



SKRIPSI

**ANALISIS PERFORMANSI INSET DAN VADER
LEXICON PADA PELABELAN SENTIMEN
MENGUNAKAN ALGORITMA SVM DAN
OPTIMASI PSO**

RAFI ADITYA MAHENDRA

NPM 20081010208

DOSEN PEMBIMBING

Yisti Vita Via, S.ST. M.Kom.

Chrystia Aji Putra, S.Kom, M.T

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2025



SKRIPSI

**ANALISIS PERFORMANSI INSET DAN VADER
LEXICON PADA PELABELAN SENTIMEN
MENGUNAKAN ALGORITMA SVM DAN
OPTIMASI PSO**

RAFI ADITYA MAHENDRA

NPM 20081010208

DOSEN PEMBIMBING

Yisti Vita Via, S.ST. M.Kom.

Chrystia Aji Putra, S.Kom, M.T

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

ANALISIS PERFORMANSI INSET DAN VADER LEXICON PADA PELABELAN SENTIMEN MENGUNAKAN ALGORITMA SVM DAN OPTIMASI PSO

Oleh :

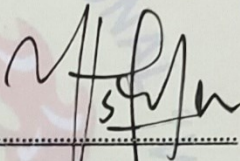
RAFI ADITYA MAHENDRA

NPM. 20081010208

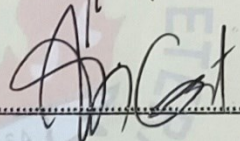
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 21 Januari 2025

Menyetujui

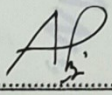
Yisti Vita Via, S.ST. M.Kom.
NIP. 19860425 2021212 001


..... (Pembimbing I)

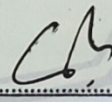
Chrystia Aji Putra, S.Kom, M.T
NIP. 19861008 2021211 001


..... (Pembimbing II)

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT
NIP. 222198 60 816400

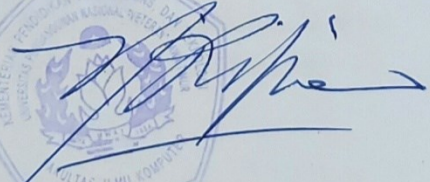

..... (Ketua Penguji)

Eva Yulia Puspaningrum, S.Kom, M.Kom
NIP. 19890705 2021212 002


..... (Anggota Penguji)

Mengetahui

Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT

NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**ANALISIS PERFORMANSI INSET DAN VADER LEXICON PADA PELABELAN SENTIMEN
MENGUNAKAN ALGORITMA SVM DAN OPTIMASI PSO**

Oleh:

RAFI ADITYA MAHENDRA

NPM. 20081010208



Menyetujui,

**Koordinator Program Studi Informatika
Fakultas Ilmu Komputer**

Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 2021212 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama mahasiswa / NPM : RAFI ADITYA MAHENDRA
Program Studi : Informatika
Dosen Pembimbing : 1. Yisti Vita Via, S.ST. M.Kom
2. Chrystia Aji Putra, S.Kom, M.T

dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul.

“ANALISIS PERFORMANSI INSET DAN VADER LEXICON PADA PELABELAN SENTIMEN MENGGUNAKAN ALGORITMA SVM DAN OPTIMASI PSO”

adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah. Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Universitas Pembangunan Nasional Veteran Jawa Timur



Surabaya, 21 Januari 2025

Mahasiswa



Rafi Aditya Mahendra

NPM. 20081010208

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Rafi Aditya Mahendra / 20081010208
Judul Skripsi : Analisis Performansi InSet Dan VADER Lexicon
Pada Pelabelan Sentimen Menggunakan
Algoritma SVM Dan Optimasi PSO
Dosen Pembimbing : 1. Yisti Vita Via, S.ST. M.Kom.
2. Chrystia Aji Putra, S.Kom, M.T

