

HALAMAN PERNYATAAN KEASLIAN TULISAN

Hari/tanggal : Rabu, 03 Juli 2024
Pukul : 03.00 WIB
Nama Mahasiswa : Nurun Nashihah Zahro'
NIM : 2120200498
Judul : Implementasi *Synthetic Minority Over-Sampling Technique* (SMOTE) Dalam Analisis Sentimen Program Pemerintah Food Estate Pada *Twitter* Menggunakan Algoritma *Support Vector Machine* (SVM)

Menyatakan dengan sebenarnya bahwa skripsi yang di tulis untuk memenuhi tugas akhir Program Studi S1 Teknik Informatika ini tidak mempunyai persamaan dengan skripsi lain atau bebas dari plagiasi. Demikian pernyataan ini dibuat tanpa paksaan dari pihak manapun. Apabila pernyataan ini tidak benar, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Bojonegoro, 03 Juli 2024



Nurun Nashihah Zahro'
NIM 2120200498

HALAMAN PERSETUJUAN

Yang bertanda tangan di bawah ini Dosen Pembimbing dari:

Nama : Nurun Nashihah Zahro'
NIM : 2120200498
Program Studi/Fakultas : Teknik Informatika/Sains dan Teknologi
Judul : Implementasi *Synthetic Minority Over-Sampling Technique* (SMOTE) Dalam Analisis Sentimen Program Pemerintah Food Estate Pada *Twitter* Menggunakan Algoritma *Support Vector Machine* (SVM)

Menyatakan bahwa Mahasiswa tersebut telah disetujui dan memenuhi syarat untuk diajukan dalam Ujian Skripsi.

Bojonegoro, 04 Juli 2024

Pembimbing I



Hastie Audytra, S.Kom.,M.T.

NIDN. 07080449004

Pembimbing II



Jauhara Rana Budiani, M. Stat.

NIDN. 0705079201

HALAMAN PENGESAHAN

Nama : Nurun Nashihah Zahro'
Nim : 2120200498
Judul : Implementasi *Synthetic Minority Over-Sampling Technique* (SMOTE) Dalam Analisis Sentimen Program Pemerintah Food Estate Pada *Twitter* Menggunakan Algoritma *Support Vector Machine* (SVM)

Telah dipertahankan dalam Ujian Skripsi pada tanggal 13 Juli 2024

Dewan Penguji
Penguji 1



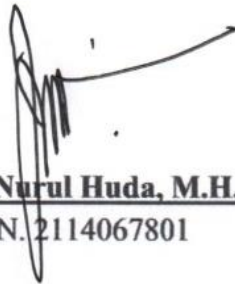
M. Jauhar Vikri, M.Kom.
NIDN. 0712078803

Tim Pembimbing
Pembimbing I



Hastie Audytra, S.Kom., M.T.
NIDN. 0708049004

Dewan Penguji
Penguji 2



Dr. Nurul Huda, M.H.I.
NIDN. 2114067801

Pembimbing II



Jauhara Rana Budiani, M. Stat.
NIDN. 0705079201

Mengetahui,
Dekan Fakultas Sains dan Teknologi



M. Jauhar Vikri, M.Kom.
NIDN. 0712078803

Mengetahui,
Ketua Prodi Teknik Informatika



Mula Agung Barata, S.S.T., M.Kom.
NIDN. 0711049301

HALAMAN MOTTO DAN PERSEMBAHAN

Motto :

"Allah tidak akan membebani seseorang, melainkan sesuai dengan kesanggupannya".

(Q.S Al-Baqarah:286)

Kuncinya, Libatkan Allah dalam setiap persoalan apapun.

