

**ANALISIS SENTIMEN ULASAN KONSUMEN  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*  
DAN KOMBINASI *SYNTHETIC MINORITY OVERSAMPLING*  
*TECHNIQUE (SMOTE)* DENGAN *TOMEK LINKS***

**Tesis  
untuk memenuhi sebagian persyaratan  
mencapai derajat Sarjana S-2 Program Studi  
Magister Sistem Informasi**



**Dody Indra Sumantiawan  
30000319410009**

**SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2022**

HALAMAN PENGESAHAN

TESIS

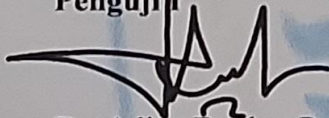
**ANALISIS SENTIMEN ULASAN KONSUMEN  
MENGUNAKAN METODE *SUPPORT VECTOR  
MACHINE* DAN KOMBINASI *SYNTHETIC MINORITY  
OVERSAMPLING TECHNIQUE (SMOTE)* DENGAN  
*TOMEK LINKS***

Oleh:  
**Dody Indra Sumantiawan**  
30000319410009

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 24 Januari 2023 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

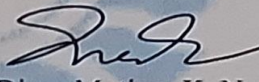
Semarang, 24 Januari 2023  
Mengetahui,

**Penguji I**



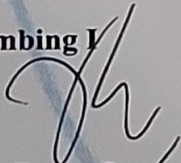
Dr. Adian Fatchur Rochim, S.T., M.T.  
NIP. 197302261998021001

**Penguji II**



Dinar Mutiara K. N., S.T.,  
M.InfoTech.(Comp), Ph.D  
NIP. 197601102009122002

**Pembimbing I**



Jatmiko Endro Suseno, M.Si., Ph.D  
NIP. 197211211998021001

**Pembimbing II**



Dr. Eng. Wahyul Amfen Syafei, ST., MT  
NIP. 197112181995121001

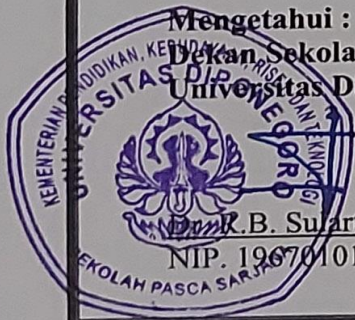
**Mengetahui :**

**Dekan Sekolah Pascasarjana  
Universitas Diponegoro**

R.B. Sularto, S.H., M.Hum  
NIP. 196701011991031005

**Ketua Program Studi  
Magister Sistem Informasi**

Drs. Bayu Surarso, M.Sc., Ph.D  
NIP. 196311051988031001





**HALAMAN PERSETUJUAN**

**Tesis dengan judul :**

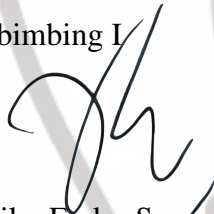
**ANALISIS SENTIMEN ULASAN KONSUMEN  
MENGUNAKAN METODE *SUPPORT VECTOR  
MACHINE* DAN KOMBINASI *SYNTHETIC MINORITY  
OVERSAMPLING TECHNIQUE (SMOTE)* DENGAN  
TOMEK LINKS**

**Oleh:  
Dody Indra Sumantiawan  
30000319410009**

Telah dilakukan pembimbingan tesis dan dinyatakan layak untuk mengikuti ujian tesis pada Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 10 November 2022  
Menyetujui,

Pembimbing I



Jatmiko Endro Suseno, M.Si., Ph.D  
NIP. 197211211998021001

Pembimbing II



Dr. Eng. Wahyul Amien Syafei, ST., MT  
NIP. 197112181995121001

**Sekolah Pascasarjana**

**PERNYATAAN PERSETUJUAN**  
**PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

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

Nama : Dody Indra Sumantiawan  
NIM : 30000319410009  
Program Studi : Magister Sistem Informasi  
Program : Sekolah Pascasarjana  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul:

**ANALISIS SENTIMEN ULASAN KONSUMEN MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE* DAN KOMBINASI *SYNTHETIC  
MINORITY OVERSAMPLING TECHNIQUE (SMOTE)* DENGAN *TOMEK  
LINKS***

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang  
Pada tanggal : 3 Januari 2023  
Yang menyatakan



Dody Indra Sumantiawan  
NIM. 30000319410009

Sekolah Pascasarjana

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 3 Januari 2023



Dody Indra Sumantiawan

# Sekolah Pascasarjana

## KATA PENGANTAR

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penyayang, penulis memanjatkan puji syukur atas limpahan rahmat, hidayah dan karunia-Nya sehingga dapat menyelesaikan penelitian Tesis dengan judul “Analisis Sentimen Ulasan Konsumen Menggunakan Metode *Support Vector Machine* dan kombinasi *Synthetic Minority Oversampling Technique* (SMOTE) dengan *Tomek Links*” Tesis ini disusun untuk memenuhi salah satu syarat menyelesaikan studi dan memperoleh gelar Magister Komputer (M.Kom) pada Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro Semarang.