Analisis sentimen adalah sebuah studi yang menganalisis pendapat, sentimen, penilaian, sikap, evaluasi, dan emosi seseorang terhadap entitas seperti sebuah masalah dan peristiwa. Salah satu pendekatan yang umum digunakan dalam analisis sentimen adalah metode *lexicon-based*, sebuah teks dianalisis berdasarkan kamus sentimen yang mengelompokkan kata-kata ke dalam kategori sentimen positif atau negatif. Kamus sentimen yang dapat digunakan antara lain yaitu InSet dan VADER Lexicon. Perbandingan akurasi yang dihasilkan dari data dengan pelabelan InSet dan VADER Lexicon dilakukan dengan tujuan untuk mengetahui *lexicon* mana yang lebih efektif untuk digunakan dalam proses pelabelan data. Dataset yang digunakan dalam pengujian pada penelitian ini yaitu *tweet* pada aplikasi X (Twitter) mengenai pengungsi rohingya di Indonesia. Selain itu, klasifikasi data akan dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dan optimasi menggunakan metode *Particle Swarm Optimization* (PSO) untuk mencari parameter terbaik yang dapat meningkatkan nilai akurasi yang dihasilkan dari proses pengujian. Penelitian ini juga membandingkan nilai akurasi yang dihasilkan dari klasifikasi data menggunakan SVM tanpa optimasi dan klasifikasi data menggunakan SVM serta optimasi menggunakan PSO (SVM+PSO). Hasil dari proses pelabelan data menggunakan InSet Lexicon yaitu 886 sentimen positif, 2906 sentimen negatif, dan 353 sentimen netral. Sedangkan, pelabelan data menggunakan VADER Lexicon menghasilkan 1353 sentimen positif, 1942 sentimen negatif, dan 850 netral. Hasil akurasi tertinggi didapatkan pada data dengan pelabelan InSet Lexicon serta klasifikasi data menggunakan SVM dan optimasi menggunakan PSO yang menghasilkan akurasi sebesar 90.7%.

Kata Kunci : Support Vector Machine, InSet Lexicon, VADER Lexicon, Particle Swarm Optimization

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Rafi Aditya Mahendra / 20081010208
Thesis Title : Performance Analysis of InSet and VADER
Lexicons in Sentiment Labeling Using SVM
Algorithm and PSO Optimization
Advisor : 1. Yisti Vita Via, S.ST. M.Kom.
2. Chrystia Aji Putra, S.Kom, M.T

Sentiment analysis is a study that analyzes people's opinions, sentiments, judgments, attitudes, evaluations, and emotions towards entities such as problems and events. One of the commonly used approaches in sentiment analysis is the lexicon-based method, where a text is analyzed based on a sentiment dictionary that groups words into positive or negative sentiment categories. Sentiment dictionaries that can be used include InSet and VADER Lexicon. A comparison of the accuracy generated from data with InSet and VADER Lexicon labeling is carried out with the aim of knowing which lexicon is more effective to use in the data labeling process. The dataset used in testing in this research is tweets on application X (Twitter) regarding rohingya refugees in Indonesia. In addition, data classification will be carried out using the Support Vector Machine (SVM) algorithm and optimization using the Particle Swarm Optimization (PSO) method to find the best parameters that can increase the accuracy value resulting from the testing process. This research also compares the accuracy value generated from data classification using SVM without optimization and data classification using SVM and optimization using PSO (SVM+PSO). The results of the data labeling process using InSet Lexicon are 886 positive sentiments, 2906 negative sentiments, and 353 neutral sentiments. Meanwhile, data labeling using VADER Lexicon produces 1353 positive sentiments, 1942 negative sentiments, and 850 neutral sentiments. The highest accuracy results are obtained on data with InSet Lexicon labeling and data classification using SVM and optimization using PSO which produces an accuracy of 90.7%.

Keywords : Support Vector Machine, InSet Lexicon, VADER Lexicon, Particle Swarm Optimization

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Analisis Performansi InSet Dan VADER Lexicon Pada Pelabelan Sentimen Menggunakan Algoritma SVM Dan Optimasi PSO”** dapat terselesaikan dengan baik.

Penulis banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Dosen Wali penulis dan Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran“ Jawa Timur.
4. Ibu Yisti Vita Via, S.ST. M.Kom. selaku dosen pembimbing utama yang telah memberikan dukungan, arahan, serta wawasan bagi penulis selama proses penelitian skripsi berlangsung.
5. Bapak Chrystia Aji Putra, S.Kom, M.T selaku dosen pembimbing yang telah memberikan dukungan, arahan, serta wawasan bagi penulis selama proses penelitian skripsi berlangsung.
6. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom. selaku penanggung jawab skripsi yang telah mengkoordinasikan seluruh proses skripsi ini sehingga dapat berjalan dengan lancar.
7. Seluruh Dosen Program Studi Informatika yang telah memberikan ilmu serta menambah wawasan kepada penulis selama masa perkuliahan
8. Kedua orang tua serta adik penulis yang selalu memberikan semangat, dorongan, dan fasilitas sehingga penulis dapat menyelesaikan penelitian ini dengan lancar.
9. Kepada seluruh sahabat baik penulis yang selalu memberikan motivasi dan semangat sehingga penulis dapat menyelesaikan proses pengerjaan skripsi dengan baik.

10. Semua teman-teman program studi Informatika Angkatan 2020 yang tidak dapat penulis sebutkan satu per satu yang telah saling berbagi pengalaman selama proses penulisan skripsi.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 21 Januari 2025

Rafi Aditya Mahendra

DAFTAR ISI

LEMBAR PENGESAHAN	iii
SURAT PERNYATAAN ORISINALITAS.....	vii
ABSTRAK.....	ix
ABSTRACT.....	xi
KATA PENGANTAR	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	5
1.5. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terdahulu.....	7
2.2. Analisis Sentimen.....	9
2.3. X (Twitter).....	10
2.4. <i>Text Preprocessing</i>	11
2.5. <i>Lexicon-based</i>	13
2.6. InSet Lexicon.....	13
2.7. VADER Lexicon	14
2.8. <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	16
2.9. <i>Support Vector Machine (SVM)</i>	17
2.10. <i>Particle Swarm Optimization (PSO)</i>	19
2.11. <i>Confusion Matrix</i>	20
BAB III DESAIN DAN IMPLEMENTASI SISTEM	23
3.1. Metode Penelitian.....	23
3.2. Desain Sistem	24
3.2.1. Pengumpulan Data	24
3.2.2. Preprocessing Data.....	24
A. <i>Case-Folding</i>	25
B. <i>Cleaning</i>	25
C. <i>Tokenizing</i>	26
D. <i>Normalisasi</i>	27
E. <i>Stopword Removal</i>	27

F.	<i>Stemming</i>	28
3.2.3.	Pelabelan Data.....	29
A.	InSet Lexicon	29
B.	VADER Lexicon.....	30
3.2.5.	Klasifikasi SVM.....	35
3.2.6.	Optimasi PSO.....	37
3.2.7.	Evaluasi.....	42
3.2.8.	Skenario Pengujian.....	43
BAB IV PENGUJIAN DAN ANALISA		45
4.1.	Metode Pengujian.....	45
4.2.	Hasil Pengujian.....	46
4.2.1.	Pengumpulan Data	46
4.2.2.	<i>Preprocessing Data</i>	47
4.2.3.	Pelabelan Data.....	59
A.	InSet Lexicon	59
B.	VADER Lexicon.....	62
4.2.4.	Pembobotan Data	65
4.2.5.	Pembagian Data	67
4.2.6.	Klasifikasi SVM.....	68
4.2.7.	Optimasi PSO.....	69
4.2.8.	Evaluasi.....	70
4.2.9.	Hasil Pengujian Skenario	72
A.	InSet Lexicon	72
B.	VADER Lexicon.....	92
4.2.10.	Perbandingan Hasil Pengujian Skenario.....	111
BAB V PENUTUP		117
5.1.	Kesimpulan.....	117
5.2.	Saran Pengembangan.....	118
DAFTAR PUSTAKA		119

DAFTAR GAMBAR

Gambar 2. 1 <i>Support Vector Machine</i>	18
Gambar 3. 1 Alur Metode Penelitian	23
Gambar 3. 2 Alur <i>Preprocessing Data</i>	25
Gambar 3. 3 Alur TF-IDF	32
Gambar 3. 4 Alur SVM.....	35
Gambar 3. 5 Alur PSO	38
Gambar 4. 1 Hasil Pengumpulan Data	47
Gambar 4. 2 Hasil Penghapusan Kolom	47
Gambar 4. 3 Contoh 1 Kalimat Mentah.....	48
Gambar 4. 4 Contoh 1 Kalimat Hasil <i>Case-Folding</i>	48
Gambar 4. 5 Hasil Proses <i>Case-Folding</i>	49
Gambar 4. 6 Contoh 1 Kalimat Hasil <i>Cleaning</i>	51
Gambar 4. 7 Hasil Proses <i>Cleaning</i>	52
Gambar 4. 8 Contoh 1 Kalimat Hasil <i>Tokenizing</i>	53
Gambar 4. 9 Hasil Proses <i>Tokenizing</i>	53
Gambar 4. 10 Contoh 1 Kalimat Proses Normalisasi	55
Gambar 4. 11 Hasil Proses Normalisasi.....	55
Gambar 4. 12 Contoh 1 Kalimat Hasil <i>Stopword Removal</i>	56
Gambar 4. 13 Hasil Proses <i>Stopwords Removal</i>	56
Gambar 4. 14 Contoh 1 Kalimat Hasil <i>Stemming</i>	57
Gambar 4. 15 Hasil Proses <i>Stemming</i>	58
Gambar 4. 16 Hasil Proses Hapus Data Duplikat	59
Gambar 4. 17 Contoh 1 Kalimat Hasil Pelabelan InSet Lexicon.....	61
Gambar 4. 18 Hasil Pelabelan InSet Lexicon	61
Gambar 4. 19 Hasil Proses <i>Translate Data</i>	63
Gambar 4. 20 Contoh 1 Kalimat Hasil Pelabelan VADER Lexicon	64
Gambar 4. 21 Hasil Pelabelan Data VADER Lexicon	64
Gambar 4. 22 Contoh 1 Kalimat Hasil Proses TF-IDF	66
Gambar 4. 23 Hasil TF-IDF	66
Gambar 4. 24 Hasil TF-IDF Pelabelan InSet Lexicon.....	67
Gambar 4. 25 Hasil TF-IDF Pelabelan VADER Lexicon	67

Gambar 4. 26 <i>Confusion Matrix</i> Skenario 1	73
Gambar 4. 27 <i>Classification Report</i> Skenario 1	73
Gambar 4. 28 <i>Confusion Matrix</i> Skenario 2	74
Gambar 4. 29 <i>Classification Report</i> Skenario 2.....	75
Gambar 4. 30 <i>Confusion Matrix</i> Skenario 3	76
Gambar 4. 31 <i>Classification Report</i> Skenario 3.....	76
Gambar 4. 32 <i>Confusion Matrix</i> Skenario 4	77
Gambar 4. 33 <i>Classification Report</i> Skenario 4.....	78
Gambar 4. 34 <i>Confusion Matrix</i> Skenario 5	79
Gambar 4. 35 <i>Classification Report</i> Skenario 5.....	79
Gambar 4. 36 <i>Confusion Matrix</i> Skenario 6	80
Gambar 4. 37 <i>Classification Report</i> Skenario 6.....	81
Gambar 4. 38 <i>Confusion Matrix</i> Skenario 7	82
Gambar 4. 39 <i>Classification Report</i> Skenario 7.....	82
Gambar 4. 40 <i>Confusion Matrix</i> Skenario 8	83
Gambar 4. 41 <i>Classification Report</i> Skenario 8.....	84
Gambar 4. 42 <i>Confusion Matrix</i> Skenario 9	85
Gambar 4. 43 <i>Classification Report</i> Skenario 9.....	85
Gambar 4. 44 <i>Best Parameter</i> Skenario 10.....	86
Gambar 4. 45 <i>Confusion Matrix</i> Skenario 10	86
Gambar 4. 46 <i>Classification Report</i> Skenario 10.....	87
Gambar 4. 47 <i>Best Parameter</i> Skenario 11.....	88
Gambar 4. 48 <i>Confusion Matrix</i> Skenario 11	88
Gambar 4. 49 <i>Classification Report</i> Skenario 11.....	89
Gambar 4. 50 <i>Best Parameter</i> Skenario 12.....	90
Gambar 4. 51 <i>Confusion Matrix</i> Skenario 12	90
Gambar 4. 52 <i>Classification Report</i> Skenario 12.....	91
Gambar 4. 53 <i>Confusion Matrix</i> Skenario 13	92
Gambar 4. 54 <i>Classification Report</i> Skenario 13.....	93
Gambar 4. 55 <i>Confusion Matrix</i> Skenario 14	94
Gambar 4. 56 <i>Classification Report</i> Skenario 14.....	94
Gambar 4. 57 <i>Confusion Matrix</i> Skenario 15	95

Gambar 4. 58 <i>Classification Report</i> Skenario 15.....	96
Gambar 4. 59 <i>Confusion Matrix</i> Skenario 16	97
Gambar 4. 60 <i>Classification Report</i> Skenario 16.....	97
Gambar 4. 61 <i>Confusion Matrix</i> Skenario 17	98
Gambar 4. 62 <i>Classification Report</i> Skenario 17.....	99
Gambar 4. 63 <i>Confusion Matrix</i> Skenario 18	100
Gambar 4. 64 <i>Classification Report</i> Skenario 18.....	100
Gambar 4. 65 <i>Confusion Matrix</i> Skenario 19	101
Gambar 4. 66 <i>Classification Report</i> Skenario 19.....	102
Gambar 4. 67 <i>Confusion Matrix</i> Skenario 20	103
Gambar 4. 68 <i>Classification Report</i> Skenario 20.....	103
Gambar 4. 69 <i>Confusion Matrix</i> Skenario 21	104
Gambar 4. 70 <i>Classification Report</i> Skenario 21.....	105
Gambar 4. 71 <i>Best Parameter</i> Skenario 22.....	106
Gambar 4. 72 <i>Confusion Matrix</i> Skenario 22	106
Gambar 4. 73 <i>Classification Report</i> Skenario 22.....	107
Gambar 4. 74 <i>Best Parameter</i> Skenario 23.....	108
Gambar 4. 75 <i>Confusion Matrix</i> Skenario 23	108
Gambar 4. 76 <i>Classification Report</i> Skenario 23.....	109
Gambar 4. 77 <i>Best Parameter</i> Skenario 24.....	109
Gambar 4. 78 <i>Confusion Matrix</i> Skenario 24	110
Gambar 4. 79 <i>Classification Report</i> Skenario 24.....	110

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2. 1 Contoh Kata Positif dan Negatif InSet Lexicon	14
Tabel 2. 2 Contoh Bobot Kata VADER Lexicon.....	15
Tabel 2. 3 Confusion Matrix	21
Tabel 3. 1 Contoh Proses <i>Case-Folding</i>	25
Tabel 3. 2 Contoh Proses <i>Cleaning</i>	26
Tabel 3. 3 Contoh Proses <i>Tokenizing</i>	26
Tabel 3. 4 Contoh Proses Normalisasi	27
Tabel 3. 5 Contoh Proses <i>Stopword Removal</i>	27
Tabel 3. 6 Contoh Proses <i>Stemming</i>	28
Tabel 3. 7 Contoh Pelabelan Data InSet Lexicon	29
Tabel 3. 8 Contoh Pelabelan Data VADER Lexicon.....	30
Tabel 3. 9 Contoh Kalimat	33
Tabel 3. 10 Hasil Perhitungan TF-IDF	34
Tabel 3. 11 Hasil Perhitungan Kernel.....	36
Tabel 3. 12 Inisialisasi Partikel	40
Tabel 3. 13 Hasil Evaluasi Fungsi Objektif	40
Tabel 3. 14 Hasil Iterasi ke-1 PSO.....	41
Tabel 3. 15 Hasil Iterasi ke-2 PSO.....	41
Tabel 3. 16 <i>Confusion Matrix</i>	42
Tabel 3. 17 Tabel Skenario Pengujian	43
Tabel 4. 1 Hasil Pengujian Skenario	112

Halaman ini sengaja dikosongkan