

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
ANALISIS SENTIMEN DAN LATENT DIRICHLET ALLOCATION
UNTUK PEMODELAN TOPIK TERKAIT GENERATIVE AI**

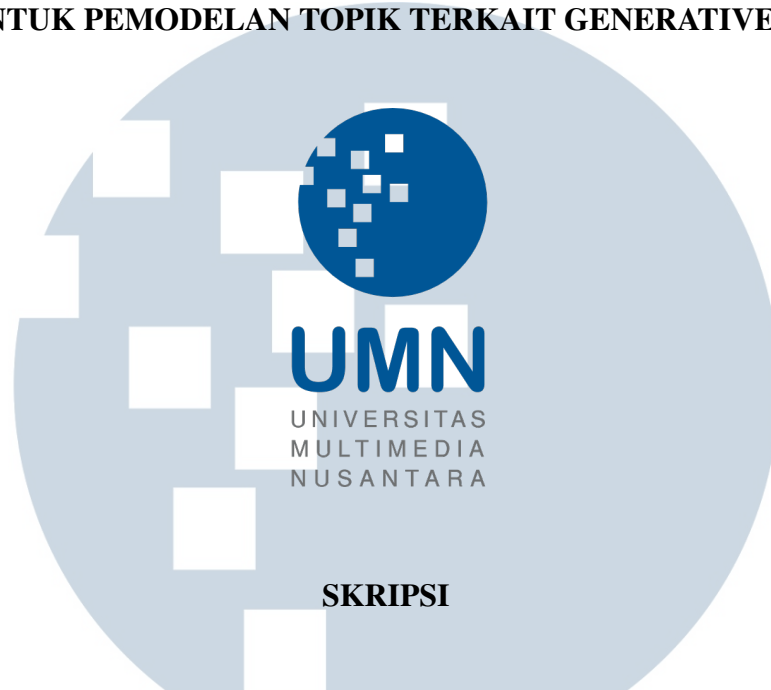


SKRIPSI

**Andrew Wijaya
00000042946**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
ANALISIS SENTIMEN DAN LATENT DIRICHLET ALLOCATION
UNTUK PEMODELAN TOPIK TERKAIT GENERATIVE AI**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Andrew Wijaya

00000042946

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Andrew Wijaya

NIM : 00000042946

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Skripsi saya yang berjudul:
Implementasi Algoritma Support Vector Machine untuk Analisis Sentimen dan Latent Dirichlet Allocation untuk Pemodelan Topik Terkait Generative AI

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 20 Mei 2024



(Andrew Wijaya)

UNIVERSI
MULTIMEDIA
NUSANTARA

Skripsi dengan judul

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK
ANALISIS SENTIMEN DAN LATENT DIRICHLET ALLOCATION
UNTUK PEMODELAN TOPIK TERKAIT GENERATIVE AI**

oleh

Nama : Andrew Wijaya
NIM : 00000042946
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Rabu, 29 Mei 2024

Pukul 10.00 s/s 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Adhi Kusnadi, S.T, M.Si.)

NIDN: 303037304

Penguji



(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

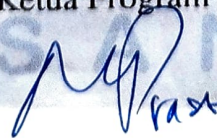
Pembimbing



(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

Pjs. Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc.)

NIDN: 0419128203

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Andrew Wijaya
NIM : 00000042946
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Skripsi

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 20 Mei 2024

Yang menyatakan



Andrew Wijaya

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"If you want something done right, do it yourself."

Charles-Guillaume Étienne

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Algoritma Support Vector Machine untuk Analisis Sentimen dan Latent Dirichlet Allocation untuk Pemodelan Topik Terkait Generative AI dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Pjs. Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dennis Gunawan, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini. (kalau ada).

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 20 Mei 2024



Andrew Wijaya

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN DAN LATENT DIRICHLET ALLOCATION UNTUK PEMODELAN TOPIK TERKAIT GENERATIVE AI

Andrew Wijaya

ABSTRAK

Generative AI merupakan sebuah teknologi *machine learning* yang dapat membuat suatu konten baru berupa teks, gambar, musik, atau pun video dengan menganalisis data konten lain yang telah ada sebelumnya. Kemunculan *generative AI* mengundang pro dan kontra terkait dengan kepemilikan, tanggung jawab, serta kebijakan yang mengatur tentang teknologi yang satu ini. Salah satu bahan pertimbangan dalam pembuatan kebijakan adalah persepsi masyarakat. Analisis sentimen merupakan metode yang dapat digunakan untuk mendapatkan persepsi serta topik yang dibahas oleh masyarakat. Analisis sentimen dapat dikombinasikan dengan metode pemodelan topik untuk mendapatkan pemahaman lebih terkait dengan topik yang diteliti. Penelitian ini berhasil menerapkan algoritma SVM dan LDA untuk analisis sentimen dan pemodelan topik melalui tahapan pengumpulan data, *pre-processing*, *data splitting*, pembuatan model SVM, evaluasi model, pembuatan model LDA, serta visualisasi model LDA. Hasilnya diperoleh model SVM dengan *macro average precision* = 0.8886, *recall* = 0.8985, *f1-score* = 0.8930, dan *accuracy* = 0.8959 yang dibuat menggunakan *kernel linear* dengan nilai $c = 10$. Sementara itu, model LDA yang dibuat berhasil menampilkan 30 *keyword* untuk masing-masing kluster topik pada komentar berlabel sentimen '*Positive*', '*Negative*', dan '*Neutral*'.