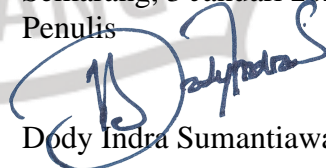
Penyusunan Tesis ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Penulis ingin mengucapkan terima kasih kepada:

1. Bapak Dr. R.B. Sularto, S.H., M.Hum. Selaku Dekan Sekolah Pascasarjana Universitas Diponegoro Semarang.
2. Bapak Drs. Bayu Surarso, M.Sc., Ph.D selaku Ketua Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.
3. Ibu Dr. Oky Dwi Nurhayati, S.T., M.T. Selaku Sekretaris Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang.
4. Bapak Jatmiko Endro Suseno, M.Si., Ph.D Selaku dosen pembimbing 1 yang senantiasa meluangkan waktu untuk memberikan bimbingan, arahan, semangat dan motivasi kepada penulis selama penyusunan Tesis.
5. Bapak Dr.Eng. Wahyul Amien Syafei, ST, MT Selaku dosen pembimbing 2 yang senantiasa meluangkan waktu untuk memberikan bimbingan, arahan, semangat dan motivasi kepada penulis selama penyusunan Tesis.

Penulis menyadari masih terdapat banyak kekurangan dari penyusunan Tesis ini sehingga penulis mengharapkan kritik dan saran yang bersifat membangun demi kesempurnaan penelitian di masa mendatang. Semoga tesis ini dapat bermanfaat untuk menambah pengetahuan ilmu. Akhir kata, penulis mengucapkan permohonan maaf apabila terdapat kata-kata yang kurang berkenan

Semarang, 3 Januari 2023

Penulis



Dody Indra Sumantiawan

# Sekolah Pascasarjana

## DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
HALAMAN PERSETUJUAN.....	iii
PERNYATAAN PERSETUJUAN .....	iv
PERNYATAAN.....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR .....	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR ARTI LAMBANG DAN SINGKATAN .....	xii
ABSTRAK .....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Tujuan Penelitian .....	5
1.3 Manfaat Penelitian .....	5
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Dasar Teori.....	14
2.2.1 Pasar <i>Online</i> .....	14
2.2.2 <i>Web Scraping</i> .....	15
2.2.3 Analisis Sentimen.....	16
2.2.4 <i>Text Mining</i> .....	17
2.2.5 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	19
2.2.6 <i>Word Cloud</i> .....	20
2.2.7 <i>Synthetic Minority Oversampling Technique (SMOTE)</i> .....	20
2.2.8 <i>Tomek Links</i> .....	21
2.2.9 <i>SMOTE -Tomek Links</i> .....	22
2.2.10 <i>Support Vector Machine (SVM)</i> .....	23
2.2.11 Ukuran Keباikan Model.....	26
BAB III METODE PENELITIAN.....	29
3.1 Bahan dan Alat Penelitian.....	29
3.2 Prosedur Penelitian.....	30
3.2.1 Studi literatur.....	31



3.2.2 Pengambilan Data .....	31
3.2.3 Pelabelan Data.....	31
3.2.4 <i>Preprocessing</i> .....	32
3.2.5 TF-IDF .....	37
3.2.6 Kombinasi SMOTE dan <i>Tomek Links</i> .....	38
3.2.7 Pemodelan dengan metode SVM.....	40
3.2.8 Evaluasi.....	41
3.3 Kerangka Sistem Informasi.....	41
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN .....</b>	<b>44</b>
4.1 Hasil Penelitian .....	44
4.1.1 Analisis Pembelajaran.....	44
4.1.2 Proses Analisis Sistem .....	44
4.1.3 <i>Word Cloud</i> .....	48
4.1.4 Sistem Informasi .....	51
4.2 Pembahasan.....	53
4.2.1 Evaluasi Sistem .....	53
4.2.2 Hasil Evaluasi Sistem.....	57
4.2.3 Validasi Hasil.....	59
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>60</b>
5.1 Kesimpulan .....	60
5.2 Saran.....	61
<b>DAFTAR PUSTAKA .....</b>	<b>62</b>
<b>LAMPIRAN .....</b>	<b>67</b>

Sekolah Pascasarjana



## DAFTAR GAMBAR

Gambar 2. 1 Proses SMOTE.....	21
Gambar 2. 2 Proses Tomek Links.....	22
Gambar 2. 3 Konsep SVM.....	24
Gambar 3. 1 Prosedur Penelitian Sentimen Analisis.....	30
Gambar 3. 2 Proses Scraping.....	31
Gambar 3. 3 <i>Library</i> TF-IDF.....	37
Gambar 3. 4 <i>Library</i> Kombinasi SMOTE dan Tomek Links.....	38
Gambar 3. 5 <i>Library</i> SVM.....	40
Gambar 3. 6 Kerangka Sistem Informasi.....	42
Gambar 4. 1 Tampilan awal data yang sudah diberi label.....	45
Gambar 4. 2 Hasil data dari proses cleaning, tokenisasi dan normalisasi.....	45
Gambar 4. 3 Hasil PreProcessing.....	46
Gambar 4. 4 Kelas Distribusi sebelum SMOTE- <i>Tomek Links</i> .....	47
Gambar 4. 5 Sentimen Distribusi dengan SMOTE- <i>Tomek Links</i> .....	48
Gambar 4. 6 <i>Word cloud</i> sentimen positif.....	49
Gambar 4. 7 <i>Word cloud</i> sentimen negatif.....	50
Gambar 4. 8 Halaman Navigasi SI.....	51
Gambar 4. 9 Tampilan Home SI.....	52
Gambar 4. 10 Halaman Train SI.....	52
Gambar 4. 11 Halaman Model SI.....	53
Gambar 4. 12 Hasil Confusion Matrix tanpa SMOTE- Tomek Links.....	54
Gambar 4. 13 Confusion Matrix dengan SMOTE- Tomek Links.....	55
Gambar 4. 14 ROC AUC Tanpa Kombinasi SMOTE dan Tomek Links.....	56
Gambar 4. 15 ROC AUC dengan Kombinasi SMOTE dan Tomek Links.....	57

Sekolah Pascasarjana

## DAFTAR TABEL

Tabel 2. 1 Daftar Penelitian terkait .....	9
Tabel 2. 2 Ilustrasi Confusion Matrix .....	27
Tabel 2. 3 Nilai rentang AUC dan tingkatan klasifikasi menurut Gorunescu (2011) .....	28
Tabel 3. 1 Case Folding .....	33
Tabel 3. 2 Ilustrasi Data Cleaning.....	33
Tabel 3. 3 Ilustrasi Tokenisasi .....	34
Tabel 3. 4 Ilustrasi Normalisasi .....	35
Tabel 3. 5 Ilustrasi Filtering.....	36
Tabel 3. 6 Ilustrasi Stemming.....	36
Tabel 4.1 Komparasi Hasil Confusion Matrix .....	57
Tabel 4.2 Komparasi Skor ROC AUC.....	58



Sekolah Pascasarjana

## DAFTAR LAMPIRAN

Lampiran 1. Surat Keterangan Validasi Bahasa .....	67
Lampiran 2. Daftar Istilah .....	68



# Sekolah Pascasarjana

## DAFTAR ARTI LAMBANG DAN SINGKATAN

### DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
$f_{t,d}$	Frekuensi ( $f$ ) <i>term</i> ( $t$ ) pada dokumen ( $d$ )
$df_t$	Jumlah dokumen yang mengandung <i>term</i> ( <i>document frequency</i> )
$N$	Jumlah keseluruhan dokumen
$W_{t,d}$	Rumus bobot <i>term Weighting</i>
$tf_{t,d}$	Jumlah kemunculan <i>term</i> dalam dokumen
$idf_t$	Jumlah dokumen yang mengandung <i>term</i>
$\vec{w}$	Bidang <i>hyper</i>
$\vec{x}$	Memetakan setiap vektor masukan ke dalam ruang dimensi
$b$	Bias
$y_i$	Korespondensi (penyampaian maksud) kelas
$\vec{x}_i$	vektor input
<i>Recall</i>	Rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif
<i>Precision</i>	Rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif
$\log_{10} \frac{N}{df_t}$	Log dari jumlah keseluruhan dokumen dibagi dengan jumlah kata yang sama pada semua dokumen
$\tau(w)$	<i>Quadratic Programming (QP) problem</i>
0, if $tf_{t,d} = 0$	Jika <i>term</i> tidak terdapat dalam dokumen, bobotnya adalah nol (0)

Sekolah Pascasarjana



## DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
NLP	<i>Natural Language Processing</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
NLTK	<i>Natural Language Tool Kit</i>
HTML	<i>Hypertext Markup Language</i>
CSV	<i>Comma-Separated Values</i>
SVM	<i>Support Vector Machine</i>
SMOTE	<i>Synthetic Minority Oversampling Technique</i>
ROC	<i>Receiver Operating Characteristic</i>
AUC	<i>Area Under the Curve</i>
FPR	<i>False Positive Rate</i>
TPR	<i>True Positive Rate</i>
TF-IDF	<i>True Positive Rate</i>

Sekolah Pascasarjana

# ANALISIS SENTIMEN ULASAN KONSUMEN MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN KOMBINASI *SYNTHETIC MINORITY OVERSAMPLING* *TECHNIQUE (SMOTE)* DENGAN *TOMEK LINKS*

## ABSTRAK

Aktivitas belanja di pasar *online* terutama tren *fashion* dengan segala upaya promo yang ditawarkan terus meningkat. Salah satu pertimbangan untuk membeli produk di pasar *online* adalah dengan membaca ulasan. Setiap ulasan konsumen menunjukkan tingkat ketertarikan terhadap produk. Banyaknya ulasan negatif dan timbulnya banyak ulasan yang bervariasi menjadi masalah dalam mengkategorikan ulasan. Analisis sentimen sebagai cara melihat polaritas ulasan untuk mengklasifikasi ulasan positif dan negatif. Metode *Support Vector Machine* dan kombinasi *Synthetic Minority Oversampling Technique (SMOTE)* dengan *Tomek Links* diterapkan dalam penelitian ini. Klasifikasi menggunakan metode *Support Vector Machine* dan kombinasi *Synthetic Minority Oversampling Technique (SMOTE)* dengan *Tomek Links* menunjukkan hasil yang lebih baik dengan akurasi 0,92, *Precision* 0,89, *Recall* 0,89, dan *f1-score* 0,89 daripada tanpa kombinasi *Synthetic Minority Oversampling Technique (SMOTE)* dengan *Tomek Links* dengan akurasi 0,68, *Precision* 0,55, *Recall* 0,99, dan *f1-score* 0,71 Hasil word cloud sentiment positif menunjukan ulasan ‘bagus’, ‘warna’, ‘barang’, ‘suka’, ‘kirim’, ‘cepat’, ‘langganan’, ‘diskon’, sedangkan ulasan negatif ‘jelek’, ‘celana’, ‘jahit’, ‘ukur’, ‘tipis’, ‘kecewa’, ‘kecil’, ‘pinggang’. Berdasarkan hasil klasifikasi penyeimbangan data dan word cloud dilakukan perbaikan dengan meningkatkan *quality control* produk secara keseluruhan, menyertakan kebijakan keterlambatan/kerusakan barang pada toko atau deskripsi produk dan meningkatkan kinerja admin dalam melayani pelanggan.

Kata kunci: *Analisis Sentimen, Klasifikasi, Support Vector Machine, SMOTE, Tomek Links*

SEMARANG

Sekolah Pascasarjana

# SENTIMENT ANALYSIS CUSTUMER REVIEWS USING *SUPPORT VECTOR MACHINE* METHOD AND COMBINATION SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) WITH *TOMEK LINKS*

## ABSTRACT

Shopping activities in the online market especially fashion trends with all the promo efforts offered continue to increase. One of the considerations for buying products on the online market is to read reviews. Each consumer review shows the level of interest in the product. The number of negative reviews and the emergence of many varied reviews pose a problem in categorizing reviews. Sentiment analysis is a way of looking at the polarity of reviews to classify positive and negative reviews. The *Support Vector Machine* method and the combination of the Synthetic Minority Oversampling Technique (SMOTE) with *Tomek Links* are applied in this study. Classification using the *Support Vector Machine* method and the combination of the Synthetic Minority Oversampling Technique (SMOTE) with *Tomek Links* showed better results with an accuracy of 0.92, precision of 0.89, recall of 0.89, and f1-score of 0.89 than without combination the Synthetic Minority oversampling Technique (SMOTE) with *Tomek Links* with an accuracy of 0.68, precision of 0.55, recall of 0.99, and an f1-score of 0.71. Positive sentiment word cloud results show reviews 'bagus', 'warna', 'barang', 'suka', 'kirim', 'cepat', 'langganan', 'diskon', while the reviews are negative 'jelek', 'celana', 'jahit', 'ukur', 'tipis', 'kecewa', 'kecil', 'pinggang'. Based on the results of the classification of balancing data and word clouds, improvements are made by increasing overall product quality control, including policies for delays/damage to goods in stores or product descriptions and improving admin performance in serving customers.

Keywords: *Sentiment Analysis, Classification, Support Vector Machine, SMOTE, Tomek Links.*

SEMARANG  
Sekolah Pascasarjana