



**ANALISIS SENTIMEN PADA ULASAN PRODUK MASKER DI
MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR
MACHINE DAN SELEKSI FITUR CHI SQUARE**

SKRIPSI

Willy Winata

1810511095

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022



**ANALISIS SENTIMEN PADA ULASAN PRODUK MASKER DI
MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR
MACHINE DAN SELEKSI FITUR CHI SQUARE**

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

Komputer

Willy Winata

1810511095

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

PERNYATAAN ORISINALITAS

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

Nama : Willy Winata

NIM : 1810511095

Tanggal : 3 Juni 2022

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

Jakarta, 3 Juni 2022

Yang Menyatakan,

A yellow rectangular meter stamp with a Garuda emblem at the top. The text "1000" is printed vertically on the left. Below the emblem, it says "METER TEMPEL". A signature is written across the stamp. At the bottom, the serial number "SD137AJX866180741" is printed.

(Willy Winata)

PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama : Willy Winata
NIM : 1810511095
Fakultas : Ilmu Komputer
Program Studi : S1 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:

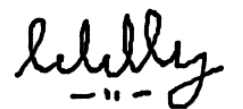
**ANALISIS SENTIMEN PADA ULASAN PRODUK MASKER DI
MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR
MACHINE DAN SELEKSI FITUR CHI SQUARE**

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 kata (Basis data), 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 : 3 Juni 2022

Yang Menyatakan,



(Willy Winata)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Willy Winata
NIM : 1810511095
Program Studi : Informatika
Judul : Analisis Sentimen Pada Ulasan Produk Masker di
*Marketplace Shopee Menggunakan Support Vector
Machine Dan Seleksi Fitur Chi Square*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian dari 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., MTI
Penguji 1



Mavanda Mega Santoni, S.Komp, M.Kom.
Penguji 2



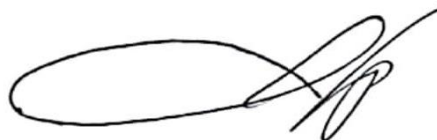
Ati Zaidiah, S.Kom., MTI.
Pembimbing 1



Nurul Chamidah, S.Kom, M.Kom.
Pembimbing 2



Dr. Ermatita, M. Kom.
Dekan



Desta Sandya Prasvita, S. Komp., M.Kom.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Pengesahan : 13 Juli 2022

LEMBAR PERSETUJUAN

Dengan ini dinyatakan bahwa Skripsi berikut :

Nama : Willy Winata
NIM : 1810511095
Program Studi : Informatika
Judul : Analisis Sentimen Pada Ulasan Produk Masker di
Marketplace Shopee Menggunakan *Support Vector*
Machine Dan Seleksi Fitur *Chi Square*

Sebagai bagian persyaratan yang diperlukan untuk mengikuti ujian Sidang Proposal Tugas Akhir/Skripsi pada Program Studi S1 Informatika Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Mengetahui,



Desta Sandya Prasvita, S.Komp., M.Kom.

Ketua Program Studi

Menyetujui,



Ati Zaidiah, S.Kom., MTI.

Pembimbing 1



Nurul Chamidah, S.Kom, M.Kom.

Pembimbing 2

Ditetapkan di : Jakarta

Tanggal Persetujuan : 23 Juni 2022

**ANALISIS SENTIMEN PADA ULASAN PRODUK MASKER DI
MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR
MACHINE DAN SELEKSI FITUR *CHI SQUARE***

Willy Winata

ABSTRAK

Masker menjadi hal wajib saat ingin keluar rumah yang membuat penjualan masker meningkat drastis. Namun dengan adanya pembatasan kegiatan tersebut membuat aktivitas jual beli secara konvensional menjadi terhambat. Oleh karena itu banyak masyarakat beralih dari berbelanja langsung menjadi berbelanja *online*. *Marketplace* Shopee merupakan salah satu *platform* belanja *online*. Penjualan masker di *marketplace* Shopee mengalami peningkatan yang sangat drastis yang membuat variasi penilaian ulasan produk yang berbeda-beda dari pembeli. Pada penelitian dilakukan analisis sentimen terhadap 2 kelas positif dan negatif berdasarkan *rating* ulasan. Algoritma klasifikasi yang digunakan adalah *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) serta menerapkan seleksi fitur *Chi Square*. Tingkat performa yang didapat menggunakan seleksi fitur *Chi Square* dengan nilai taraf nyata (α) 0.3 menghasilkan nilai akurasi sebesar 91,32%, presisi sebesar 93,28%, *recall* sebesar 93,14% dan *f1-score* 93,21%. Dibanding dengan tingkat performa tanpa menggunakan seleksi fitur hanya menghasilkan nilai akurasi sebesar 90,75%, presisi sebesar 92,08%, *recall* sebesar 93,59% dan *f1-score* 92,83%.

Kata kunci: Masker, Analisis Sentimen, *Marketplace*, Shopee, Ulasan Produk, *Support Vector Machine*, *Chi Square*

SENTIMENT ANALYSIS OF MASK PRODUCT REVIEWS AT THE SHOPEE MARKETPLACE USING SUPPORT VECTOR MACHINE AND CHI SQUARE FEATURE SELECTION

Willy Winata

ABSTRACT

Masks are mandatory when you want to walk away from home, which makes mask sales increase drastically. However, with the limitation of these activities, conventional selling activities are hampered. Therefore, many people switch from direct shopping to online shopping. Shopee Marketplace is an online shopping platform. The sale of masks in the Shopee marketplace has experienced a very drastic increase which has made variations in the assessment of different product reviews from buyers. In this study, sentiment analysis is divided into 2 positive and negative classes based on the rating review. The classification algorithm used is a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel and applies the Chi Square feature selection. The result of performance level using the Chi Square feature selection with a significance level value (α) of 0.3 obtained an accuracy of 91,32%, precision of 93,28%, recall of 93,14% and f1-score of 93,21%. Compared to the result of performance level without using feature selection, it only obtained an accuracy of 90.75%, precision of 92.08%, recall of 93.59% and f1-score of 92.83%.

Keywords: Mask, Sentiment Analysis, Marketplace, Shopee, Product Review, Support Vector Machine, Chi Square

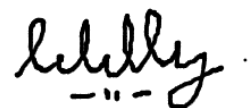
KATA PENGANTAR

Alhamdulillah, segala puji dan syukur untuk Allah SWT yang telah memberikan rahmat dan karunia-Nya yang berlimpah, sehingga skripsi dengan judul “Analisis Sentimen Pada Ulasan Produk Masker Di *Marketplace* Shopee Menggunakan *Support Vector Machine* Dan Seleksi Fitur *Chi Square*” berhasil diselesaikan. Penulis menyadari bahwa apa yang telah diperoleh hingga saat ini tidak lepas dari adanya do’a, bimbingan, motivasi serta bantuan dari berbagai pihak, untuk itu perlunya penulis ucapkan terimakasih kepada:

1. Ibu Ati Zaidiah, S.Kom., MTI. selaku dosen pembimbing 1 dan Nurul Chamidah, S.Kom, M.Kom. selaku dosen pembimbing 2 yang telah membimbing, memberi masukan dan nasihat dalam penulisan proposal hingga skripsi.
2. Ibu Dr. Ermatita, M. Kom. selaku dekan Fakultas Ilmu Komputer.
3. Bapak Desta Sandya Prasvita.S.Kom., M.Kom. selaku ketua program studi Fakultas Ilmu Komputer.
4. Seluruh Bapak/Ibu Dosen Pengajar Fakultas Ilmu Komputer atas ilmu-ilmu bermanfaat yang diberikan.
5. Kedua orang tua tercinta dan kakak yang selalu memberikan penulis dukungan baik materi maupun non materi dalam penulisan skripsi.
6. Teman-teman, saudara dan pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu, terima kasih atas saran dan dukungan yang telah diberikan.

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

Jakarta, 3 Juni 2022



Penulis

DAFTAR ISI

HALAMAN JUDUL.....	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
LEMBAR PENGESAHAN	v
LEMBAR PERSETUJUAN.....	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR SIMBOL.....	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	2
1.3. Ruang Lingkup	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.6. Luaran Yang Diharapkan	4
1.7. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. <i>Marketplace</i> Shopee	6
2.2. Ulasan Produk	6

2.3.	<i>Analisis Sentimen</i>	6
2.4.	<i>Text Mining</i>	7
2.5.	<i>Web Scraping</i>	7
2.6.	Pembobotan TF-IDF.....	8
2.7.	<i>Chi Square</i>	8
2.8.	<i>Support Vector Machine</i>	10
2.8.1.	<i>Kernel</i>	11
2.9.	<i>Confusion Matrix</i>	12
2.10.	Penelitian Terkait.....	13
BAB III METODOLOGI PENELITIAN.....		15
3.1.	Kerangka Pikir.....	15
3.2.	Tahapan Penelitian	15
3.2.1.	Identifikasi Masalah.....	15
3.2.2.	Studi Literatur	16
3.2.3.	Pengumpulan Data	16
3.2.4.	<i>Labeling</i>	17
3.2.5.	<i>Preprocessing</i>	18
3.2.6.	Pembobotan TF-IDF	19
3.2.7.	Seleksi Fitur <i>Chi Square</i>	19
3.2.8.	Klasifikasi <i>Support Vector Machine</i>	20
3.2.9.	Evaluasi.....	20
3.2.10.	Visualisasi	20
3.3.	Alat Bantu Penelitian.....	21
3.3.1.	Perangkat keras	21
3.3.2.	Perangkat Lunak.....	21
3.4.	Jadwal Rencana Penelitian	21

BAB IV HASIL DAN PEMBAHASAN	22
4.1. Pengumpulan Data.....	22
4.2. <i>Labeling</i>	23
4.3. <i>Preprocessing</i>	25
4.3.1. <i>Case folding</i>	25
4.3.2. <i>Cleansing</i>	26
4.3.3. <i>Stopword removal</i>	27
4.3.4. <i>Stemming</i>	28
4.3.5. <i>Normalization</i>	29
4.3.6. <i>Tokenization</i>	30
4.4. Pembobotan TF-IDF.....	32
4.5. Seleksi Fitur <i>Chi Square</i>	36
4.6. Klasifikasi.....	39
4.6.1. Tanpa Seleksi Fitur	41
4.6.2. Seleksi Fitur dengan α 0.5.....	42
4.6.3. Seleksi Fitur dengan α 0.3.....	43
4.6.4. Seleksi Fitur dengan α 0.25.....	44
4.6.5. Seleksi Fitur dengan α 0.1	45
4.7. Evaluasi	46
4.7.1. Perbandingan Akurasi	46
4.7.2. Perbandingan Presisi	47
4.7.3. Perbandingan <i>Recall</i>	47
4.7.4. Perbandingan <i>F1-Score</i>	48
4.7.5. Analisis Hasil	49
4.8. Visualisasi.....	50
4.8.1. Visualisasi Sentimen Positif.....	50

4.8.2. Visualisasi Sentimen Negatif	51
BAB V PENUTUP.....	53
5.1. Kesimpulan.....	53
5.2. Saran.....	54
DAFTAR PUSTAKA	55
RIWAYAT HIDUP	57
LAMPIRAN.....	58