Kata kunci: Analisis sentimen, *Generative AI*, *Latent Dirichlet Allocation*, Pemodelan topik, *Support Vector Machine*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

***Implementation of Support Vector Machine Algorithms for Sentiment Analysis
and Latent Dirichlet Allocation for Modeling Generative AI Related Topics***

Andrew Wijaya

ABSTRACT

Generative AI is a machine learning technology that can create new content in the form of text, images, music or videos by analyzing other pre-existing content data. The emergence of generative AI invites pros and cons related to ownership, responsibility and policies governing this technology. One of the considerations in policy making is public perception. Sentiment analysis is a method that can be used to obtain perceptions and topics discussed by the public. Sentiment analysis can be combined with topic modeling methods to gain a deeper understanding of the topic under study. This research succeeded in applying the SVM and LDA algorithms for sentiment analysis and topic modeling through the stages of data collection, pre-processing, data splitting, SVM model creation, model evaluation, LDA model creation, and LDA model visualization. The results obtained were an SVM model with macro average precision = 0.8886, recall = 0.8985, f1-score = 0.8930, and accuracy = 0.8959 which was created using a linear kernel with a value of $c = 10$. Meanwhile, the LDA model created successfully displayed 30 keywords for each. Each topic cluster in the comment is labeled 'Positive', 'Negative', and 'Neutral' sentiment.

Keywords: *Generative AI, Latent Dirichlet Allocation, Sentiment analysis, Support Vector Machine, Topic Modeling*



DAFTAR ISI

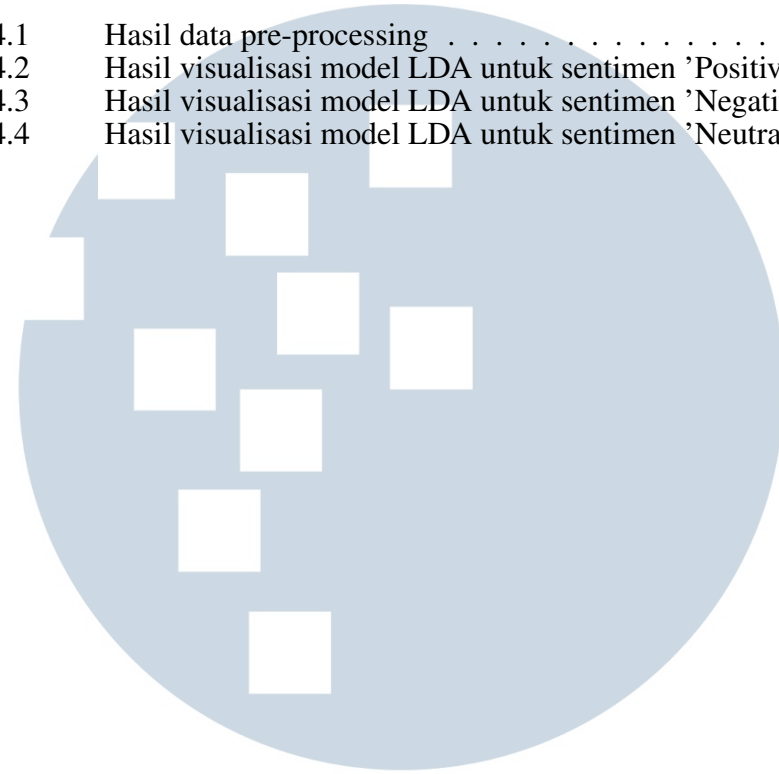
HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
Listings	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Permasalahan	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	6
2.1 Generative AI	6
2.2 Analisis Sentimen	6
2.3 Textblob	6
2.4 Support Vector Machine	7
2.5 SMOTE	9
2.6 Confusion Matrix	10
2.7 Pemodelan Topik	11
2.8 Latent Dirichlet Allocation	12
2.9 PyLDAvis	13
2.10 Text Pre-processing	14
2.10.1 Data Cleaning	14
2.10.2 Tokenization	14
2.10.3 Removing Stop Word	15
2.10.4 Lemmatization	15
2.10.5 Labeling	15
2.11 TF-IDF	15
BAB 3 METODOLOGI PENELITIAN	17
3.1 Alur Penelitian	17
3.2 Studi Literatur	18
3.3 Pengumpulan Data	19
3.4 Text Pre-processing	20
3.4.1 Data Cleaning	21
3.4.2 Tokenization	22
3.4.3 Labeling (Textblob)	24
3.5 TF-IDF	25
3.6 Data Splitting	25
3.7 Analisis Sentimen dengan SVM	26

3.8	Pengujian Model SVM	27
3.9	Pemodelan Topik dengan LDA dan PyLDAvis	28
3.10	Dokumentasi	29
BAB 4	HASIL DAN DISKUSI	30
4.1	Source Code Hasil Implementasi	30
4.1.1	Pengumpulan Data	31
4.1.2	Data Pre-processing	34
4.1.3	Data Splitting	43
4.1.4	Analisis Sentimen Menggunakan SVM	44
4.1.5	Pengujian Model SVM	47
4.1.6	Pemodelan Topik dengan LDA dan PyLDAvis	48
4.2	Skenario Uji Coba	51
4.3	Hasil dan Diskusi Uji Coba	51
BAB 5	SIMPULAN DAN SARAN	73
5.1	Simpulan	73
5.2	Saran	73
DAFTAR PUSTAKA		75



DAFTAR TABEL

Tabel 4.1	Hasil data pre-processing	43
Tabel 4.2	Hasil visualisasi model LDA untuk sentimen 'Positive' . .	64
Tabel 4.3	Hasil visualisasi model LDA untuk sentimen 'Negative' . .	67
Tabel 4.4	Hasil visualisasi model LDA untuk sentimen 'Neutral' . .	69



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2.1	Cara kerja SVM	8
Gambar 2.2	Confusion matrix yang digunakan	11
Gambar 2.3	Tampilan PyLDAvis	13
Gambar 3.1	Teknis alur penelitian	17
Gambar 3.2	Flowchart alur penelitian	18
Gambar 3.3	Flowchart pengumpulan data	19
Gambar 3.4	Data frame dari CSV yang digunakan	20
Gambar 3.5	Flowchart text pre-processing	20
Gambar 3.6	Flowchart data cleaning	21
Gambar 3.7	Flowchart tokenization	22
Gambar 3.8	Flowchart removing stop word	23
Gambar 3.9	Flowchart lemmatization	24
Gambar 3.10	Flowchart labeling (textblob)	24
Gambar 3.11	Flowchart proses TF-IDF	25
Gambar 3.12	Flowchart analisis sentimen dengan SVM	26
Gambar 3.13	Flowchart pembuatan model SVM	26
Gambar 3.14	Flowchart pemodelan topik dengan LDA dan PyLDAvis	28
Gambar 3.15	Flowchart pembuatan model LDA	28
Gambar 4.1	Developer key untuk API Youtube	31
Gambar 4.2	Video id pada video Youtube	32
Gambar 4.3	File CSV dari pengumpulan data	33
Gambar 4.4	Dataframe komentar sebelum di pre-process	34
Gambar 4.5	Dataframe komentar setelah menjalankan remove HTML	35
Gambar 4.6	Dataframe komentar setelah menjalankan remove symbol dan diubah ke lowercase	36
Gambar 4.7	Dataframe komentar setelah menjalankan detect language	37
Gambar 4.8	List mapping dari singkatan yang umum digunakan	38
Gambar 4.9	Dataframe komentar setelah menjalankan abbreviation mapping	39
Gambar 4.10	Isi dari stop word Bahasa Inggris yang disediakan library nltk	40
Gambar 4.11	Dataframe komentar setelah menjalankan tokenization dan menghapus data yang kosong (null)	40
Gambar 4.12	Dataframe komentar setelah menjalankan proses labeling menggunakan textblob	42
Gambar 4.13	Line chart proses hyperparameter tuning (gamma Auto)	52
Gambar 4.14	Line chart proses hyperparameter tuning (gamma Scale)	52
Gambar 4.15	Classification report model SVM tanpa menggunakan SMOTE	54
Gambar 4.16	Classification report model SVM menggunakan SMOTE	54
Gambar 4.17	Classification report model SVM menggunakan Undersampling	54
Gambar 4.18	Confusion matrix model SVM terbaik	56
Gambar 4.19	Word cloud semua komentar pada kolom 'new_comment'	58
Gambar 4.20	Word cloud komentar positif pada kolom 'new_comment'	58
Gambar 4.21	Word cloud komentar negatif pada kolom 'new_comment'	59
Gambar 4.22	Word cloud komentar netral pada kolom 'new_comment'	59

