

**PERBANDINGAN PERFORMA LABELING LEXICON INSET DAN  
LEXICON VADER PADA ANALISA SENTIMEN ROHINGYA DI  
APLIKASI X DENGAN SUPPORT VECTOR MACHINE**

**SKRIPSI**



**Oleh :**

**MUHAMMAD FERNANDA NAUFAL FATHONI**

**20081010257**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA  
TIMUR**

**2024**

# LEMBAR PENGESAHAN

## SKRIPSI

**Judul : PERBANDINGAN PERFORMA LABELING LEXICON  
INSET DAN LEXICON VADER PADA ANALISA SENTIMEN  
ROHINGYA DI APLIKASI X DENGAN SUPPORT VECTOR  
MACHINE**

**Oleh : MUHAMMAD FERNANDA NAUFAL FATHONI**

**NPM : 20081010257**

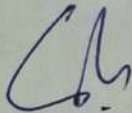
**Telah Diseminarkan Dalam Ujian Skripsi Pada:**

**Hari Selasa, Tanggal 22 Mei 2024**

### Mengetahui

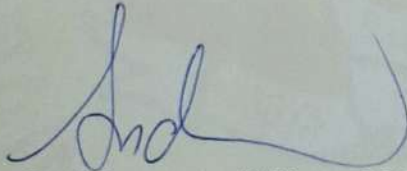
#### Dosen Pembimbing

1.



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

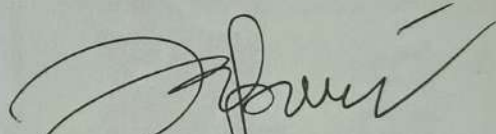
2.



Andreas Nugroho Sihananto, S.Kom, M.Kom.  
NPT. 211199 00 412271

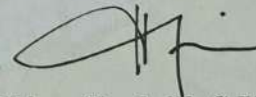
#### Dosen Penguji

1.



Dr. Rr. Ani Dijan Rahajoe, ST., M.Cs.  
NIP. 19730512 200501 2003

2.



Afina Lina Nurlaili, S.Kom, M.Kom.  
NIP.1993121 3202203 2010

### Menyetujui

#### Dekan

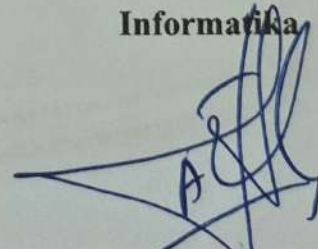
#### Fakultas Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.  
NIP. 19681126 199403 2 001

#### Koordinator Program Studi

#### Informatika



Fetty Tri Anggraeny, S.Kom, M.Kom.  
NIP. 19820211 2021212 005

## SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Muhammad Fernanda Naufal Fathoni

NPM : 20081010257

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

“Perbandingan Performa Labeling *Lexicon* Inset dan *Lexicon* Vader pada Analisa Sentimen Rohingya di Aplikasi X dengan *Support Vector Machine*”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.



Surabaya, 2 Juni 2024

Hormat saya,



Muhammad Fernanda Naufal Fathoni

NPM. 20081010257

# PERBANDINGAN PERFORMA LABELING LEXICON INSET DAN LEXICON VADER PADA ANALISA SENTIMEN ROHINGYA DI APLIKASI X DENGAN SUPPORT VECTOR MACHINE

**Nama Mahasiswa** : **Muhammad Fernanda Naufal Fathoni**  
**NPM** : **20081010257**  
**Program Studi** : **Teknik Informatika**  
**Dosen Pembimbing** : **1. Eva Yulia Puspaningrum, S.Kom, M.Kom**  
**2. Andreas Nugroho Sihananto, S.Kom., M.Kom.**

## Abstrak

Masuknya Rohingya di Indonesia menjadi bahan perbincangan di sosial media. Salah satu cara untuk melihat bagaimana respon masyarakat terhadap hal tersebut dibutuhkan analisa sentimen. Banyaknya data menjadikan masalah efisiensi waktu dalam proses pelabelan, maka dari itu dibutuhkannya kamus *lexicon* untuk proses pelabelan tersebut. Kamus *lexicon* merupakan adalah suatu kamus yang berisikan data kata dengan nilai atau bobot kata yang kemudian dikalkulasikan untuk menentukan suatu sentimen sebuah kalimat. Adanya *lexicon* membuat proses pelabelan data training jadi lebih cepat sehingga mempermudah proses pengolahan data. Di era saat ini data yang berkembang dan beredar sangat pesat sehingga dibutuhkan waktu yang cepat dan efisien. Meskipun cepat dan mempermudah dalam penyelesaian masalah masih perlu dipertanyakan akurasi yang dihasilkan apabila menggunakan pelabelan *lexicon* tersebut.

Dilakukan perbandingan proses pelabelan antara *lexicon* InSet dan *lexicon* VADER untuk mengetahui akurasi dari pelabelan tersebut. Hal itu dilakukan kombinasi *lexicon* dengan machine learning metode support vector machine dan pembobotan TF-IDF. Hasil yang didapatkan akan dievaluasi kinerjanya dengan menggunakan confusion matrix dan dilakukan perhitungan akurasi dari hasil tersebut.

Data yang didapatkan dari sosial media X sebanyak 9117 baris dan pada saat dilabeli dengan *lexicon* InSet menghasilkan 5241 sentimen negatif, 1369 sentimen positif, dan 521 sentimen netral. Lalu hasil pelabelan dengan *lexicon* VADER menghasilkan 2749 sentimen positif, 2523 sentimen negatif, dan 1881 sentimen netral. Setelah diberikan label dilakukan

pembobotan lalu diproses dengan SVM dan dihitung *confusion* matrixnya dengan hasil akurasi InSet *lexicon* memiliki rata rata sebesar 85,8% sedangkan *lexicon* VADER SVM memiliki rata rata 82,65%

***Kata Kunci : Lexicon, InSet, VADER, TF-IDF, SVM***

## KATA PENGANTAR

Alhamdulillahirobilalamin, Segala puji Bagi Allah Tuhan yang Maha Esa, Maha Pengasih dan Penyayang yang telah memberikan rahmat serta petunjuk sehingga skripsi ini dapat terselesaikan. Skripsi yang berjudul “PERBANDINGAN PERFORMA LABELING LEXICON INSET DAN LEXICON VADER PADA ANALISA SENTIMEN ROHINGYA DI APLIKASI X DENGAN SUPPORT VECTOR MACHINE” adalah skripsi yang ditulis berdasarkan ketertarikan penulis sejak belajar mengenai pengolahan data sehingga penulis tertarik mengambil judul yang berkaitan terkait data mining. Pada waktu itu berita mengenai pengungsi rohingya yang tiba di aceh dan sempat menjadi trending topik di sosial media X menjadikan penulis tertarik mengambil data tersebut untuk diteliti.

Ucapan terima kasih tak terhingga saya haturkan kepada Dosen Pembimbing 1 ibu Eva Yulia Puspaningrum yang telah membantu serta meluangkan waktu dalam mengarahkan serta memberikan nasihat kepada penulis dalam penyusunan hingga skripsi ini dapat terselesaikan dengan baik. Ucapan terima kasih juga dihaturkan kepada dosen Pembimbing 2 bapak Andreas Nugroho Sihananto yang sudah meluangkan waktunya untuk diskusi terkait penelitian tersebut sehingga penulis tidak tersesat dijalan.

Dalam prosesnya penulis juga mengucapkan terima kasih atas bantuan dari berbagai pihak lainnya. Maka dari itu penulis mengucapkan terima kasih kepada berbagai pihak sebagai berikut :

1. Kedua orang tua yang telah membantu dalam proses belajar selama ini
2. Bapak Ibu dosen prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur
3. Para jajaran staff Universitas Pembangunan Nasional Veteran Jawa Timur yang turut membantu dalam proses belajar selama ini
4. Universitas Pembangunan Nasional Veteran Jawa Timur sebagai kampus yang memberikan wadah untuk belajar
5. Teman teman angkatan 20 informatika universitas pembangunan nasional veteran jawa timur
6. Freedive malang sebagai wadah dalam menyalurkan stress selama pengerjaan berlangsung
7. Galuh Alfisyah mahasiswa matematika universitas negeri surabaya telah membantu dan mensupport penuh dalam penulisan ini.

8. Lintang sebagai teman selama perkuliahan dan diskusi selama penelitian berlangsung.
9. Saudara yang telah membantu support dan doa
10. Semua pihak yang belum disebutkan membantu dalam pengerjaan dan proses pengerjaan selama ini.

Sebagai penutup, penulis mengharapkan tulisan ini menjadi nilai manfaat dan pahala jariyah bagi penulis dan menambah wawasan serta pengetahuan bagi pembacanya. Penulis menyadari bahwa tulisan ini masih jauh dari kata sempurna sehingga penulis terbuka menerima kritik dan saran sebagai sarana evaluasi penulis kedepannya.

Surabaya,

Penulis

Muhammad Fernanda Naufal Fathoni

NPM. 20081010257

# DAFTAR ISI

<b>LEMBAR PENGESAHAN</b> .....	ii
<b>KATA PENGANTAR</b> .....	v
<b>DAFTAR ISI</b> .....	vii
<b>DAFTAR TABEL</b> .....	x
<b>DAFTAR GAMBAR</b> .....	xii
<b>DAFTAR KODE PROGRAM</b> .....	xiii
<b>BAB I PENDAHULUAN</b> .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan.....	3
1.4. Manfaat.....	3
1.5. Batasan Masalah.....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	5
2.1 Penelitian Sebelumnya .....	5
2.2 Rohingya .....	5
2.3 Aplikasi X .....	6
2.4 Analisa Sentimen.....	7
2.3.1 Lexicon Based.....	8
2.3.2 Lexicon InSet .....	9
2.3.3 <i>Lexicon VADER</i> .....	10
2.5 Machine Learning .....	11
2.6 Text Preprocessing .....	11
2.7 TF-IDF (Term Frequency Inverse Document Frequency).....	12
2.8 SVM (Support Vector Machine) .....	13
2.9 Data Latih dan Uji .....	15
2.10 Confusion Matrix .....	15
<b>BAB III METODOLOGI PENELITIAN</b> .....	17
3.1 Metode Penelitian.....	17
3.2 Pengumpulan Data .....	17
3.3 Preprocessing .....	18



3.3.1	Penggabungan Data .....	19
3.3.2	Cleaning Data.....	19
3.3.3	Tokenizing .....	20
3.3.4	Normalisasi .....	21
3.3.5	Translate Data .....	22
3.3.6	Filtering / Stopword Removal.....	23
3.3.7	Stemming .....	25
3.4	Lexicon Based.....	26
3.4.1	InSet Lexicon.....	26
3.4.2	VADER Lexicon.....	28
3.5	Split Data Uji dan Data Latih.....	29
3.6	TF-ID.....	29
3.7	Support Vector Machine .....	33
3.8	Skenario Pengujian.....	38
3.9	Confusion Matrix .....	39
<b>BAB IV</b>	<b>PEMBAHASAN .....</b>	<b>40</b>
4.1	Proses Penelitian .....	40
4.1.1	Perangkat Keras .....	40
4.1.2	Perangkat Lunak .....	40
4.2	Pengumpulan Data .....	41
4.3	Data Preprocessing.....	41
4.2.1	Penggabungan Data .....	42
4.2.2	Cleaning Data .....	43
4.2.3	Tokenizing .....	47
4.2.4	Normalisasi.....	48
4.2.5	Translate Data.....	50
4.2.6	Filtering / Stopword Removal .....	51
4.2.7	Stemming.....	55
4.4	Lexicon Based.....	59
4.3.1	InSet Lexicon.....	59
4.3.2	VADER Lexicon .....	62
4.5	Split Data Uji dan Data Latih.....	64
4.6	TF-IDF .....	66

4.7	TF-IDF Standarisasi .....	67
4.8	Support Vector Machine .....	68
4.9	Confusion Matrix .....	69
<b>BAB V PENUTUP.....</b>		<b>75</b>
5.1	Kesimpulan.....	75
5.2	Saran.....	76
<b>DAFTAR PUSTAKA.....</b>		<b>77</b>