"Letakan aku dalam hatinu, maka aku akan meletakanmu dalam hatiku"

(Q.S Al-Baqarah:152)

“Yang Penting Sudah Berusaha dan Berdo’a Perihal Hasil
Biarlah Menjadi Urusan-NYA”

-Dr. Fahrudin Faiz-

“Orang Beriman gak Panikan”

-Ustadzah Halimah Alaydrus-

Persembahan :

Tiada lembar skripsi yang paling indah dalam skripsi ini kecuali lembar persembahan. Dengan mengucapkan syukur Alhamdulillahirobbil’alamin, sungguh perjuangan yang cukup panjang telah penulis lalui, untuk karya yang sederhana ini dipersembahkan untuk:

Allah SWT yang telah memberikan kemudahan dan pertolongan sehingga saya dapat menyelesaikan skripsi saya dengan baik.

Cinta pertama sekaligus menjadi sosok yang menginspirasi penulis yaitu Ayahanda Mustofa (Alm). Beliau memang tidak sempat melihat

putrinya sampai kependidikan bangku perkuliahan. Beliau adalah inspirasi penulis untuk menjadi sosok perempuan yang mandiri. Terimakasih atas do'a dan ridhonya sehingga penulis sudah sampai dititiki ini.

Pintu surgaku dan sekaligus menjadi panutan penulis untuk menjadi sosok perempuan yang kuat, penyayang dan memiliki kesabaran yang tinggi yaitu Ibunda Shoifah. Dan tak lupa kepada kakak-kakakku (Umi Sa'adatul Khoiriyah, Ista Hamidatin, dan Durrotun Nafisah) yang selalu mensupport dalam segala hal. Terimakasih atas kasih sayang, semangat, ridho dan do'a yang selalu dilantarkan disetiap sholatnya demi keberhasilan penulis untuk mewujudkan impian dan harapannya.

Keluarga besar saya terimakasih atas do'a dan dukungan yang selalu diberikan kepada saya.

Abah Yai Kholiq Munir beserta Ibu Nyai Maria Ulfa dan jajaran Dzurriyah Abu Dzarrin yang selalu saya harapkan do'a dan restunya.

Teman seperjuangan angkatan 2020, terimakasih atas segala waktu, usaha, dan dukungan moral maupun material yang telah diberikan.

Siti Novi Mudayani, sahabat yang selalu mendukung setiap keputusan yang penulis ambil, sahabat yang selalu mengulurkan tangan agar penulis bangkit kebalikan dan tidak terus berlarut dalam kesedihan. Terimakasih sudah menemani penulis dalam setiap prosesnya.

Akhir kata semoga skripsi ini dapat menambah wawasan dan dapat bermanfaat bagi kita semua, Aamiin Ya Robbal Alamin.

KATA PENGANTAR

Bismillahirohmanirrohim

Puji syukur kita panjatkan kehadirat Allah SWT, karena atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi dengan judul "Implementasi *Synthetic Minority Over-Sampling Technique* (SMOTE) Dalam Analisis Sentimen Program Pemerintah *Food Estate* Pada *Twitter* Menggunakan Algoritma *Support Vector Machine* (SVM)" diberi kemudahan, kelancaran dan juga sesuai dengan apa yang penulis harapkan.

Dengan terselesaikannya penyusunan Skripsi ini, penulis tidak lupa mengucapkan banyak terima kasi kepada :

1. Bapak M. Jauharul Ma'arif, M.Pd.I. selaku Rektor Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
2. Dekan Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Bapak M. Jauhar Vikri, M.Kom
3. Bapak Mula Agung Barata, S.S.T.,M.Kom. selaku Ketua Program Studi Teknik Informatika Universtias Nahdlatul Ulama Sunan Giri Bojonegoro.
4. Ibu Hastie Audytra, S.Kom., M.T. selaku Dosen Pembimbing 1 yang memberikan banyak arahan, bimbingan, serta motivasi yang luar biasa, dan juga banyak dukungan dalam penyusunan skripsi.
5. Ibu Jauhara Rana Budiani. M. Stat. selaku Dosen Pembimbing 2 yang memberikan bimbingan dan juga dukungan dalam penyusunan skripsi.
6. Dosen penguji yang telah memberikan bimbingan dan arahan.
7. Bapak dan Ibu Dosen Teknik Informatika yang selama ini telah memberikan banyak sekali ilmu pengetahuan, pengalaman, dan juga memberi dukungan.

Penulis menyadari sepenuhnya bahwa penulisan Proposal Skripsi ini masih memiliki banyak kekurangan. Maka dari itu, penulis banyak mengharapkan kritik dan saran yang bersifat membangun dari semua pihak kepada semua pihak.

Bojonegoro, 10 Maret 2024

Penulis

Nurun Nashihah Zahro'



ABSTRACT

Zahro', Nurun Nashihah. Implementation of SMOTE in Sentiment Analysis of the Government Food Estate Program on Twitter using the SVM Algorithm. Thesis, Department of Informatics Engineering, Faculty of Science and Technology, Nahdlatul Ulama' Sunan Giri University. Main Supervisor Hastie Audytra, S.Kom., M.T. and the Assistant Advisor forjauhara Rana Budiani, M.Stat.

Sentiment analysis is a method used to understand public perceptions and opinions on a particular topic. In this study, the researchers implemented the Synthetic Minority Over-sampling Technique (SMOTE) to address the issue of data imbalance in sentiment analysis of the Food Estate program on Twitter. Data imbalance often occurs when the amount of data in one class is significantly larger than in another class, which can lead to bias in the analysis results. The method used in this research is the Support Vector Machine (SVM) algorithm, known for its capability in data classification. The data used consists of tweets related to the Food Estate program, which were collected through web scraping from the Twitter platform. The analysis process includes data collection, data preprocessing, applying the SMOTE technique to balance the data, and finally training the SVM model for sentiment analysis. The study results indicate that the use of SMOTE successfully improved the performance of the SVM model in classifying the sentiment of tweets related to the Food Estate program. The application of SMOTE significantly contributed to addressing the data imbalance issue, as evidenced by the improvement in accuracy, precision, and recall of the SVM model. The study demonstrated that the implemented SVM model achieved an accuracy of 67.4% in classifying public sentiment towards the Food Estate program. This research makes a significant contribution to the field of sentiment analysis, particularly in the context of data imbalance and the application of the SVM algorithm to social media data.

Keywords: Sentiment Analysis, Food Estate, SMOTE, SVM, Twitter, Data Imbalance.

UNUGIRI

ABSTRAK

Zahro', Nurun Nashihah. Implementasi SMOTE dalam Analisis Sentimen Program Pemerintah *Food Estate* pada Twitter menggunakan Algoritma SVM. Skripsi Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Nahdlatul Ulama' Sunan Giri. Pembimbing Utama Hastie Audytra, S.Kom.,M.T. dan Pembimbing Pendamping Jauhara Rana Budiani, M. Stat.

Analisis sentimen merupakan metode yang digunakan untuk memahami persepsi dan opini publik terhadap suatu topik. Dalam penelitian ini, peneliti mengimplementasikan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menangani masalah ketidakseimbangan data dalam analisis sentimen program *Food Estate* di Twitter. Ketidakseimbangan data sering terjadi ketika jumlah data dalam satu kelas jauh lebih banyak dibandingkan dengan kelas lainnya, yang dapat menyebabkan bias dalam hasil analisis. Metode yang digunakan dalam penelitian ini adalah algoritma *Support Vector Machine* (SVM), yang terkenal dengan kemampuannya dalam klasifikasi data. Data yang digunakan adalah *tweet* yang berkaitan dengan program *Food Estate*, yang diambil melalui *web scraping* dari platform Twitter. Proses analisis meliputi pengumpulan data, pra-pemrosesan data, penerapan teknik SMOTE untuk menyeimbangkan data, dan akhirnya pelatihan model SVM untuk analisis sentimen. Hasil penelitian menunjukkan bahwa penggunaan SMOTE berhasil meningkatkan kinerja model SVM dalam mengklasifikasikan sentimen *tweet* terkait program *Food Estate*. Penerapan SMOTE memberikan kontribusi signifikan dalam mengatasi masalah ketidakseimbangan data, yang terbukti dari peningkatan akurasi, *presisi*, dan *recall* pada model SVM. Hasil penelitian menunjukkan bahwa model SVM yang diimplementasikan berhasil mencapai akurasi sebesar 67.4% dalam mengklasifikasikan sentimen masyarakat terhadap program *food estate*. Penelitian ini memberikan kontribusi penting dalam bidang analisis sentimen, khususnya dalam konteks ketidakseimbangan data dan penerapan algoritma SVM untuk data dari media sosial.

Kata Kunci: Analisis Sentimen, *Food Estate*, SMOTE, SVM, *Twitter*, Ketidakseimbangan Data.

UNUGIRI

DAFTAR ISI

HALAMAN SAMPUL DALAM.....	i
HALAMAN PERNYATAAN KEASLIAN TULISAN.....	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN MOTTO DAN PERSEMBAHAN	v
KATA PENGANTAR.....	vii
ABSTRACT	ix
ABSTRAK	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Masalah.....	5
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	8
2.1 Penelitian Terkait.....	8
2.2. <i>Critical Review</i>	19
2.3. Landasan Teori.....	19
2.3.1 Analisis Sentimen	20
2.3.2 Data Mining	20
2.3.2.1 Tahapan Data Mining	22
2.3.3 <i>Text Mining</i>	23
2.3.4 <i>Text Preprocessing</i>	23
2.3.4.1 <i>Cleaning</i>	23
2.3.4.2 <i>Case folding</i>	24
2.3.4.3 <i>Tokenizing</i>	24
2.3.4.4 <i>Remove stopword atau stopword removal</i>	24

2.3.4.5	<i>Stemming</i>	24
2.3.5	Klasifikasi	25
2.3.6	Algoritma <i>Support Vector Machine</i> (SVM).....	25
2.3.7.	<i>Term Frequency-Inverse Document Frequency</i> (TF-IDF).	29
2.3.7.1.	<i>Term Frequency</i> (TF)......	29
2.3.7.2.	<i>Invers Document Frequency</i> (IDF).....	30
2.3.7.3.	<i>Term Frequency-Inverse Document Frequency</i> (TF-IDF).	30
2.3.8.	<i>Confusion Matrix</i>	30
2.3.9.	<i>Synthetic Minority Over-sampling Technique</i> (SMOTE).....	32
2.3.10.	<i>Food Estate</i>	33
BAB III METODOLOGI PENELITIAN		34
3.1	Subjek dan Objek Penelitian.....	34
3.2	Identifikasi Masalah.....	34
3.3	Waktu Penelitian.....	35
3.4	Metode atau Model yang Diusulkan.....	36
3.4.1	Metode Penelitian.....	36
3.4.2.	Pengumpulan Data.....	37
3.4.3.	<i>Pre-processing</i>	39
3.4.3.1.	<i>Cleaning</i>	39
3.4.3.2	<i>Case Folding</i>	40
3.4.3.3	<i>Tokenizing</i>	41
3.4.3.4.	<i>Stopword Removal</i>	42
3.4.3.5.	<i>Stemming</i>	43
3.4.4	<i>Imbalance Data</i> dengan SMOTE.....	44
3.4.4.1.	Langkah-langkah dalam SMOTE.....	46
3.4.4.2.	Implementasi SMOTE pada Data Sentimen.	46
3.4.5.	TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	46
3.4.5.1.	Menghitung <i>Term Freugency</i> (TF)	47
3.4.5.2.	Menghitung <i>Inverse Document Frequency</i> (IDF).....	48
3.4.5.3.	Menghitung TF-IDF.....	48
3.4.6.	Klasifikasi dengan algoritma SVM.....	51
3.4.6.1.	Rancangan pelatihan <i>Support Vector Machine</i>	51
3.4.7.	<i>Evaluation</i>	52
3.4.7.1.	<i>Confusion Matrix</i>	52
3.5.	Model Pengembangan Sistem.....	53

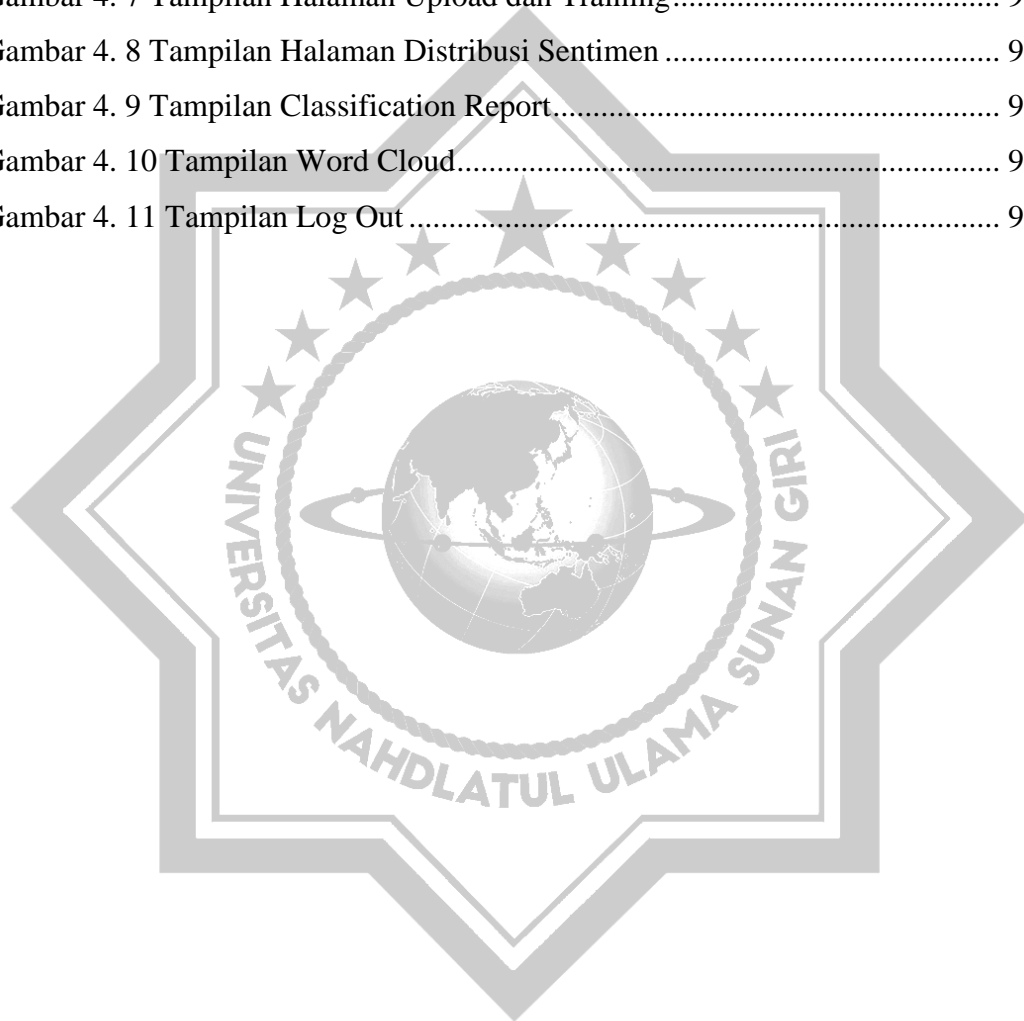
3.5.1	<i>Analysis</i>	54
3.5.1.1.	Analisis Kebutuhan Fungsional dan Non-fungsional	54
3.5.1.2.	Kebutuhan Perangkat Lunak	57
3.5.1.1.	Kebutuhan Perangkat Keras	57
3.5.1.2.	Analisis Pengguna	57
3.5.1.3.	Analisis Sistem	57
3.5.2	<i>Design</i>	58
3.5.2.1.	<i>Use Case</i>	58
3.5.2.2.	<i>Mock Up</i>	59
3.5.2.2.1.	<i>Menu Login</i>	59
3.5.2.2.2.	<i>Menu Dashboard</i>	60
3.5.2.2.3.	<i>Menu Input Data</i>	60
3.5.2.2.4.	<i>Menu Klasifikasi</i>	61
3.5.3	<i>Implementation</i>	61
3.5.4	<i>Testing</i>	61
3.6	Rencana Pengujian	61
3.6.1	Rencana Pengujian <i>Blackbox</i>	61
BAB IV HASIL DAN PEMBAHASAN		63
4.1.	Implementasi Analisis Sentimen Program Pemerintah <i>Food Estate</i>	63
4.1.1.	<i>Web Scrapping</i>	63
4.1.2.	Implementasi Pelabelan Pada Data	65
4.1.3.	<i>Input Dataset</i>	66
4.1.4.	<i>Preprocessing</i>	66
4.1.4.1.	<i>Cleaning</i>	67
4.1.4.2.	<i>Case Folding</i>	68
4.1.4.3.	<i>Tokenizing</i>	69
4.1.4.4.	<i>Stopword Removal</i>	71
4.1.4.5	<i>Stemming</i>	72
4.1.5.	<i>Ekstraksi Fitur (Term-Weighting TF-IDF)</i>	74
4.1.6.	<i>Data Mining</i>	77
4.1.6.1.	<i>Split Data</i>	77
4.1.7.	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	77
4.1.7.1.	Data Penelitian	77

4.1.7.2.	<i>Balancing Data</i> (Menyeimbangkan Data) Menggunakan Excel Dengan Menghitung Jarak <i>Euclidean</i> Dan Temukan Tetangga Terdekat.	79
4.1.7.3.	Visualisasi Proses SMOTE.	84
4.1.7.3.1.	Visualisasi Data Sebelum Proses SMOTE.....	84
4.1.7.3.2.	Visualisasi Data Setelah Proses SMOTE.....	84
4.1.8.	<i>Support Vector Machine</i> (SVM).....	86
4.1.9.	Evaluasi dengan <i>Confusion Matrix</i>	91
4.1.10.	Implementasi Sistem.....	93
4.1.10.1.	Buka <i>Browser</i>	94
4.1.10.2.	Tampilan Halaman <i>Login</i>	94
4.1.10.3.	Tampilan Halaman <i>Dashboard</i>	95
4.1.10.4.	Tampilan Halaman <i>Training Data</i>	96
4.1.10.5.	Tombol <i>Import</i> dan <i>Training Data</i>	96
4.1.10.6.	Tampilan Halaman <i>Upload</i> dan <i>Training</i>	96
4.1.10.7.	Tampilan Halaman <i>Distribusi Sentimen</i>	97
4.1.10.8.	Tampilan <i>Classification Report</i>	97
4.1.10.9.	Tampilan <i>Word Cloud</i>	98
4.1.10.10.	Tampilan <i>Log Out</i>	99
4.2.	Hasil Pengujian.....	99
4.2.1.	Hasil Uji <i>Black Box</i>	99
4.2.2.	Hasil Uji Kelayakan.....	101
BAB V	KESIMPULAN DAN SARAN	104
5.1.	Kesimpulan.....	104
5.2.	Saran Pengembangan Selanjutnya.....	104
DAFTAR PUSTAKA	105
LAMPIRAN	108

