

**ANALISIS SENTIMEN ULASAN MOBIL LISTRIK PADA
YOUTUBE MENGGUNAKAN ALGORITMA NB SVM DAN
KNN**



SKRIPSI

Meisha Giovanni Mulin

00000055487

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**ANALISIS SENTIMEN ULASAN MOBIL LISTRIK PADA
YOUTUBE MENGGUNAKAN ALGORITMA NB SVM DAN
KNN**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer

Meisha Geovanni Mulin

00000055487

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Meisha Geovanni Mulin

Nomor Induk Mahasiswa : 00000055487

Program Studi : Sistem Informasi

Skripsi dengan judul:

*“ANALISIS SENTIMEN ULASAN MOBIL LISTRIK PADA YOUTUBE
MENGGUNAKAN ALGORITMA NB SVM DAN KNN”*

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 3 Juni 2025



(Meisha Geovanni Mulin)

HALAMAN PERSETUJUAN

Skripsi dengan judul

ANALISIS SENTIMEN ULASAN MOBIL LISTRIK PADA YOUTUBE MENGUNAKAN ALGORITMA NB SVM DAN KNN

Oleh

Nama : Meisha Geovanni Mulin

NIM : 00000055487

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 3 Juni 2025

Pembimbing



Ahmad Faza, S.Kom., M.T.I
0312019501

Ketua Sistem Informasi



Ririn Ikana Desanti, S.Kom., M.Kom.

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN ULASAN MOBIL LISTRIK PADA YOUTUBE
MENGUNAKAN ALGORITMA NB SVM DAN KNN

Oleh

Nama : Meisha Geovanni Mulin
NIM : 00000055487
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

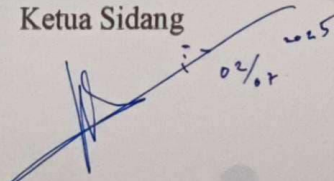
Telah diujikan pada hari Jumat, 20 Juni 2025

Pukul 13.00 s.d 14.00 dan dinyatakan

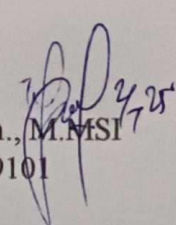
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang


Rudi Sutomo, S.Kom., M.Si., M.Kom
0222057501

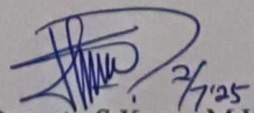
Penguji


Wella, S.Kom., M.MSI
0305119101

Pembimbing

 2/7/2025
Ahmad Faza, S.Kom., M.T.I
0312019501

Ketua Program Studi Sistem Informasi

 2/7/25
Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001

Yang bertanda tangan di bawah ini:

Nama : Meisha Geovanni Mulin

NIM : 00000055487

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Analisis Sentimen Ulasan Mobil Listrik Pada Youtube Menggunakan Algoritma Nb Svm dan Knn

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☐ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☒ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 4 Juni 2025


(Meisha Geovanni Mulin)

* Pilih salah satu

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

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yesus Kristus atas kasih karunia, penyertaan, dan kekuatan-Nya yang melimpah, sehingga penulis dapat menyelesaikan penulisan skripsi dengan tepat waktu. Penulisan skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (Strata 1) pada Program Studi Sistem Informasi. Penulis menyadari bahwa tanpa campur tangan Tuhan serta dukungan dan bantuan dari berbagai pihak, skripsi ini tidak mungkin terwujud. Oleh karena itu, dengan penuh rasa terima kasih dan hormat, penulis sampaikan kepada:

1. Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo ST, M.Sc., selaku Dekan Fakultas Teknik & Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom, selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Ahmad Faza, S.Kom., M.T.I., sebagai Pembimbing pertama yang telah memberikan bimbingan dan arahan atas terselesaikannya tugas akhir ini.
5. Keluarga di rumah dan keluarga rohani saya yang telah memberikan dukungan dalam bentuk doa, material, dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman seperjuangan saya yang telah memberikan dukungan secara penuh sebagai motivasi untuk menyelesaikan tugas akhir ini.

Saya berharap hasil tugas akhir ini dapat memberikan manfaat signifikan, menjadi pijakan bagi penelitian selanjutnya di bidang sentimen mobil listrik.

Tangerang, 4 Juni 2025


Meisha Geovanni Mulin

ANALISIS SENTIMEN KOMENTAR ULASAN MOBIL LISTRIK PADA YOUTUBE DENGAN MENGGUNAKAN ALGORITMA NB SVM DAN KNN

Meisha Geovanni Mulin

ABSTRAK

Perkembangan teknologi telah mendorong transformasi signifikan di berbagai sektor, termasuk industri otomotif, dengan munculnya mobil listrik sebagai solusi transportasi yang lebih efisien dan ramah lingkungan. Meski demikian, tingkat adopsi mobil listrik di Indonesia masih terbatas. Pemilihan “Wuling Binguo EV” dan “BYD M6” didasarkan penjualan mobil listrik terbanyak pada tahun 2024.

Penelitian ini menggunakan metode CRISP-DM dengan melakukan proses scraping untuk mengumpulkan data komentar pengguna dari platform YouTube. Tahapan analisis dimulai dengan preprocessing teks dan pelabelan sentimen dengan TextBlob. Data kemudian direpresentasikan menggunakan metode TF-IDF dan SMOTE. Beberapa algoritma *machine learning* yang digunakan termasuk *Support Vector Machine* (SVM), *Naïve Bayes* (NB), dan *K-Nearest Neighbor* (KNN).

Hasil evaluasi menunjukkan bahwa model SVM dengan *Grid Search* dan SMOTE memberikan performa terbaik dengan akurasi sebesar 93.58%, precision 90%, recall 88%, dan F1-score 89%. Hasil akhir dari analisis sentimen berupa web Streamlit. Sistem ini memberikan kontribusi nyata sebagai alat pendukung bagi regulator dan industri otomotif dalam memantau opini publik secara real-time dan menyusun strategi berbasis bukti untuk pengembangan kendaraan listrik di Indonesia.

Kata kunci: Analisis sentimen, *K-Nearest Neighbors*, Mobil listrik, *Naïve Bayes*, Youtube

SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON ELECTRIC VEHICLE REVIEWS USING NB SVM AND KNN

ALGORITHMS

Meisha Geovanni Mulin

ABSTRACT (English)

Technological advancements have brought significant changes across many sectors, including the automotive industry, with electric vehicles emerging as a more efficient and eco-friendly transportation. However, EV adoption in Indonesia is still limited. The selection of “Wuling Binguo EV” and “BYD M6” for this research is based on the top-selling electric cars in 2024.

This research follows the CRISP-DM methodology and involves scraping user comments from YouTube as the primary data source. The analysis process begins with text preprocessing and sentiment labeling using TextBlob. The data is then transformed using the TF-IDF method and balanced with SMOTE. Several machine learning algorithms were applied, including Support Vector Machine (SVM), Naïve Bayes (NB), and K-Nearest Neighbor (KNN).

Evaluation results show that the SVM model, optimized with Grid Search and SMOTE, achieved the best performance—with an accuracy of 93.58%, precision of 90%, recall of 88%, and an F1-score of 89%. The final outcome of the sentiment analysis is presented through a Streamlit-based web application. This system serves as a practical tool to help policymakers and the automotive industry monitor public opinion in real time and develop evidence-based strategies for advancing electric vehicle adoption in Indonesia.

Keywords: Sentiment Analysis, K-Nearest Neighbors, Electric Cars, Naïve Bayes, Support Vector Machine.

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iiiv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT (English)	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	7
1.4 Tujuan dan Manfaat Penelitian	7
1.4.1 Tujuan Penelitian	7
1.4.2 Manfaat Penelitian	8
1.5 Sistematika Penulisan	9
BAB II LANDASAN TEORI	11
2.1 Penelitian Terdahulu	11
2.2 Teori Penelitian	17
2.2.1 Analisis Sentimen	17
2.2.2 Mobil Listrik	18
2.2.3 Machine Learning	18
2.2.4 Youtube	19
2.2.5 Cross-Industry Standard Process for Data Mining (CRISP-DM)	20
2.2.6 KDD	21
2.2.7 Web Scrapping	22
2.2.8 TF-IDF	23

2.2.9	<i>Text preprocessing</i>	24
2.2.10	SMOTE	25
2.3	Framework dan Algoritma Penelitian	26
2.3.1	<i>Naïve Bayes Classifier</i>	26
2.3.2	<i>Support Vector Machine</i>	27
2.3.3	<i>K-Nearest Neighbours</i>	28
2.4	Tools Penelitian	29
2.4.1	Python	29
2.4.2	Google Colaboratory	29
2.4.3	Microsoft Excel	30
2.4.4	Visual Studio Code	30
BAB III METODOLOGI PENELITIAN		31
3.1.	Objek Penelitian	31
3.2	Metode Penelitian	33
3.2.1	Alur Penelitian	36
3.3	Teknik Pengumpulan Data	43
3.4	Variabel Penelitian	44
3.4.1	Variabel Independen	44
3.4.2	Variabel Dependen	44
3.5	Teknik Analisis Data	45
BAB IV ANALISIS DAN HASIL PENELITIAN		46
4.1	<i>Business Understanding</i>	46
4.2	<i>Data Understanding</i>	47
4.3	<i>Data Preparation</i>	50
4.3.8	<i>Normalization</i>	63
4.3.9	<i>Join Words</i>	65
4.3.10	<i>Labelling Data</i>	67
4.3.11	TF-IDF	69
4.3.12	<i>Label Encoder</i>	70
4.3.13	<i>Data Splitting</i>	71
4.3.14	SMOTE	72
4.4	Data Mining	74
4.4.1	<i>Support Vectore Machine (SVM)</i>	74

4.4.2	Naïve Bayes (NB).....	78
4.4.3	<i>K-Nearest Neighbour (KNN)</i>	80
4.5	<i>Evaluation</i>	80
4.5.1	<i>Support Vector Machine Model Evaluation</i>	82
4.5.2	<i>SVM With Grid Search Model Evaluation</i>	84
4.5.3	<i>SVM With Grid Search and SMOTE Model Evaluation</i>	85
4.5.4	<i>NB With Grid Search Model Evaluation</i>	88
4.5.5	<i>NB With Grid Search and SMOTE Model Evaluation</i>	90
4.5.6	<i>KNN With Grid Search Model Evaluation</i>	91
4.5.7	<i>KNN With Grid Search and SMOTE Model Evaluation</i>	93
4.6	<i>Deployment</i>	95
4.7	<i>Hasil dan Pembahasan</i>	99
BAB V	SIMPULAN DAN SARAN	102
5.1	Simpulan	102
5.2	Saran	103
DAFTAR	PUSTAKA	104
LAMPIRAN		111



DAFTAR TABEL

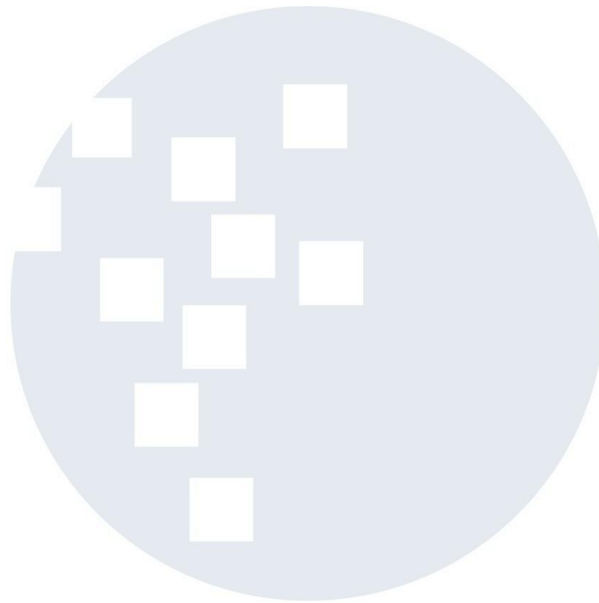
Tabel 1.1 Mobil listrik terlaris sepanjang tahun 2024	3
Tabel 2.1 Penelitian terdahulu.....	11
Tabel 2.2 Rumus kernel	25
Tabel 3.1 Dataset yang diambil pada tanggal 22 Februari 2025.....	31
Tabel 3.2 Contoh komentar pada video ulasan mobil listrik	32
Tabel 3.3 Perbandingan algoritma yang digunakan	33
Tabel 3.4 Perbandingan crisp-dm dan kdd.....	36
Tabel 4.1 Perbandingan sebelum dan setelah <i>lowercase</i>	54
Tabel 4.2 Perbandingan sebelum dan setelah <i>cleaning</i>	56
Tabel 4.3 Perbandingan sebelum dan setelah <i>tokenizing</i>	58
Tabel 4.4 Perbandingan sebelum dan setelah <i>stopword removal</i>	60
Tabel 4.5 Perbandingan sebelum dan setelah <i>stemming</i>	62
Tabel 4.6 Perbandingan sebelum dan setelah <i>normalization</i>	64
Tabel 4.7 Perbandingan sebelum dan setelah <i>join words</i>	66
Tabel 4.8 Komparasi model	99