## DAFTAR TABEL

Tabel 2. 1 Daftar Bobot Kata pada Lexicon InSet.....	9
Tabel 2.2 Daftar Bobot Kata Lexicon Vader.....	10
Tabel 2.3 Confusion Matrix .....	15
Tabel 3.1 Data Cleaning .....	20
Tabel 3.2 Tokenizing.....	21
Tabel 3.3 Normalisasi .....	22
Tabel 3.4 Kamus Normalisasi .....	22
Tabel 3.5 Tabel Normalisasi.....	23
Tabel 3.6 Filtering Bahasa Indonesia.....	24
Tabel 3.7 Filtering Bahasa Inggris .....	25
Tabel 3. 8 Stemming .....	25
Tabel 3.9 Perhitungan Bobot Lexicon InSet.....	27
Tabel 3.10 Perhitungan Bobot VADER Lexicon .....	28
Tabel 3.11 Kalimat Pada Setiap Dokumen.....	30
Tabel 3.12 Perhitungan Term Frequency.....	31
Tabel 3.13 Perhitungan Kemunculan Kata .....	31
Tabel 3.14 Hasil Perhitungan TF-IDF .....	32
Tabel 3.15 Hasil Perhitungan Kernel.....	34
Tabel 3.16 Skenario Pengujian .....	38
Tabel 4.1 Tabel Perangkat Lunak.....	40
Tabel 4.2 Hasil Data Cleaning .....	45
Tabel 4.3 Hasil Tokenize .....	47
Tabel 4.4 Hasil Normalisasi Data .....	49
Tabel 4.5 Hasil Translate Data .....	50
Tabel 4.6 Hasil Filtering Data Bahasa Indonesia.....	53
Tabel 4.7 Hasil Filtering Data Bahasa Inggris .....	54
Tabel 4.8 Hasil Stemming Data Bahasa Indonesia .....	56
Tabel 4.9 Hasil Stemming Bahasa Inggris.....	58
Tabel 4.10 Hasil Labeling InSet Lexicon .....	61
Tabel 4.11 Tabel Jumlah Labeling Lexicon InSet .....	61
Tabel 4.12 Hasil Labeling VADER Lexicon .....	63

<b>Tabel 4.13 Jumlah Labeling Lexicon VADER.....</b>	<b>64</b>
<b>Tabel 4.14 Hasil Split Data Latih dan Uji.....</b>	<b>66</b>
<b>Tabel 4.15 Hasil Confusion Matrix .....</b>	<b>69</b>
<b>Tabel 4.16 Hasil Akurasi Model .....</b>	<b>72</b>

## DAFTAR GAMBAR

Gambar 2.1 Migrasi Rohingya .....	6
Gambar 2.2 Penggunaan Analisa Sentimen .....	7
Gambar 2.3 Pendekatan Analisa Sentimen .....	8
Gambar 2.4 Ilustrasi Support Vector Machine .....	14
Gambar 3.1 Diagram Metodologi Penelitian.....	17
Gambar 3.2 Pengumpulan Data .....	17
Gambar 3.3 Flowchart Preprocessing.....	18
Gambar 3.4 Proses Penggabungan Data .....	19
Gambar 3.5 Proses Normalisasi Data .....	21
Gambar 3.6 Flowchart Stopword Removal.....	23
Gambar 3.7 Proses Labeling .....	26
Gambar 3.8 Flowchart Perhitungan TF-IDF .....	29
Gambar 3.9 Flowchart Support Vector Machine .....	332
Gambar 4.1 Kumpulan Dataset.....	42
Gambar 4.2 Hasil Penggabungan Data.....	43
Gambar 4.3 Hasil Pengambilan Kolom yang Digunakan .....	44
Gambar 4.4 Hasil Labeling InSet Lexicon .....	62
Gambar 4.5 Gambar Hasil Labeling VADER .....	64
Gambar 4.6 Hasil Hapus Netral VADER Lexicon .....	65
Gambar 4.7 Hasil Hapus Netral InSet Lexicon.....	65
Gambar 4.8 Hasil TF-IDF .....	67

## DAFTAR KODE PROGRAM

Kode Program 3.1 Kode Program Daftar Stopwords Bahasa Indonesia .....	24
Kode Program 4.1 Penggabungan Data.....	42
Kode Program 4.2 Explorasi Data .....	43
Kode Program 4.3 Penghapusan Data yang Tidak Digunakan.....	44
Kode Program 4.4 Data Cleaning.....	45
Kode Program 4.5 Tokenizing Data.....	47
Kode Program 4.6 Normalisasi Data .....	48
Kode Program 4.7 Translate Bahasa .....	50
Kode Program 4.8 Filtering Data Bahasa Indonesia .....	52
Kode Program 4.9 Filtering Data Bahasa Inggris .....	54
Kode Program 4.10 Stemming Bahasa Indonesia.....	56
Kode Program 4.11 Stemming Bahasa Inggris .....	58
Kode Program 4.12 Membaca Kamus Lexicon .....	60
Kode Program 4.13 Fungsi Menghitung Bobot Nilai InSet Lexicon.....	60
Kode Program 4.14 Menerapkan Fungsi dan Menyimpan Hasil InSet Lexicon.....	61
Kode Program 4.15 Labeling VADER Lexicon .....	63
Kode Program 4.16 Hapus Data Sentimen Netral.....	65
Kode Program 4.17 Split Data Uji dan Latih.....	66
Kode Program 4.18 TF-IDF .....	67
Kode Program 4.19 Support Vector Machine.....	68
Kode Program 4.20 Confusion Matrix.....	69