



**ANALISIS SENTIMEN MASYARAKAT TERHADAP FILM
“DIRTY VOTED” DI YOUTUBE MENGGUNAKAN MODEL
SVM (SUPPORT VECTOR MACHINE) DAN LSTM (LONG
SHORT-TERM MEMORY)**



LAPORAN TUGAS AKHIR

**UNIVERSITAS
MERCU BUANA
MUHAMAD RIZKY ARDIANSYAH
41520110050**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2024**



**ANALISIS SENTIMEN MASYARAKAT TERHADAP FILM
“DIRTY VOTED” DI YOUTUBE MENGGUNAKAN MODEL
SVM (SUPPORT VECTOR MACHINE) DAN LSTM (LONG
SHORT-TERM MEMORY)**



LAPORAN TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**UNIVERSITAS
MERCU BUANA
MUHAMAD RIZKY ARDIANSYAH
41520110050**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2024**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan dibawah ini:

Nama : Muhamad Rizky Ardiansyah
NIM : 41520110050
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Analisis Sentimen Masyarakat Terhadap Film “Dirty Vote” di YouTube Menggunakan Model SVM (Support Vector Machine) dan LSTM (Long Short-Term Memory)

Menyatakan bahwa Laporan Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta 20 Juli 2024



Muhamad Rizky Ardiansyah

HALAMAN PENGESAHAN

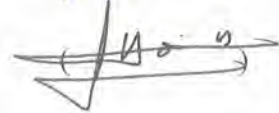
Laporan Skripsi ini diajukan oleh:

Nama : Muhamad Rizky Ardiansyah
NIM : 41520110050
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Analisis Sentimen Masyarakat Terhadap Film 'Dirty Vote' di YouTube Menggunakan Model SVM (Support Vector Machine) dan LSTM (Long Short-Term Memory)

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Suhendra S.Kom., M.Kom
NIDN : 0308019002
Ketua Penguji : Dr. Hadi Santoso, S.Kom., M.Kom
NIDN : 0225067701
Penguji 1 : Roy Mubarak, S.T., M.Kom
NIDN : 310027402
Penguji 2 : Lukman Hakim, S.T., M.Kom
NIDN : 327107701



Jakarta, 20 Juli 2024

Mengetahui,

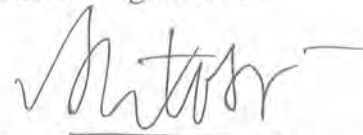
Dekan



Dr. Bambang Jokonowo, S.Si., M.TI

NIDN : 0320037002

Ketua Progam Studi



Dr. Hadi Santoso, S.Kom., M.Kom

NIDN : 0225067701


KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah Swt, Tuhan Yang Maha Pengasih dan Maha Penyayang, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana Jakarta.
2. Bapak Dr. Bambang Jokonowo selaku Dekan Fakultas Ilmu Komputer Universitas Mercu Buana Jakarta.
3. Bapak Hadi Dr. Hadi Santoso, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana Jakarta.
4. Bapak Suhendra S.Kom., M.Kom., selaku Dosen Pembimbing Tugas Akhir yang telah menyediakan waktu, tenaga, dan pikiran untuk mengerahkan dalam penyusunan skripsi ini.
5. Bapak Roy Mubarak, S.T., M.Kom., selaku Dosen Penguji 1 Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Bapak Lukman Hakim, S.T., M.Kom., selaku Dosen Penguji 2 Tugas Akhir atas koreksi dan arahan serta masukannya.
7. Keluarga tercinta khususnya mama saya yang telah memberi doa, semangat, motivasi, dan materialnya kepada saya.

Akhir kata, saya berharap Allah SWT, Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 20 Juli 2024


Muhamad Rizky Ardiansyah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Muhamad Rizky Ardiansyah
NIM : 41520110050
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Analisis Sentimen Masyarakat Terhadap Film “Dirty Vote” di YouTube Menggunakan Model SVM (Support Vector Machine) dan LSTM (Long Short-Term Memory)

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 20 Juli 2024

Yang menyatakan,



Muhamad Rizky Ardiansyah

ABSTRAK

Nama	:	Muhamad Rizky Ardiansyah
NIM	:	41520110050
Program Studi	:	Teknik Informatika
Judul Laporan Skripsi	:	Analisis Sentimen Masyarakat Terhadap Film “Dirty Vote” di YouTube Menggunakan Model SVM (Support Vector Machine) dan LSTM (Long Short-Term Memory)
Pembimbing	:	Suhendra S. Kom., M. Kom.

Film dokumenter "Dirty Vote" yang tersedia di kanal YouTube PSHK Indonesia, telah menjadi salah satu video yang memperoleh perhatian dan reaksi yang beragam di *platform* YouTube. Analisis sentimen dilakukan untuk mengetahui sentimen masyarakat dalam memberikan komentar terhadap film dokumenter ini. Dalam menganalisis sentimen masyarakat, digunakan *InSet Lexicon* sebagai kamus untuk proses pelabelan otomatis bahasa Indonesia, *SVM* dan *LSTM* sebagai algoritma klasifikasi sentimen dan *SMOTE* untuk mengatasi *overfitting*. Penelitian ini melakukan perbandingan hasil pelabelan dan model di tiap tahap pra-pemrosesan teks. Hasil penelitian menunjukkan penggunaan *InSet Lexicon* untuk pelabelan pada tahap terakhir di pra-pemrosesan *stemming* menghasilkan distribusi sentimen yang paling timpang, sentimen positif sebesar 20,26% dan negatif mencapai 79,74%. Selain itu, hasil evaluasi dari kinerja *SVM* mencatat akurasi tertinggi sebesar 97,58% menggunakan data di proses *stemming* dengan *test size* 10% dan *random state* 1, sementara *LSTM* mencatat akurasi tertinggi sebesar 96,3% dan nilai *loss function* yaitu 0,098 dengan *batch size* 16 dengan menggunakan data ditahap *stemming*. Penggunaan metode *SMOTE* untuk mengatasi *overfitting* menunjukkan hasil yang baik untuk model *SVM* sedangkan hasil yang negatif untuk *LSTM*.

Kata kunci: Analisis sentimen, *SVM*, *LSTM*, *SMOTE*, *Lexicon*

ABSTRACT

Nama : Muhamad Rizky Ardiansyah
NIM : 41520110050
Program Studi : Teknik Informatika
Judul Laporan Skripsi : *Analysis of Public Sentiment towards the Movie "Dirty Vote" on YouTube Using SVM Model (Support Vector Machine) and LSTM (Long Short-Term Memory) models.*
Pembimbing : Suhendra S. Kom., M. Kom.

The documentary "Dirty Vote" on PSHK Indonesia's YouTube channel has become one of the videos that has gained attention and mixed reactions on the YouTube platform. Sentiment analysis was conducted to find out people's sentiments in commenting on this documentary. In analyzing public sentiment, InSet Lexicon is used as a dictionary for the automatic labeling process of Indonesian language, SVM and LSTM as sentiment classification algorithms and SMOTE to overcome overfitting. This research compares the labeling results and models at each stage of text pre-processing. The results showed that the use of InSet Lexicon for labeling at the last stage of stemming pre-processing resulted in the most unequal sentiment distribution, with positive sentiment accounting for 20,26% and negative sentiment accounting for 79,74%. In addition, the evaluation results of SVM performance recorded the highest accuracy of 97,58% using data in the stemming process with test size 10% and random state 1, while LSTM recorded the highest accuracy of 96,3% and loss function value of 0,098 with batch size 16 using data in the stemming stage. The use of the SMOTE method to overcome overfitting showed good results for the SVM model while negative results for the LSTM.

Keywords: Sentiment Analysis, SVM, LSTM, SMOTE, Lexicon

DAFTAR ISI

HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	v
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiv
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 Penelitian	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu.....	7
2.2 Teori Pendukung	17
2.2.1 <i>Data mining</i>	17
2.2.2 <i>Text Mining</i>	17
2.2.3 Analisis Sentimen	18
2.2.4 <i>Natural Language Processing</i>	18
2.2.5 <i>Word Embedding</i>	19
2.2.6 <i>Machine learning</i>	20
2.2.7 <i>Deep learning</i>	20
2.2.8 <i>SVM (Support Vector Machine)</i>	21
2.2.9 <i>LSTM (Long Short-Term Memory)</i>	22
2.2.10 Pelabelan berbasis leksikon	23
2.2.11 <i>TF-IDF (Term Frequency-Inverse Document Frequency)</i> <i>Vectorizer</i>	23