Gambar 4.23	Line chart proses pencarian coherence value untuk sentimen 'Positive'	60
Gambar 4.24	Line chart proses pencarian coherence value untuk sentimen 'Negative'	60
Gambar 4.25	Line chart proses pencarian coherence value untuk sentimen 'Neutral'	61
Gambar 4.26	Visualiasi pemodelan topik untuk sentimen positif menggunakan pyLDAvis	61
Gambar 4.27	Topik positif 1	62
Gambar 4.28	Topik positif 2	63
Gambar 4.29	Topik positif 3	63
Gambar 4.30	Visualiasi pemodelan topik untuk sentimen negatif menggunakan pyLDAvis	64
Gambar 4.31	Topik negatif 1	65
Gambar 4.32	Topik negatif 2	66
Gambar 4.33	Topik negatif 3	66
Gambar 4.34	Visualiasi pemodelan topik untuk sentimen neutral menggunakan pyLDAvis	67
Gambar 4.35	Topik netral 1	68
Gambar 4.36	Topik netral 2	69
Gambar 4.37	Venn diagram untuk keyword-keyword hasil pemodelan topik	70
Gambar 4.38	Komentar sentimen positif terkait generative AI pada keyword copyright	71
Gambar 4.39	Komentar sentimen negatif terkait generative AI pada keyword copyright	72



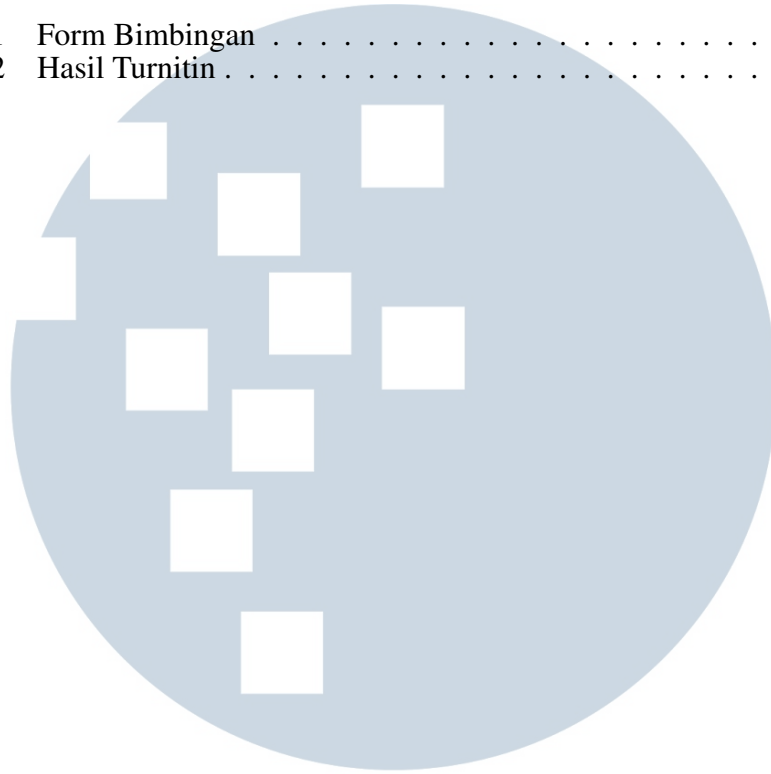
Listings

4.1	Berbagai library yang digunakan dalam source code	30
4.2	Source code pengumpulan data	32
4.3	Source code untuk menghapus tag HTML	34
4.4	Source code untuk menghapus emoji, simbol, dan melakukan lowercase	35
4.5	Source code untuk mendeteksi bahasa dari komentar	36
4.6	Source code untuk melakukan mapping singkatan	38
4.7	Source code untuk proses tokenization	39
4.8	Menghapus data field 'new_comment' yang kosong	40
4.9	Melakukan labeling menggunakan library textblob	41
4.10	Source code untuk data splitting	43
4.11	Source code untuk menghitung jumlah dari masing-masing class	44
4.12	Source code untuk proses analisis sentimen menggunakan SVM	45
4.13	Source code visualisasi proses hyperparameter tuning menggunakan line chart	46
4.14	Source code confusion matrix model SVM	47
4.15	Source code untuk menampilkan Word cloud	47
4.16	Melakukan prediksi terhadap setiap komentar menggunakan model SVM	48
4.17	Source code untuk membuat dictionary dan corpus untuk setiap class sentimen	49
4.18	Source code untuk mencari coherence value terbaik	49
4.19	Source code proses pembuatan model LDA dan visualisasi hasil topic modeling	50



DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan	82
Lampiran 2	Hasil Turnitin	83



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA