



**ANALISIS SENTIMEN MEDIA SOSIAL *TWITTER* MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE* (SVM)  
(STUDI KASUS: MASKAPAI PENERBANGAN PT GARUDA INDONESIA  
(PERSERO) TBK )**

**SKRIPSI**

**TANTRI AYU PRASETIARINI**

**1610511018**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2020**



**ANALISIS SENTIMEN MEDIA SOSIAL *TWITTER* MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE* (SVM)**

**(STUDI KASUS: MASKAPAI PENERBANGAN PT GARUDA INDONESIA  
(PERSERO) TBK)**

**SKRIPSI**

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

**TANTRI AYU PRASETIARINI**

**1610511018**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2020**

## PERNYATAAN ORISINALITAS

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

Nama : Tantri Ayu Prasetiari

NIM :1610511018

Tanggal : 27 Mei 2020

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

Jakarta, 27 Mei 2020

Yang Menyatakan,



(Tantri Ayu Prasetiari)

## **PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

---

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

Nama : Tantri Ayu Prasetiari

NIM :1610511018

Fakultas : Ilmu Komputer

Program Studi : Informatika

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

**Analisis Sentimen Media Sosial Twitter Menggunakan Metode Support  
Vector Machine (SVM)**

**(Studi Kasus: Maskapai Penerbangan PT Garuda Indonesia (Persero)  
Tbk)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/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 sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 27 Mei 2020

Yang Menyatakan,



(Tantri Ayu Prasetiari)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Tantri Ayu Prasetiari  
NIM : 1610511018  
Program Studi : S1 Informatika  
Judul Skripsi : ANALISIS SENTIMEN MEDIA SOSIAL *TWITTER*  
MENGUNAKAN METODE *SUPPORT VECTOR*  
*MACHINE* (SVM)

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Henki Bayu Seta, S.Kom., M.TI.  
Penguji I



I Wayan Widi P., S.Kom., M.TI.  
Penguji II



Iin Ernawati, S.Kom., M.Si.  
Pembimbing I



Nurul Chamidah, S.Kom, M.Kom.  
Pembimbing II



Dr. Ermatita, M.Kom.  
Dekan



Anita Muliawati, S.Kom., M.TI.  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 23 Juni 2020



**ANALISIS SENTIMEN MEDIA SOSIAL *TWITTER* MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE* (SVM)**

**(STUDI KASUS: MASKAPAI PENERBANGAN PT GARUDA INDONESIA  
(PERSERO) TBK)**

Tantri Ayu Prasetiari

**Abstrak**

Media sosial *Twitter* merupakan wadah untuk bertukar informasi. Informasi tersebut dapat berupa opini yang ditujukan oleh suatu perusahaan tertentu. Opini tersebut dapat dianalisis menjadi analisis sentimen yang nantinya akan membantu dalam memahami sentimen opini masyarakat berbentuk teks yang awalnya tidak terstruktur menjadi terstruktur. Data opini tersebut dilabelkan dan diklasifikasikan menjadi kelas positif, dan kelas negatif, dengan menggunakan metode *support vector machine* (SVM), dimana data yang digunakan berasal dari sosial media *twitter* @indonesiagaruda. Dari hasil metode klasifikasi menggunakan *Support Vector Machine* (SVM) dilakukan evaluasi terhadap model dengan menggunakan *confussion matrix* terhadap perhitungan, sehingga mendapatkan nilai akurasi sebesar 88,75% dengan menggunakan kernel linear, sedangkan pada kernel *polynomial* mendapatkan nilai akurasi sebesar 75,625%. Model dengan akurasi tertinggi digunakan untuk prediksi data baru dengan menggunakan data pada bulan Februari 2020 sebanyak 564 *tweet*, untuk dianalisis sentimennya dalam bentuk visualisasi berdasarkan kategori tingkat kepuasan terhadap pelayanan berdasarkan fasilitas, ketepatan waktu, penilaian terhadap tarif tiket, penilaian terhadap *customer service*, serta penilaian terhadap GarudaMiles dan *website* Garuda Indonesia. Dari hasil model prediksi tersebut terdapat 291 sentimen positif dan 273 sentimen negatif, dengan penilaian sentimen positif tinggi pada kategori fasilitas, ketepatan waktu, *customer service*, sedangkan penilaian sentimen negatif tinggi pada kategori tarif tiket, GarudaMiles dan *website* Garuda Indonesia pada bulan Februari 2020.

**Kata kunci:** *twitter*, *support vector machine*, analisis sentimen

***SENTIMENT ANALYSIS SOCIAL MEDIA TWITTER USING THE SUPPORT  
VECTOR MACHINE (SVM) METHOD***

***(CASE STUDY: MASKAPAI PENERBANGAN PT GARUDA INDONESIA  
(PERSERO) TBK)***

Tantri Ayu Prasetiari

**Abstract**

Twitter social media is a place to exchange information. The information can be in the form of an opinion addressed by a particular company. These opinions can be analyzed into sentiment analysis which will help in understanding the sentiments of public opinion in the form of texts that were initially unstructured into structured. The opinion data is labeled and classified into positive classes, and negative classes, using the support vector machine (SVM) method, where the data used comes from the social media twitter @indonesiagaruda. From the results of the classification method using the Support Vector Machine (SVM) an evaluation of the model is done using a confusion matrix against calculations, so that the accuracy value of 88.75% is obtained by using a linear kernel, while the polynomial kernel gets an accuracy value of 75.625%. The highest accuracy model is used for predicting new data using 564 tweets in February 2020, to analyze sentiment in the form of visualization based on the level of satisfaction with services based on facilities, timeliness, assessment of ticket rates, assessment of customer service, and rating against GarudaMiles and the Garuda Indonesia website. From the results of the prediction model, there are 291 positive sentiments and 273 negative sentiments, with a high positive sentiment rating in the category of facilities, timeliness, customer service, while high negative sentiment ratings in the ticket fare category, GarudaMiles and Garuda Indonesia website in February 2020.

**Keywords:** *twitter, support vector machine, sentiment analysis*

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala karunia-Nya, sehingga Skripsi ini berhasil diselesaikan. Penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tua penulis, Suprayetno (bapak) dan Nani Supriatiningsih (bunda), serta keluarga yang selalu memberi dorongan dan doa sehingga dapat menyelesaikan skripsi ini.
2. Ibu Iin Ernawati S.Kom., M.Si. dan Ibu Nurul Chamidah, S.Kom, M.Kom., selaku dosen pembimbing yang telah memberikan saran yang bermanfaat.
3. Ibu Dr. Ermatita, M. Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
4. Ibu Anita Muliawati, S. Kom., M. T. I. selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional Veteran Jakarta.
5. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta terima kasih atas ilmu-ilmu yang bermanfaat.
6. Teman-teman Informatika 2016, rekan-rekan KSM Robotika, dan seluruh rekan mahasiswa yang tidak dapat penulis sebutkan namanya satu persatu, terima kasih atas saran dan dukungan yang telah diberikan.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 27 Mei 2020

Penulis



# DAFTAR ISI

HALAMAN JUDUL.....	ii
PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
LEMBAR PENGESAHAN .....	v
Abstrak.....	vi
Abstract.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xiii
DAFTAR TABEL.....	xiv
DAFTAR SIMBOL .....	xv
DAFTAR LAMPIRAN .....	xvi
BAB 1 PENDAHULUAN .....	17
1.1 Latar Belakang .....	17
1.2 Rumusan Masalah .....	18
1.3 Batasan Masalah.....	18
1.4 Tujuan Penelitian.....	19
1.5 Manfaat Penelitian.....	19
1.6 Luaran yang Diharapkan .....	20
1.7 Sistematika Penulisan.....	20
BAB 2 TINJAUAN PUSTAKA.....	22
2.1 <i>Twitter</i> .....	22
2.2 Analisis Sentimen.....	22
2.3 <i>Data Mining</i> .....	23
2.3.1 Pengertian <i>Text Mining</i> .....	23

2.4	<i>Text Analysis</i> .....	24
2.5	<i>Twitter Crawling</i> .....	24
2.6	<i>Python</i> .....	25
2.7	<i>Tableau</i> .....	25
2.8	<i>Text Preprocessing</i> .....	25
2.9	<i>Term Frequency – Inverse Document Frequency (TF IDF)</i> .....	27
2.10	<i>Support Vector Machine (SVM)</i> .....	28
2.10.1	Konsep <i>Support Vector Machine (SVM)</i> .....	28
2.11	Evaluasi.....	31
2.12	Penelitian Terkait .....	32
BAB 3	METODOLOGI PENELITIAN .....	34
3.1	Tahapan Penelitian .....	34
3.1.1	Identifikasi Masalah .....	35
3.1.2	Studi Pustaka.....	35
3.1.3	Pengumpulan Data .....	35
3.1.4	<i>Text Preprocessing</i> .....	35
3.1.5	Pembobotan <i>Term</i> .....	36
3.1.6	Proses .....	36
3.1.7	Evaluasi .....	36
3.1.8	Prediksi.....	36
3.1.9	Visualisasi .....	37
3.2	Perangkat Penelitian .....	37
	:Intel HD Graphics 620 + NVIDIA .....	37
	GEFORCE 930MX .....	37
3.3	Jadwal Penelitian.....	37
BAB 4	HASIL DAN PEMBAHASAN .....	39

4.1	Data.....	39
4.1.2	Pelabelan Data.....	40
4.2	<i>Text Preprocessing</i> .....	41
4.2.2	<i>Cleaning</i> .....	42
4.2.3	<i>Case Folding</i> .....	43
4.2.4	<i>Tokenizing</i> .....	44
4.2.5	<i>Filtering</i> .....	45
4.2.6	<i>Stemming</i> .....	46
4.2.7	Pembobotan Kata (TF- IDF).....	48
4.3	<i>Data Mining</i> .....	51
4.4	Prediksi Data Baru.....	54
4.4.1	Kategori Tingkat Kepuasan Berdasarkan Fasilitas.....	55
4.4.2	Kategori Tingkat Ketepatan Waktu.....	55
4.4.3	Kategori Penilaian Berdasarkan Tarif Tiket.....	55
4.4.4	Kategori Penilaian Terhadap Customer Service.....	56
4.4.5	Kategori Penilaian Terhadap GarudaMiles dan Website Garuda Indonesia ..	56
4.5	Visualisasi.....	56
4.5.1	Tingkat penilaian sentimen positif dan negatif pada bulan Februari 2020.....	57
4.5.2	Tingkat Penilaian Sentimen Berdasarkan Tanggal.....	57
4.5.3	Tingkat Kepuasan Terhadap Pelayanan.....	59
BAB 5 PENUTUP.....		66
5.1	Kesimpulan.....	66
5.2	Saran.....	67
DAFTAR PUSTAKA.....		68
RIWAYAT HIDUP.....		70
LAMPIRAN.....		72

Lampiran 1 <i>Similarity Index</i> Tugas Akhir .....	73
Lampiran 2 Hasil <i>Filtering Data</i> .....	73

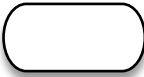


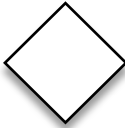
## DAFTAR GAMBAR

Gambar 1 Tahapan <i>Text Preprocessing</i> .....	26
Gambar 2 Konsep <i>Support Vector Machine (SVM)</i> .....	28
Gambar 3 Bagan Tahapan Penelitian.....	34
Gambar 4 Hasil <i>Crawling</i> dengan Format.csv .....	39
Gambar 5 Tingkat Penilaian Sentimen Positif dan Negatif Februari 2020.....	57
Gambar 6 Tingkat Penilaian Sentimen Berdasarkan Tanggal .....	58
Gambar 7 Tingkat Kepuasan Berdasarkan Fasilitas .....	60
Gambar 8 Tingkat Ketepatan Waktu.....	61
Gambar 9 Penilaian Terhadap Tarif Tiket .....	62
Gambar 10 Penilaian Terhadap <i>Customer Service</i> .....	64
Gambar 11 Penilaian Terhadap GarudaMiles dan <i>Website</i> .....	65

## DAFTAR TABEL

Tabel 1	Contoh Tabel <i>Confusion Matrix</i> .....	31
Tabel 2	Jadwal Penelitian.....	38
Tabel 3	Format Data <i>Tweet</i> yang Digunakan .....	40
Tabel 4	Contoh Pelabelan Data <i>Tweet</i> .....	41
Tabel 5	Data Sebelum Proses <i>Cleaning</i> .....	42
Tabel 6	Data Hasil Proses <i>Cleaning</i> .....	42
Tabel 7	Data Sebelum <i>Case Folding</i> .....	43
Tabel 8	Data Hasil Proses <i>Case Folding</i> .....	43
Tabel 9	Data Sebelum <i>Tokenizing</i> .....	44
Tabel 10	Data Hasil Proses <i>Tokenizing</i> .....	44
Tabel 11	Data Sebelum Proses <i>Filtering</i> .....	45
Tabel 12	Data Hasil Proses <i>Filtering</i> .....	46
Tabel 13	Data Sebelum Proses <i>Stemming</i> .....	46
Tabel 14	Data Hasil Proses <i>Stemming</i> .....	47
Tabel 15	Dokumen ( <i>tweet</i> ).....	48
Tabel 16	Perhitungan <i>TF IDF</i> .....	49
Tabel 17	Pembagian Data yang Digunakan.....	51
Tabel 18	Hasil Evaluasi Menggunakan Kernel Linear.....	52
Tabel 19	Hasil Evaluasi Menggunakan Kernel <i>Polynomial</i> .....	53

## DAFTAR SIMBOL

<b>Simbol <i>Flowchart</i></b>			
<b>NO</b>	<b>Simbol</b>	<b>Nama Simbol</b>	<b>Keterangan</b>
1		<i>Terminator</i> ( <i>start, end</i> )	Simbol yang menggambarkan dimulainya atau berakhirnya suatu kegiatan berakhir.
2		<i>Flow Line</i>	Simbol yang menggambarkan hubungan dari simbol satu ke simbol lainnya dalam suatu proses kegiatan.
3		<i>Process</i>	Simbol yang menggambarkan suatu pengolahan yang dijalankan.
4		<i>Decision</i>	Simbol yang menggambarkan pilihan dari suatu proses.

## DAFTAR LAMPIRAN

Lampiran 1	Similarity Index Tugas Akhir .....	73
Lampiran 2	Hasil Filtering Data .....	79