2.2.12	<i>SMOTE (Synthetic Minority Over-sampling Technique)</i>	23
2.2.13	Python	24
BAB III METODE PENELITIAN		25
3.1	Jenis Penelitian	25
3.2	Tahapan Penelitian	25
3.2.1	Menentukan Topik	27
3.2.2	Studi literatur	28
3.2.3	Identifikasi Masalah	28
3.2.4	Pengumpulan data	29
3.2.5	Pra-pemrosesan data	29
3.2.6	Pelabelan data	32
3.2.7	Pembagian data latih dan data uji	34
3.2.8	Ketidakeimbangan kelas dengan <i>SMOTE</i>	35
3.2.9	Ekstraksi fitur dengan <i>TF-IDF</i>	35
3.2.10	Tokenisasi dan <i>Padding</i>	36
3.2.11	<i>Trainable Word Embedding</i>	36
3.2.12	Pelatihan model dengan <i>SVM</i>	37
3.2.13	Membangun Arsitektur <i>LSTM</i>	37
3.2.14	Pelatihan model dengan <i>LSTM</i>	38
3.2.15	Uji dan Evaluasi Model	39
3.2.16	Visualisasi	39
BAB IV HASIL DAN PEMBAHASAN		41
4.1	<i>Experiment Setup</i>	41
4.2	<i>Library</i>	41
4.3	<i>Dataset</i>	43
4.4	Pra-pemrosesan data	46
4.5	Pelabelan data	54
4.6	Pemodelan <i>SVM</i>	58
4.7	Analisis Menggunakan Parameter Terbaik model <i>SVM</i>	66
4.8	Pemodelan <i>LSTM</i>	71
4.9	Analisis Menggunakan Parameter Terbaik Model <i>LSTM</i>	80
4.10	Analisis menggunakan <i>Wordcloud</i>	84
BAB V KESIMPULAN DAN SARAN		89
5.1	Kesimpulan	89

5.2	Saran	93
	DAFTAR PUSTAKA.....	95
	LAMPIRAN	99



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	7
Tabel 3.1 Sampel <i>Stopword</i> Indonesia dari <i>NLTK</i>	31
Tabel 3.2 Sampel <i>InSet Lexicon</i> (Positif).....	33
Tabel 3.3 Sampel <i>InSet Lexicon</i> (Negatif)	33
Tabel 4.1 Spesifikasi Laptop untuk Penelitian.....	41
Tabel 4.2 Library yang Digunakan.....	41
Tabel 4.3 Hasil Sampel Pengumpulan Data.....	45
Tabel 4.4 Hasil Sampel <i>Case Folding</i>	49
Tabel 4.5 Hasil Sampel <i>Text Cleansing</i>	50
Tabel 4.6 Hasil Sampel <i>Text Normalization</i>	51
Tabel 4.7 Sampel Hasil <i>Tokenization</i>	52
Tabel 4.8 Hasil Sampel <i>Stopword Removal</i>	53
Tabel 4.9 Sampel Hasil Stemming	53
Tabel 4.10 Sampel Skor Polaritas dan Labelnya Tiap Proses	55
Tabel 4.11 Sampel Skor Polaritas dan Labelnya Tiap Proses (Lanjutan)	55
Tabel 4.12 Sampel Label yang Dikonversi ke Bentuk <i>Integer</i>	55
Tabel 4.13 Hasil Jumlah Pelabelan Tiap Proses.....	56
Tabel 4.14 Hasil Pelabelan Tiap Proses	57
Tabel 4.15 Matriks Evaluasi Tanpa <i>SMOTE</i>	64
Tabel 4.16 Matriks Evaluasi Dengan <i>SMOTE</i>	64
Tabel 4.17 Rangkuman Hasil Evaluasi Skor dan <i>MSE</i> dengan model <i>SVM</i>	68
Tabel 4.18 Laporan Klasifikasi model <i>SVM</i> tanpa <i>SMOTE</i>	70
Tabel 4.19 Laporan Klasifikasi model <i>SVM</i> dengan <i>SMOTE</i>	70
Tabel 4.20 Laporan Klasifikasi model <i>LSTM</i>	83

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Tahapan Penelitian	26
Gambar 3.2 Diagram Alir Tahapan Pengolahan Data	26
Gambar 4.1 <i>Define</i> Objek <i>Youtube API</i>	44
Gambar 4.2 Fungsi Pengumpulan Data dengan <i>YouTube API</i>	44
Gambar 4.3 Memasukkan Hasil Pengumpulan Data ke <i>Dataframe</i>	45
Gambar 4.4 Fungsi untuk <i>Export Dataframe</i> dengan Format <i>CSV</i>	45
Gambar 4.5 Memanggil Fungsi <i>Export CSV</i>	45
Gambar 4.6 Fungsi Menampilkan jumlah data yang <i>null</i> di setiap kolom	47
Gambar 4.7 Memanggil fungsi jumlah data yang <i>null</i>	47
Gambar 4.8 Menghapus data yang memiliki data <i>null</i>	47
Gambar 4.9 Melihat jumlah data yang duplikat dan menghapusnya.....	47
Gambar 4.10 Fungsi filter data berdasarkan waktu komentar	48
Gambar 4.11 Pemanggilan Fungsi Filter Data Berdasarkan Waktu.....	48
Gambar 4.12 Fungsi Konversi <i>Timestamp</i> ke Format Tanggal, Bulan dan Tahun	48
Gambar 4.13 Fungsi <i>Case Folding</i>	49
Gambar 4.14 Fungsi <i>Text Cleansing</i>	49
Gambar 4.15 Fungsi <i>Text Cleansing</i> Lanjutan.....	50
Gambar 4.16 Membentuk Kamus Kata <i>Slang</i>	51
Gambar 4.17 Fungsi <i>Normalization Text</i>	51
Gambar 4.18 Fungsi <i>Tokenization</i>	52
Gambar 4.19 Fungsi <i>Stopword Removal</i>	52
Gambar 4.20 Fungsi <i>Stemming</i>	53
Gambar 4.21 Fungsi Hapus Data yang Kurang dari 3 Karakter	54
Gambar 4.22 Fungsi Pelabelan Data.....	54
Gambar 4.23 Membuat Kolom untuk Tiap Tahap Proses yang Sudah di Labeli..	55
Gambar 4.24 Mengaplikasikan Fungsi Pelabelan.....	55
Gambar 4.25 Grafik Jumlah Kategori Sentimen Positif dan Negatif di <i>Dataset</i> Setelah Pelabelan	57
Gambar 4.26 <i>Source Code</i> Pemodelan <i>SVM</i>	59
Gambar 4.27 Penerapan fungsi pemodelan dari ke-4 <i>dataset</i>	60
Gambar 4.28 Hasil Lengkap Evaluasi <i>SVM</i> di Data <i>Cleansing</i>	61