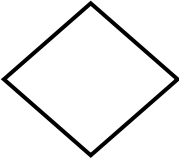
DAFTAR GAMBAR

Gambar 2. 1 <i>Hyperplane Margin Maximal</i>	10
Gambar 3. 1 Alur Penelitian.....	15
Gambar 3. 2 Alur Proses <i>Web Scraping</i>	17
Gambar 3. 3 Alur <i>Preprocessing</i>	18
Gambar 4. 1 Data Hasil <i>Web Scraping</i>	23
Gambar 4. 2 Persentase Kelas Negatif dan Positif.....	24
Gambar 4. 3 Tokenisasi Sebelum <i>Stopword Removal</i>	31
Gambar 4. 4 Tokenisasi Sesudah <i>Normalization</i>	32
Gambar 4. 5 <i>Confusion Matrix</i> Tanpa Seleksi Fitur	41
Gambar 4. 6 <i>Confusion Matrix</i> dengan Taraf Nyata 0.5.....	42
Gambar 4. 7 <i>Confusion Matrix</i> dengan Taraf Nyata 0.3.....	43
Gambar 4. 8 <i>Confusion Matrix</i> dengan Taraf Nyata 0.25.....	44
Gambar 4. 9 <i>Confusion Matrix</i> dengan Taraf Nyata 0.1.....	45
Gambar 4. 10 Grafik Perbandingan Akurasi.....	46
Gambar 4. 11 Grafik Perbandingan Presisi.....	47
Gambar 4. 12 Grafik Perbandingan <i>Recall</i>	48
Gambar 4. 13 Grafik Perbandingan <i>f1-score</i>	48
Gambar 4. 14 <i>Wordcloud</i> Sentimen Positif.....	50
Gambar 4. 15 Frekuensi Kata Sentimen Positif	51
Gambar 4. 16 <i>Wordcloud</i> Sentimen Negatif	51
Gambar 4. 17 Frekuensi Kata Sentimen Negatif	52

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	12
Tabel 4. 1 Contoh Ulasan dengan <i>Rating</i> 3.....	23
Tabel 4. 2 Ulasan dengan Kelas Negatif dan Positif.....	25
Tabel 4. 3 Ulasan Sebelum dan Sesudah <i>Case Folding</i>	26
Tabel 4. 4 Ulasan Sebelum dan Sesudah <i>Cleansing</i>	27
Tabel 4. 5 Ulasan Sebelum dan Sesudah <i>Stopword Removal</i>	28
Tabel 4. 6 Ulasan Sebelum dan Sesudah <i>Stemming</i>	28
Tabel 4. 7 Ulasan Sebelum dan Sesudah <i>Normalization</i>	29
Tabel 4. 8 Ulasan Sebelum dan Sesudah <i>Tokenization</i>	30
Tabel 4. 9 Perhitungan <i>Term Frequency</i>	33
Tabel 4. 10 Perhitungan <i>Inverse Document Frequency</i>	34
Tabel 4. 11 Perhitungan pembobotan TF-IDF	35
Tabel 4. 12 Contoh ulasan sebelum dilakukan seleksi fitur	36
Tabel 4. 13 Contoh ulasan setelah dilakukan pembobotan	37
Tabel 4. 14 Hasil perhitungan <i>p-value</i>	38
Tabel 4. 15 Hasil seleksi fitur	38
Tabel 4. 16 Pembagian data	39
Tabel 4. 17 Hasil Seleksi Fitur.....	40
Tabel 4. 18 Perbandingan hasil evaluasi	46

DAFTAR SIMBOL

<i>Flowchart Diagram</i>			
No.	Gambar	Nama	Keterangan
1.		<i>Terimal (Start, End)</i>	Menggambarkan kegiatan dimulai atau kegiatan berakhir
2.		<i>Flow Direction</i>	Menggambarkan hubungan antar simbol yang menyatakan proses jalannya sistem
3.		<i>Process</i>	Menggambarkan penjelasan dari proses yang akan dijalankan
4.		<i>Decision</i>	Menggambarkan proses yang memiliki kondisi dimana harus memilih lebih dari satu proses

DAFTAR LAMPIRAN

Lampiran 1 Korpus <i>Stopword</i> Tambahan	59
Lampiran 2 Korpus Normalisasi	63
Lampiran 3 Tabel Distribusi <i>Chi Square</i>	72