DAFTAR GAMBAR

Gambar 2. 1 Karakter Data Mining	21
Gambar 2. 2 Tahapan Data Mining.....	22
Gambar 2. 3 Hyperplane Support Vector Machine (Ii, 2022)	26
Gambar 3. 1 Alur Proses Analisis Sentimen.....	36
Gambar 3. 2 Flowchart Pengambilan Data	37
Gambar 3. 3 Proses Pengambilan Data	38
Gambar 3. 4 Review hasil pengambilan data.....	38
Gambar 3. 5 Proses pelabelan data ulasan program Food Estate.....	39
Gambar 3. 6 Preprocessing data.....	39
Gambar 3. 7 Sebelum diproses Cleaning	40
Gambar 3. 8 setelah di proses cleaning.....	40
Gambar 3. 9 sebelum diproses Case Folding	40
Gambar 3. 10 setelah diproses Case Folding	41
Gambar 3. 11 sebelum diproses Tokenizing	41
Gambar 3. 12 setelah diproses Tokenizing	42
Gambar 3. 13 sebelum diproses Stopword Removal	42
Gambar 3. 14 setelah diproses Stopword Removal	43
Gambar 3. 15 sebelum diproses Stemming.....	44
Gambar 3. 16 setelah diproses Stemming	44
Gambar 3. 17 Alur Proses SMOTE	45
Gambar 3. 18 Flowchart Klasifikasi SVM.....	52
Gambar 3. 19 Model Pengembangan Perangkat Lunak.....	54
Gambar 3. 20 Use Case Diagram.....	58
Gambar 3. 21 Tampilan Menu Login.....	59
Gambar 3. 22 Tampilan Menu Dashboard.....	60
Gambar 3. 23 Tampilan Menu Input Data	60
Gambar 3. 24 Tampilan Menu Klasifikasi	61
Gambar 4. 1 Hasil Perhitungan Titik Sintetik	82
Gambar 4. 2 Halaman Browser.....	94

Gambar 4. 3 Tampilan Halaman Login.....	95
Gambar 4. 4 Tampilan Halaman Dashboard.....	95
Gambar 4. 5 Tampilan Halaman Training data.....	96
Gambar 4. 6 Tampilan Tombol Import dan Training Data.....	96
Gambar 4. 7 Tampilan Halaman Upload dan Training.....	97
Gambar 4. 8 Tampilan Halaman Distribusi Sentimen	97
Gambar 4. 9 Tampilan Classification Report.....	98
Gambar 4. 10 Tampilan Word Cloud.....	98
Gambar 4. 11 Tampilan Log Out	99



UNUGIRI

DAFTAR TABEL

Tabel 2. 1 Penelitian terlebih dahulu.....	8
Tabel 2. 2 Persamaan Kernel pada Support Vector Machine (Ii, 2022)	28
Tabel 2. 3 Confusion Matrix	31
Tabel 3. 1 Proses Pengerjaan Skripsi	35
Tabel 3. 2 Analisis Kebutuhan Fungsional dan Non-fungsional	54
Tabel 3. 3 Perangkat Lunak	57
Tabel 3. 4 Perangkat Keras	57
Tabel 3. 5 Penjelasan Use Case Diagram.....	58
Tabel 3. 6 Hasil Rencana Pengujian Blackbox	62
Tabel 4. 1 Web Scrapping.....	63
Tabel 4.2 Labeling Data.....	65
Tabel 4.3 Proses Cleaning Data.....	67
Tabel 4.4 Hasil Sebelum dan Sesudah Cleaning.....	68
Tabel 4.5 Proses Case Folding.....	68
Tabel 4.6 Hasil Sebelum dan Sesudah Case Folding.....	69
Tabel 4.7 Proses Tokenizing.....	69
Tabel 4.8 Hasil Sebelum dan Sesudah Tokenizing.....	70
Tabel 4. 9 Proses Stopword Removal	71
Tabel 4. 10 Hasil Sebelum dan Sesudah Stopword Removal	72
Tabel 4. 11 Proses Stemming.....	73
Tabel 4. 12 Hasil Sebelum dan Sesudah Stemming.....	73
Tabel 4. 13 Hasil Ekstraksi TF-IDF.....	75
Tabel 4. 14 Data Dummy	78
Tabel 4. 15 Hasil Random pada Data Dummy	79
Tabel 4. 16 Hasil perhitungan Jarak Euclidean.....	80
Tabel 4. 17 Hasil Perhitungan Tetangga Terdekat Kedua	81
Tabel 4. 18 Hasil Penambahan Sintetik pada Data Asli	83
Tabel 4. 19 Data untuk perhitungan SVM	86
Tabel 4. 20 Hasil Perhitungan SVM	93

Tabel 4. 21 Confusion Matrix	91
Tabel 4. 22 Kasus dan Hasil Pengujian Black Box.....	100
Tabel 4. 23 Tabel Hasil Uji Kelayakan.....	101
Tabel 4. 24 Total Skor.....	102