Gambar 4.29 Hasil Lengkap Evaluasi <i>SVM</i> di Data <i>Normalization</i>	61
Gambar 4.30 Hasil Lengkap Evaluasi <i>SVM</i> di Data <i>Stopword Removal</i>	61
Gambar 4.31 Hasil Lengkap Evaluasi <i>SVM</i> di Data <i>Stemming</i>	62
Gambar 4.32 Grafik Tren Akurasi Model <i>SVM</i> Berdasarkan <i>Test Size</i> dan <i>Random State Data Cleansing</i>	62
Gambar 4.33 Grafik Tren Akurasi Model <i>SVM</i> Berdasarkan <i>Test Size</i> dan <i>Random State Data Normalization</i>	62
Gambar 4.34 Grafik Tren Akurasi Model <i>SVM</i> Berdasarkan <i>Test Size</i> dan <i>Random State Data Stopword Removal</i>	63
Gambar 4.35 Grafik Tren Akurasi Model <i>SVM</i> Berdasarkan <i>Test Size</i> dan <i>Random State Data Stemming</i>	63
Gambar 4.36 Grafik Histogram Data Latih Tanpa <i>SMOTE</i>	66
Gambar 4.37 Grafik Histogram Data Latih Dengan <i>SMOTE</i>	67
Gambar 4.38 <i>Confusion Matrix</i> model <i>SVM</i>	67
Gambar 4.39 <i>Source code</i> pemodelan <i>LSTM</i>	72
Gambar 4.40 Arsitektur Model <i>LSTM</i> (<i>Batch Size 128 dan 64</i>)	72
Gambar 4.41 Arsitektur Model <i>LSTM</i> (<i>Batch Size 32 dan 16</i>).....	73
Gambar 4.42 <i>Source Code</i> Evaluasi Pemodelan <i>LSTM</i>	75
Gambar 4.43 <i>Source Code</i> Evaluasi Pemodelan <i>LSTM</i> (<i>Lanjutan ke-1</i>).....	75
Gambar 4.44 <i>Source Code</i> Evaluasi Pemodelan <i>LSTM</i> (<i>Lanjutan ke-2</i>).....	75
Gambar 4.45 Hasil Lengkap Evaluasi <i>LSTM</i> tanpa <i>SMOTE</i>	76
Gambar 4.46 Hasil Lengkap Evaluasi <i>LSTM</i> Dengan <i>SMOTE</i>	78
Gambar 4.47 Hasil Evaluasi <i>LSTM Accuracy</i> , <i>Loss</i> , dan <i>Durasi</i>	79
Gambar 4.48 Grafik <i>History</i> Akurasi dan <i>Loss Function</i> Model <i>LSTM</i>	80
Gambar 4.49 <i>Confusion Matrix</i> Model <i>LSTM</i>	82
Gambar 4.50 <i>WordCloud</i> Semua Sentimen.....	84
Gambar 4.51 Frekuensi kata tertinggi (15 kata)	85
Gambar 4.52 <i>Wordcloud</i> Sentimen Positif.....	86
Gambar 4.53 <i>Wordcloud</i> sentimen negatif.....	87

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi.....	99
Lampiran 2 Halaman Persetujuan	100
Lampiran 3 Curriculum Vitae.....	101
Lampiran 4 Surat Pernyataan HAKI.....	102
Lampiran 5 Sertifikat BNSP	104
Lampiran 6 Hasil Cek Turnitin	105
Lampiran 7 Form Revisi Penguji 1	106
Lampiran 8 Form Revisi Penguji 2	107

