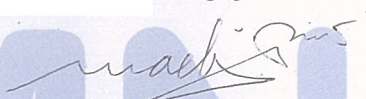
LULUS

Dengan susunan penguji sebagai berikut

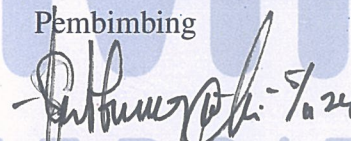
Ketua Sidang


(Dr. Adhi Kusnadi, S.T., M.Si.)
NIDN: 0303037304

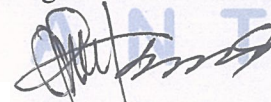
Penguji


(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)
NIDN: 0311106903

Pembimbing


(Fenina Adline Twince Tobing, S.Kom., M.Kom.)
NIDN: 0406038802

Ketua Program Studi Informatika,


(Dr. Adhi Kusnadi, S.T., M.Si.)
NIDN: 0303037304

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Christofer Alexander Kumentas
NIM : 00000034802
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

ANALISIS SENTIMEN APLIKASI WHOOSH PADA PLAY STORE DAN APP STORE MENGGUNAKAN SUPPORT VECTOR MACHINE

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 8 October 2024

Yang menyatakan



Christofer Alexander Kumentas

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Aplikasi Whoosh pada Play Store dan App Store Menggunakan Support Vector Machine dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Dr. Adhi Kusnadi, S.T., M.Si., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Fenina Adline Twince Tobing, S.Kom., M.Kom, sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 8 October 2024


Christofer Alexander Kumentas

ANALISIS SENTIMEN APLIKASI WHOOSH PADA PLAY STORE DAN APP STORE MENGGUNAKAN SUPPORT VECTOR MACHINE

Christofer Alexander Kumentas

ABSTRAK

Whoosh adalah aplikasi transportasi kereta cepat yang dikelola oleh PT KCIC (Kereta Cepat Indonesia China) sebagai bagian dari Proyek Strategis Nasional. Meskipun menjadi proyek besar negara, rating aplikasi Whoosh di platform Google Play Store(2.5) dan App Store(1.7) menunjukkan tingkat ketidakpuasan yang cukup tinggi dari pengguna. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Whoosh dengan menggunakan metode *Support Vector Machine* (SVM). Data ulasan dikumpulkan melalui proses *scraping* dari kedua platform, menghasilkan 555 ulasan. Hasil pelabelan menunjukkan bahwa 73% ulasan bersentimen negatif, dengan keluhan utama terkait proses registrasi, verifikasi email, dan kesulitan login. Sementara itu, 27% ulasan bersentimen positif, umumnya memuji kemudahan dalam melihat informasi kereta. Keluhan yang sering diungkapkan divisualisasikan menggunakan *word cloud*. Data diolah menggunakan metode pembobotan TF-IDF dan *oversampling* dengan SMOTE untuk mengatasi ketidakseimbangan data. Evaluasi model SVM dengan pembagian data latih dan uji 80/20 menghasilkan kinerja tertinggi dengan nilai akurasi 92%, presisi 93%, recall 82%, dan f1-score 87% dibandingkan dengan pembagian data latih dan uji 70/30 dan 60/40.

Kata kunci: Analisis Sentimen, *Support Vector Machine*, Google PlayStore, AppStore, *word cloud*

U M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

***Sentiment Analysis of the Whoosh Application on Play Store and App Store
Using Support Vector Machine***

Christofer Alexander Kumentas

ABSTRACT

Whoosh is a high-speed rail transportation application managed by PT KCIC (Kereta Cepat Indonesia China) as part of the National Strategic Project. Despite being a major national project, the rating of the Whoosh application on the Google Play Store (2.5) and App Store (1.7) indicates a high level of user dissatisfaction. This study aims to analyze user sentiment towards the Whoosh application using the Support Vector Machine (SVM) method. Review data was collected through scraping from both platforms, resulting in 555 reviews. Labeling results showed that 73% of the reviews were negative, with the main complaints related to the registration process, email verification, and login issues. Meanwhile, 27% of the reviews were positive, mostly praising the ease of accessing train information. Frequently mentioned complaints were visualized using a word cloud. The data was processed using the TF-IDF weighting method and oversampling with SMOTE to address data imbalance. The evaluation of the SVM model with an 80/20 train-test split achieved the highest performance with an accuracy of 92%, precision of 93%, recall of 82%, and an f1-score of 87%, compared to 70/30 and 60/40 train-test splits.

Keywords: *Sentiment Analysis, Support Vector Machine, Google PlayStore, AppStore, word cloud*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	6
2.1 Whoosh Kereta Cepat	6
2.2 Analisis Sentimen	6
2.3 Text Mining	7
2.4 Text-Preprocessing	7
2.5 Oversampling SMOTE	9
2.6 Pembobotan TF-IDF	10
2.7 Machine Learning	11
2.8 Support Vector Machine (SVM)	12
2.9 K-Fold Cross Validation	17
2.10 Labeling	18
2.11 Confusion matrix	18
BAB 3 METODOLOGI PENELITIAN	20
3.1 Studi Literatur	20
3.2 Rancangan Sistem	23
3.2.1 Data Scrapping	24
3.2.2 Text-Preprocessing	24
3.2.3 Data Labeling	29
3.2.4 Oversampling SMOTE	30
3.2.5 Pembobotan TF - IDF	30
3.2.6 Data Splitting	30
3.2.7 Pemodelan Support Vector Machine	31
3.2.8 Visualisasi dan Analisis Data	31
3.2.9 Pengujian Model	32
3.2.10 Evaluasi	32
BAB 4 HASIL DAN DISKUSI	33
4.1 Spesifikasi Sistem	33
4.2 Text Mining	33

4.3	Pengolahan Data Ulasan	35
4.4	Translasi Ulasan kedalam Bahasa Indonesia	36
4.5	Text-Preprocessing	37
4.5.1	Data Cleaning	38
4.5.2	Case Folding	39
4.5.3	Filter Stopword	40
4.5.4	Tokenizing	41
4.5.5	Stemming	42
4.5.6	Normalization	43
4.5.7	Data Labeling BERT	44
4.5.8	Sentimen Word Cloud	48
4.6	Data Splitting	49
4.7	Pembobotan TF-IDF	50
4.8	Oversampling SMOTE	58
4.9	Data Modeling SVM	59
4.9.1	Data Modeling SVM Split 80/20	65
4.9.2	Data Modeling SVM Split 70/30	65
4.9.3	Data Modeling SVM Split 60/40	65
4.10	Pengujian Model	66
4.10.1	Hasil Pengujian Model split 80/20	66
4.10.2	Pengujian Model split 70/30	67
4.10.3	Pengujian Model split 60/40	68
4.11	Penyesuaian Threshold	68
4.11.1	Hasil Penyesuaian Threshold split 80/20	69
4.11.2	Hasil Penyesuaian Threshold split 70/30	70
4.11.3	Hasil Penyesuaian Threshold split 60/40	70
4.12	Evaluasi	71
4.12.1	Evaluasi Model SVM Split 80/20	71
4.12.2	Evaluasi Model SVM Split 70/30	76
4.12.3	Evaluasi Model SVM Split 60/40	80
4.13	Hasil dan Pembahasan	85
4.13.1	Pemilahan Data untuk Peningkatan Kinerja Model	85
4.13.2	Perbandingan Kinerja Model SVM Berdasarkan Data Split pada Data Testing	85
4.13.3	Perbandingan Kinerja Model SVM Setelah Penyesuaian Threshold	86
BAB 5	SIMPULAN DAN SARAN	89
5.1	Simpulan	89
5.2	Saran	90
	DAFTAR PUSTAKA	91

DAFTAR GAMBAR

Gambar 2.1	Alur Text Preprocessing	8
Gambar 2.2	Ilustrasi SVM	14
Gambar 3.1	Alur penelitian	23
Gambar 3.2	Tahapan data cleaning	25
Gambar 3.3	Tahapan case folding	26
Gambar 3.4	Tahapan stopwords removal	27
Gambar 3.5	Tahapan tokenization	27
Gambar 3.6	Tahapan stemming	28
Gambar 3.7	Tahapan normalisasi	29
Gambar 3.8	Tahapan klasifikasi SVM	31
Gambar 4.1	Hasil pelabelan sentimen model BERT	46
Gambar 4.2	Hasil mapping data label	47
Gambar 4.3	Distribusi label sentimen	48
Gambar 4.4	Word cloud sentimen positif	49
Gambar 4.5	Word cloud sentimen negatif	49
Gambar 4.6	Matriks TF-IDF pada dokumen	51
Gambar 4.7	Hasil TF-IDF berdasarkan library scikit-learn	58
Gambar 4.8	Hasil classification report untuk split 80/20	67
Gambar 4.9	Hasil classification report untuk split 70/30	67
Gambar 4.10	Hasil classification report untuk split 60/40	68
Gambar 4.11	Classification report penyesuaian threshold split 80/20	69
Gambar 4.12	Classification report penyesuaian threshold split 70/30	70
Gambar 4.13	Classification report penyesuaian threshold - split 60/40	71
Gambar 4.14	Confusion matrix model split 80/20	72
Gambar 4.15	Performance metrics vs threshold split 80/20	74
Gambar 4.16	Confusion matrix model threshold split 80/20	74
Gambar 4.17	Confusion matrix model split 70/30	76
Gambar 4.18	Performance metrics vs threshold split 70/30	78
Gambar 4.19	Confusion matrix model threshold split 70/30	79
Gambar 4.20	Confusion matrix model split 60/40	81
Gambar 4.21	Performance metrics vs threshold split 60/40	83
Gambar 4.22	Confusion matrix model threshold split 60/40	83

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR TABEL

Tabel 2.1	Confusion Matrix	18
Tabel 3.1	Daftar penelitian terkait	20
Tabel 4.1	Perbandingan ulasan sebelum dan sesudah cleaning	39
Tabel 4.2	Perbandingan ulasan sebelum dan sesudah case folding	39
Tabel 4.3	Contoh stopwords removal	41
Tabel 4.4	Contoh hasil dari teks yang di tokenisasi	42
Tabel 4.5	Contoh hasil dari teks yang telah di stemming	43
Tabel 4.6	Contoh hasil dari teks yang telah di normalisasi	44
Tabel 4.7	Term frequency dokumen 1 and dokumen 2	52
Tabel 4.8	Term frequency dokumen 1	53
Tabel 4.9	Term frequency dokumen 2	53
Tabel 4.10	IDF calculation	54
Tabel 4.11	TF-IDF untuk dokumen 1	55
Tabel 4.12	TF-IDF untuk dokumen 2	56
Tabel 4.13	Normalization TF-IDF dokumen 1	57
Tabel 4.14	Normalization TF-IDF dokumen 2	57
Tabel 4.15	Komparasi performa model SVM dengan split berbeda untuk kelas positif dan negatif.	86
Tabel 4.16	Komparasi performa model SVM setelah threshold dengan split berbeda untuk kelas positif dan negatif.	87



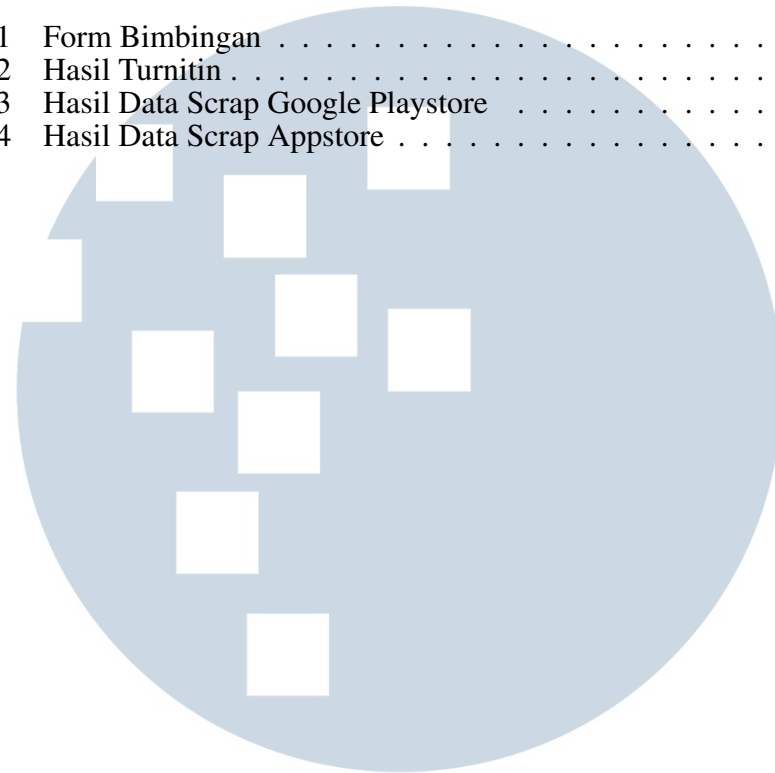
DAFTAR KODE

4.1	Mengambil ulasan dari google play store	34
4.2	Mengambil ulasan dari app store	35
4.3	Mengabungkan kedua dataset menjadi satu	35
4.4	Memeriksa duplikat dan nilai kosong dalam dataframe	36
4.5	Menemukan dan menampilkan baris duplikat dalam dataframe	36
4.6	Mendeteksi bahasa dari setiap teks dalam dataframe	36
4.7	Mengganti baris tertentu dalam dataframe dengan terjemahan	37
4.8	Fungsi untuk memproses teks dengan menghapus karakter yang tidak diinginkan	38
4.9	Mengonversi konten menjadi huruf kecil	39
4.10	Pembuatan daftar stopwords kustom	40
4.11	Fungsi menjalankan stopwords	40
4.12	Fungsi tokenisasi	41
4.13	Menjalankan stemming	42
4.14	Fungsi memuat kamus slang kbba.txt	43
4.15	Normalisasi teks menggunakan kamus slang kbba.txt	44
4.16	Fungsi pelabelan menggunakan model BERT	45
4.17	Mapping kategori sentimen	46
4.18	pemilahan manual hasil pelabelan	47
4.19	Mengubah kategori sentimen	50
4.20	Vektorisasi TF-IDF	50
4.21	Menerapkan SMOTE pada data latih	59
4.22	Hyperparameter tuning dengan gridsearchcv	60
4.23	Parameter terbaik and hasil cross validation	60
4.24	Parameter terbaik data split 80/20	65
4.25	Parameter terbaik data split 70/30	65
4.26	Parameter terbaik data split 60/40	66
4.27	Pengujian model training dengan data testing	66
4.28	Prediksi dengan threshold yang ditetapkan	68

UIN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan	97
Lampiran 2	Hasil Turnitin	99
Lampiran 3	Hasil Data Scrap Google Playstore	100
Lampiran 4	Hasil Data Scrap Appstore	135



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA