



UNIVERSITAS DIPONEGORO

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM)
UNTUK ANALISIS KLASIFIKASI SENTIMEN KOMENTAR APLIKASI
*E-TICKTETING PADA PLAY STORE***

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Teknik**

**ILHAM PUTRA ARIFA
21120116140081**

**FAKULTAS TEKNIK
DEPARTEMEN TEKNIK KOMPUTER
SEMARANG
SEPTEMBER 2020**

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh

Nama : Ilham Putra Arifa
NIM : 21120116140081
Jurusan/Program Studi : Teknik Komputer
Judul Tugas Akhir : Implementasi Algoritma *Support Vector Machine* (SVM) untuk Analisis Klasifikasi Sentimen Komentar Aplikasi *E-ticketing* pada *Play Store*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Teknik pada Jurusan/Program Studi Teknik Komputer, Fakultas Teknik, Universitas Diponegoro.

TIM PENGUJI

Pembimbing I : Agung Budi Prasetyo, S.T., MIT, PhD.
Pembimbing II : Yudi Eko Windarto, S.T., M.Kom.
Ketua Penguji : Dr. R. Rizal Isnanto, S.T., M.M., M.T.
Anggota Penguji : Risma Septiana, S.T., M.Eng.



Semarang, 18 September 2020

Ketua Departemen Teknik Komputer


Dr. R. Rizal Isnanto, S.T., M.M., M.T.

NIP. 197007272000121001

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar

Nama : Ilham Putra Arifa

NIM : 21120116140081

Tanda Tangan :



Tanggal : 18 September 2020

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademika Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Ilham Putra Arifa
NIM : 21120116140081
Jurusan/Program Studi : Teknik Komputer
Fakultas : Teknik
Jenis Karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

Implementasi Algoritma Support Vector Machine (SVM) untuk Analisis Klasifikasi Sentimen Komentar Aplikasi *E-ticketing* pada *Play Store*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini, Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai Penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Semarang

Pada Tanggal : 18 September 2020

Yang menyatakan



Ilham Putra Arifa

KATA PENGANTAR

Puji syukur Alhamdulillah ke hadirat Allah SWT yang telah melimpahkan rahmat-Nya sehingga Penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul “Implementasi Algoritma Support Vector Machine (SVM) untuk Analisis Klasifikasi Sentimen Komentar Aplikasi *E-ticketing* pada *Play Store*” .

Laporan Tugas Akhir ini merupakan salah satu syarat untuk menyelesaikan studi di Departemen Teknik Komputer Fakultas Teknik Universitas Diponegoro. Diharapkan penyusunan laporan Tugas Akhir ini dapat memberikan manfaat dalam bidang Pendidikan.

Dalam penyusunan laporan Tugas Akhir ini Penulis banyak mendapatkan dukungan, doa, bimbingan dan dorongan dari berbagai pihak. Oleh karenanya, melalui kesempatan ini Penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Agung Budi Prasetyo, S.T., MIT, PhD. selaku dosen pembimbing I dan Bapak Yudi Eko Windarto, S.T., M.Kom. selaku dosen pembimbing II yang telah memberikan petunjuk serta bimbingan dalam pembuatan Tugas Akhir.
2. Bapak Dr. R. Rizal Isnanto, S.T., M.M., M.T. selaku Ketua Departemen Teknik Komputer yang telah memberikan dorongan untuk terus belajar dan berkarya.
3. Ibu Ike Pertiwi Windasari, S.T., M.T. selaku Koordinator Tugas Akhir yang telah memberikan petunjuk serta bimbingan pelaksanaan rangkaian kegiatan Tugas Akhir
4. Bapak dan Ibu dosen Departemen Teknik Komputer yang telah memberikan perhatian dan motivasi yang terus menerus sehingga Penulis dapat menyelesaikan Tugas Akhir ini.
5. Kedua orangtua dan keluarga yang selalu mendoakan dan mendukung Penulis.
6. Sahabat-sahabat Penulis yang sudah ikut berkontribusi dalam membuat dataset yang akan digunakan sebagai bahan penelitian Tugas Akhir ini.

7. Teman-teman Teknik Komputer, khususnya angkatan 2016 yang Penulis sayangi dan banggakan.
8. Serta semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu hingga terselesaikannya Tugas Akhir ini.

Penulis menyadari bahwa laporan Tugas Akhir ini masih perlu perbaikan, kritik, saran dan masukan di masa yang akan datang. Penulis berharap laporan Tugas Akhir ini dapat bermanfaat bagi Penulis maupun bagi orang banyak. Akhir kata Penulis mengucapkan terima kasih.

Semarang, 18 September 2020



Ilham Putra Arifa

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
ABSTRAK	xiii
<i>ABSTRACT</i>	xiv
BAB I	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Tugas Akhir.....	3
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan	4
BAB II.....	6
2.1 Penelitian Terdahulu.....	6
2.2 <i>Review Play Store</i>	8
2.3 <i>Sentiment Analysis</i>	8
2.4 <i>Machine Learning</i>	10
2.5 <i>Data mining</i>	11
2.6 <i>Web Scraping</i>	13
2.7 <i>Text Preprocessing</i>	14

2.7.1	<i>Cleansing Text</i>	14
2.7.2	<i>Stemming</i>	14
2.7.3	<i>Filtering</i>	15
2.7.4	<i>Tokenization</i>	15
2.8	Pembobotan	15
2.9	<i>N-Gram</i>	16
2.10	<i>Support Vector Machine</i>	17
2.11	<i>K-fold Cross Validation</i>	20
2.12	Evaluasi Performansi.....	21
2.13	<i>Flask</i>	23
BAB III		24
3.1	Spesifikasi Perangkat.....	24
3.2	Sumber Data	24
3.3	Teknik Analisi Data.....	25
3.4	Langkah Penelitian	26
3.4.1	Pengumpulan Data	26
3.4.2	Melakukan <i>Text Preprocessing</i>	30
3.4.3	Klasifikasi menggunakan <i>Support Vector Machine (SVM)</i>	33
3.4.4	Hasil dan pembahasan.....	34
3.4.5	Implementasi pada aplikasi web	34
BAB IV		37
4.1	Pengumpulan Data.....	37
4.2	<i>Text Preprocessing</i>	38
4.2.1	Melakukan <i>Text Cleansing</i>	38
4.2.2	Melakukan <i>Stemming</i>	41

4.2.3	Melakukan <i>Filtering/Stopword Removal</i>	42
4.2.4	Melakukan <i>Tokenizing</i>	43
4.3	Klasifikasi menggunakan <i>Support Vector Machine</i>	43
4.3.2	Pembobotan Kata	43
4.3.3	Implementasi Model SVM.....	45
4.3.4	Penerapan klasifikasi pada data uji <i>review</i> aplikasi <i>e-ticketing</i>	56
4.4	Hasil dan Pembahasan	60
4.5	Implementasi pada Aplikasi Web.....	64
BAB V.....		68
5.1	Kesimpulan.....	68
5.2	Saran	69
DAFTAR PUSTAKA		70
LAMPIRAN.....		73

DAFTAR GAMBAR

Gambar 2. 1 <i>Data Mining</i> sebagai langkah dalam proses dari <i>Knowledge Discovery</i>	11
Gambar 2. 2 Logo Aplikasi Octoparse.....	13
Gambar 2. 3 SVM berusaha menemukan <i>hyperplane</i> terbaik dari kedua <i>class</i> –1 dan +1.....	18
Gambar 2. 4 <i>K-fold cross validation</i>	20
Gambar 2. 5 Tabel <i>Confusion Matrix</i>	21
Gambar 3. 1 Diagram Alir Sistem.....	26
Gambar 3. 2 Tampilan dashboard aplikasi Octoparse.....	27
Gambar 3. 3 Tampilan pengaturan alur sistem program saat <i>scraping</i> pada aplikasi Octoparse.....	28
Gambar 3. 4 Tampilan saat proses <i>scraping</i> pada aplikasi <i>Octoparse</i>	28
Gambar 3. 5 Tampilan saat proses <i>scraping</i> telah selesai pada aplikasi <i>Octoparse</i>	29
Gambar 3. 6 Tampilan contoh data hasil <i>scraping</i>	29
Gambar 3. 7 Tampilan data setelah pelabelan.....	30
Gambar 3. 8 Proses Prapengolahan Data.....	31
Gambar 3. 9 Proses <i>Text Cleansing</i>	31
Gambar 3. 10 Proses <i>Stemming</i>	31
Gambar 3. 11 Proses <i>Filtering</i>	32
Gambar 3. 12 Proses <i>Tokenization</i>	32
Gambar 3. 13 Proses pembobotan dengan fitur Unigram dan TF-IDF.....	33
Gambar 3. 14 Perancangan antarmuka aplikasi.....	35
Gambar 3. 15 <i>Use Case Diagram</i> Aplikasi Analisis Sentimen.....	36
Gambar 4. 1 Hasil <i>Confusion Matrix</i> pengujian model dengan komposisi data 50:50	49
Gambar 4. 2 Hasil <i>Confusion Matrix</i> pengujian model dengan komposisi data 60:40	50

Gambar 4. 3 Hasil <i>Confusion Matrix</i> pengujian model dengan komposisi data 70:30	51
Gambar 4. 4 Hasil <i>Confusion Matrix</i> dengan komposisi label komentar berbeda 70:30.....	53
Gambar 4. 5 Hasil <i>Confusion Matrix</i> dengan komposisi label komentar berbeda 50:50.....	54
Gambar 4. 6 Hasil <i>Confusion Matrix</i> dengan komposisi label komentar berbeda 20:80.....	55
Gambar 4. 7 <i>Confusion Matrix</i> dari model	56
Gambar 4. 8 Hasil penerapan klasifikasi terhadap <i>review</i> aplikasi Mister Aladin	57
Gambar 4. 9 Hasil penerapan klasifikasi terhadap <i>review</i> aplikasi Pegipegi.....	57
Gambar 4. 10 Hasil penerapan klasifikasi terhadap <i>review</i> aplikasi Tiket.com ...	58
Gambar 4. 11 Hasil penerapan klasifikasi terhadap <i>review</i> aplikasi Tiket.Com ..	59
Gambar 4. 12 Diagram hasil klasifikasi <i>review</i> Mister Aladin	62
Gambar 4. 13 Diagram hasil klasifikasi <i>review</i> Pegipegi	62
Gambar 4. 14 Diagram hasil klasifikasi <i>review</i> Tiket.com.....	63
Gambar 4. 15 Diagram hasil klasifikasi <i>review</i> Traveloka	63
Gambar 4. 16 Tampilan awal aplikasi.....	65
Gambar 4. 17 Tampilan klasifikasi komentar	65
Gambar 4. 18 Tampilan klasifikasi file komentar.....	66
Gambar 4. 19 Tampilan hasil klasifikasi <i>review</i>	67

DAFTAR TABEL

Tabel 2. 1 N-gram kata.....	16
Tabel 2. 2 N-gram kalimat	17
Tabel 4. 1 Daftar aplikasi dan URL yang digunakan.....	37
Tabel 4. 2 Variabel Hasil Keluaran dari <i>Scraping</i>	37
Tabel 4. 3 Tampilan sample data hasil <i>Scraping</i>	38
Tabel 4. 4 Contoh Proses <i>Case Folding</i>	39
Tabel 4. 5 Contoh Proses <i>Remove Noise</i>	40
Tabel 4. 6 Contoh Proses <i>Stemming</i>	41
Tabel 4. 7 Contoh Proses <i>Filtering</i>	42
Tabel 4. 8 Contoh Proses <i>Filtering</i>	43
Tabel 4. 9 Tabel <i>Term Frequency (TF)</i>	44
Tabel 4. 10 Tabel <i>Inverse Document Frequency (IDF)</i>	44
Tabel 4. 11 Tabel <i>Term Frequency-Inverse Document (TF-IDF)</i>	44
Tabel 4. 12 Import library	45
Tabel 4. 13 Melakukan TF-IDF	45
Tabel 4. 14 Melakukan pencarian komposisi data latih dan uji melalui <i>k-fold validation</i>	46
Tabel 4. 15 Penerapan model menggunakan library <i>GridSearchCV</i>	46
Tabel 4. 16 Hasil Pengujian <i>K-Fold</i>	47
Tabel 4. 17 Hasil Pengujian model melalui jumlah komposisi data <i>test</i> dan data <i>train</i>	48
Tabel 4. 18 Hasil pengujian model melalui komposisi pelabelan komentar berbeda	52
Tabel 4. 19 Hasil Pengujian <i>K-Fold</i>	60
Tabel 4. 20 Hasil Pengujian model melalui komposisi data <i>test</i> dan data <i>train</i> ...	60
Tabel 4. 21 Hasil pengujian model melalui komposisi pelabelan komentar berbeda	61

ABSTRAK

Penting bagi penyedia produk atau jasa untuk mengetahui tanggapan konsumen mengenai produk atau jasa yang mereka tawarkan. Tidak bisa dipungkiri bahwa tanggapan konsumen yang muncul dapat memengaruhi citra dari penyedia produk atau jasa. Akan tetapi, memantau dan mengklasifikasikan tanggapan dari konsumen bukanlah hal yang mudah. Pada perusahaan besar seperti Traveloka, Tiket.com, Pegipegi dan Mr.Aladin tanggapan yang dimuat jumlahnya terlalu banyak untuk diproses apabila dilakukan secara manual. Oleh karena itu, diperlukan sebuah metode atau teknik khusus yang mampu menganalisis serta mengelompokkan tanggapan-tanggapan konsumen secara otomatis, apakah termasuk sentimen positif atau negatif. Dengan memanfaatkan machine learning dapat dilakukan klasifikasi secara massif bagaimana tanggapan atau opini pengguna aplikasi terhadap layanan atau produk tersebut melalui penerapan Algoritma klasifikasi Support Vector Machine.

Dalam penelitian ini dilakukan pengambilan data review pada PlayStore menggunakan teknik scraping yang kemudian data yang didapat akan dilakukan prapengolahan data. Setelah dilakukan prapengolahan data, data akan dikelompokkan menjadi data uji dan data latih dimana data latih akan digunakan untuk proses pelatihan model yang dibuat menggunakan algoritma klasifikasi Support Vector Machine (SVM) melalui beberapa tahap pengujian model sehingga didapatkan hasil paling optimal pada model.

Hasil penelitian berdasarkan dari tahapan pengujian model yang telah dilakukan mendapatkan accuracy 91,30%, Precision 91,64%, dan Recall 90,99%. Hasil akurasi tersebut dipengaruhi oleh beberapa hal, yaitu pertama komposisi jumlah data train dan data test melalui k-fold validation, jumlah dataset yang digunakan dan komposisi pelabelan komentar positif dan negatif

Kata Kunci : *Klasifikasi, Analisis Sentimen, scraping, prapengolahan data, Support Vector Machine (SVM), Traveloka, Tiket.com, Pegipegi, Mr.Aladin*

ABSTRACT

It is important for product or service providers to find out what consumers think about the products or services they offer. It cannot be denied that consumer responses that emerge can affect the image of the product or service provider. However, monitoring and classifying responses from consumers is not easy. In large companies such as Traveloka, Tiket.com, Pegipegi and Mr. Aladin, the responses that are loaded are too large to be processed if done manually. Therefore, we need a special method or technique that is able to analyze and classify consumer responses automatically, including positive or negative sentiments. By utilizing machine learning, it can be done massively classification of how the responses or opinions of application users to these services or products through the application of the Support Vector Machine classification algorithm.

In this study, data collection reviews were carried out on the PlayStore using scraping techniques, then the data obtained would be pre-processed. After pre-processing the data, the data will be grouped into test data and training data where the training data will be used for the model training process created using the Support Vector Machine (SVM) classification algorithm through several stages of model testing so that the most optimal results are obtained in the model.

The results of the research are based on the testing stages of the model that has been carried out to produce 91.30% accuracy, 91.64% precision, and 90.99% recall. The results of this accuracy are influenced by several things, namely the composition of the number of train data and test data through k-fold validation, the number of datasets used and the composition of the labeling of positive and negative comments.

Keyword : *Classification, Sentymnt Analyst, Scraping, pre-processing, Support Vector Machine (SVM), Traveloka, Tiket.com, Pegipegi, Mr.Aladin*