DAFTAR GAMBAR

Gambar 1.1 Indeks kualitas udara terburuk di dunia [3].....	2
Gambar 2.1 Algoritma SVM [33]	27
Gambar 2.2 Ilustasi cara kerja algoritma KNN [34]	28
Gambar 3.1 Alur penelitian.....	37
Gambar 3.2 Alur proses <i>data preparation</i>	39
Gambar 4.1 Instalasi <i>library data scraping</i>	47
Gambar 4.2 Mengimpor modul dari <i>library</i>	48
Gambar 4.3 Proses pengambilan data komentar dari youtube	49
Gambar 4.4 Hasil pengambilan data komentar dari youtube	50
Gambar 4.5 Informasi dataset	51
Gambar 4.6 Hasil pengecekan duplikasi data	52
Gambar 4.7 Hasil menghapus duplikasi data	53
Gambar 4.8 Proses <i>lowercase</i>	53
Gambar 4.9 Proses <i>cleaning data</i>	55
Gambar 4.10 Proses <i>tokenization</i>	57
Gambar 4.11 Proses <i>stopword removal</i>	59
Gambar 4.12 Proses <i>stemming</i>	62
Gambar 4.13 Proses <i>normalization</i>	63
Gambar 4.14 Menggabungkan teks	65
Gambar 4.15 Menyimpan hasil <i>pre-processing</i>	67
Gambar 4.16 Proses <i>labeling</i>	68
Gambar 4.17 Hasil distribusi proses <i>labeling</i>	69
Gambar 4.18 Proses TF-IDF	70
Gambar 4.19 Proses <i>label encoder</i>	70
Gambar 4.20 Proses <i>split data</i>	71
Gambar 4.21 Proses SMOTE	73
Gambar 4.22 Visualisasi perbandingan data tiap kelas dengan SMOTE.....	72
Gambar 4.23 Implementasi SVM dengan 4 kernel.....	75
Gambar 4.24 Hasil Implementasi SVM dengan 4 kernel	76
Gambar 4.25 GridSearchCV pada SVM.....	77
Gambar 4.26 Implementasi GridSearchCV pada NB	79
Gambar 4.27 Implementasi GridSearchCV5 pada KNN	80
Gambar 4.28 Implementasi GridSearchCV10 pada KNN	81
Gambar 4.29 <i>Confusion matrix pure SVM</i>	82
Gambar 4.30 <i>Classification report pure SVM</i>	83
Gambar 4.31 <i>Confusion matrix grid search SVM</i>	84
Gambar 4.32 <i>Classification report grid search SVM</i>	85
Gambar 4.33 <i>Confusion matrix grid search-smote SVM</i>	86
Gambar 4.34 <i>Classification report grid search-smote SVM</i>	87
Gambar 4.35 <i>Confusion matrix grid search NB</i>	88
Gambar 4.36 <i>Classification report grid search NB</i>	89
Gambar 4.37 <i>Confusion matrix grid search-SMOTE NB</i>	90

Gambar 4.38 <i>Classification report grid search-SMOTE NB</i>	91
Gambar 4.39 <i>Confusion matrix grid search KNN</i>	92
Gambar 4.40 <i>Classification report grid search KNN</i>	93
Gambar 4.41 <i>Confusion matrix grid search-SMOTE KNN</i>	94
Gambar 4.42 <i>Classification report grid search-SMOTE KNN</i>	95
Gambar 4.43 Halaman utama website streamlit	96
Gambar 4.44 Menu teks preprocessing pada website streamlit	97
Gambar 4.45 Menu sentiment analysis pada website streamlit	98
Gambar 4.46 Menu visualisasi pada website streamlit	98



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

Rumus 2. 1 Formula TF	23
Rumus 2. 2 Formula IDF	23
Rumus 2. 3 Formula TF-IDF	24
Rumus 2. 4 Formula Naïve Bayes	26



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	111
Lampiran B Form Konsultasi Bimbingan	118



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA