

**ANALISIS SENTIMEN OPINI PENGGUNA X TERHADAP
BERITA BENSIN PERTAMAX DICAMPUR PERTALITE
DENGAN ALGORITMA NAIVE BAYES**



SKRIPSI

**JOSHUA ARDIN PUTRA PERDANA
00000061685**

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

**ANALISIS SENTIMEN OPINI PENGGUNA X TERHADAP
BERITA BENSIN PERTAMAX DICAMPUR PERTALITE
DENGAN ALGORITMA NAIVE BAYES**



SKRIPSI

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

**JOSHUA ARDIN PUTRA PERDANA
00000061685**

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Joshua Ardin Putra Perdana
Nomor Induk Mahasiswa : 00000061685
Program Studi : Informatika

Skripsi dengan judul:

Analisis sentimen opini pengguna X terhadap berita bensin Pertamina dicampur Peralite dengan Algoritma Naive Bayes

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, 4 Juli 2025



(Joshua Ardin Putra Perdana)

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN OPINI PENGGUNA X TERHADAP BERITA BENSIN PERTAMAX DICAMPUR PERTALITE DENGAN ALGORITMA NAIVE BAYES

oleh

Nama : Joshua Ardin Putra Perdana
NIM : 00000061685
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 18 Juli 2025

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

(Januar Wahjudi, S.Kom., M.Sc.)

NIDN: 0330017201

Penguji

(Anak Agung Ngurah Ananda Kusuma,

B.Eng., M.Eng., Ph.D.)

NIDK: 08984101024

Pembimbing

(Alethea Suryadibrata, S.Kom., M.Eng.)

NIDN: 0322099201

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

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

Yang bertanda tangan di bawah ini:

Nama : Joshua Ardin Putra Perdana
NIM : 00000061685
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Analisis sentimen opini pengguna
X terhadap berita bensin Pertamina
dicampur Pertalite dengan Algoritma
Naive Bayes

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 tiga tahun.

Tangerang, 4 Juli 2025

Yang menyatakan



Joshua Ardin Putra Perdana

HALAMAN PERSEMBAHAN / MOTTO

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis sentimen opini pengguna X terhadap berita bensin Pertamina dicampur Pertalite dengan Metode Naive Bayes 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. Ir. Andrey Andoko, M.Sc., 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 Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Alethea Suryadibrata, S.Kom., M.Eng., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

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

Tangerang, 4 Juli 2025



Joshua Ardin Putra Perdana

**ANALISIS SENTIMEN OPINI PENGGUNA X TERHADAP BERITA
BENSIN PERTAMAX DICAMPUR PERTALITE DENGAN ALGORITMA
NAIVE BAYES**

Joshua Ardin Putra Perdana

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu pencampuran bahan bakar Pertamina dan Pertalite melalui media sosial X (sebelumnya Twitter) dengan menggunakan algoritma Naïve Bayes. Isu ini mencuat sebagai respons atas dugaan praktik pencampuran BBM yang menimbulkan kekhawatiran publik. Data dikumpulkan melalui Tweet Harvest menggunakan kata kunci relevan, kemudian dilakukan tahap pra-proses teks yang meliputi translasi, normalisasi, stemming, dan stopword removal. Sentimen diklasifikasikan menggunakan lexicon-based labeling sebelum data diolah dengan metode TF-IDF dan diklasifikasikan dengan dua varian algoritma: Multinomial Naïve Bayes dan Complement Naïve Bayes. Hasil penelitian menunjukkan bahwa Multinomial Naïve Bayes memperoleh akurasi sebesar 76% dan skor F1 tertinggi pada sentimen positif sebesar 82%, sedangkan Complement Naïve Bayes mencatatkan akurasi 75% dengan skor F1 tertinggi sebesar 79%. Meskipun akurasi keduanya hampir setara, Complement Naïve Bayes menunjukkan kestabilan performa yang lebih baik. Penelitian ini menunjukkan efektivitas metode Naïve Bayes dalam klasifikasi sentimen berbasis media sosial, serta pentingnya tahapan pra-proses dan pelabelan dalam meningkatkan performa model klasifikasi.

Kata kunci: Analisis Sentimen, Naive Bayes, Pertamina, Pertalite, TF-IDF, X(twitter).

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

**ANALYSIS OF SENTIMENT OF X USER OPINIONS ON THE NEWS OF
PERTAMAX GASOLINE MIXED WITH PERTALITE USING THE NAIVE
BAYES ALGORITHM**

Joshua Ardin Putra Perdana

ABSTRACT

This study aims to analyze public sentiment regarding the issue of Pertamina and Pertalite fuel blending on the social media platform X (formerly Twitter) using the Naïve Bayes algorithm. The issue emerged in response to allegations of fuel adulteration, triggering widespread public concern. Data were collected through Tweet Harvest using relevant keywords, followed by a series of text preprocessing stages including translation, normalization, stemming, and stopword removal. Sentiment labeling was performed using a lexicon-based approach. The processed data were then transformed using the TF-IDF method and classified using two algorithm variants: Multinomial Naïve Bayes and Complement Naïve Bayes. The results indicate that Multinomial Naïve Bayes achieved an accuracy of 76% and the highest F1-score of 82% for positive sentiment, while Complement Naïve Bayes recorded an accuracy of 75% and the highest F1-score of 79%. Although the accuracy scores were comparable, Complement Naïve Bayes demonstrated greater stability in performance. This research confirms the effectiveness of the Naïve Bayes method in social media-based sentiment classification and emphasizes the importance of preprocessing and labeling stages in enhancing model performance.

Keywords: Naive Bayes, Sentiment Analysis, Pertamina, Pertalite, TF-IDF, X(twitter).



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	5
2.1 Pengoplosan BBM	5
2.2 X (Twitter)	5
2.3 Analisis Sentimen	5
2.4 Text Mining	6
2.5 Text Preprocessing	6
2.5.1 Translate	6
2.5.2 Case Folding	6
2.5.3 Cleansing	7
2.5.4 Tokenization	7
2.5.5 Stopword Removal	8
2.5.6 Normalization	8
2.5.7 Stemming	8
2.5.8 Labelling	9
2.6 Data Splitting	10
2.7 Term Frequency-Inverse Document Frequency (TF-IDF)	10
2.7.1 Term Frequency (TF)	11
2.7.2 Inverse Document Frequency (IDF)	11
2.8 Algoritma Naïve Bayes	12
2.9 Confusion Matrix	13
2.9.1 Accuracy	14
2.9.2 Precision	14
2.9.3 Recall	15
2.9.4 F1-Score	15
BAB 3 METODOLOGI PENELITIAN	16
3.1 Metode Penelitian	16
3.1.1 Studi Literatur	16
3.1.2 Pengumpulan Data	16

3.1.3	Pengolahan Data	16
3.1.4	Pengujian dan Evaluasi	17
3.1.5	Penulisan Laporan	17
3.2	Perancangan Sistem	17
3.2.1	Gambaran Umum Sistem	18
3.2.2	Data Mining	19
3.2.3	Data Pre-Processing	19
3.2.4	Labeling	20
3.2.5	Data Splitting	21
3.2.6	TF-IDF	21
3.2.7	Penerapan Model Naive Bayes	21
3.2.8	Testing dan Evaluasi	22
3.3	Dokumentasi	22
BAB 4	HASIL DAN DISKUSI	23
4.1	Spesifikasi Sistem	23
4.2	Implementasi Sistem	23
4.2.1	Crawling Data	23
4.2.2	Data Preprocessing	24
4.2.3	Labeling	33
4.2.4	Data Splitting	35
4.2.5	TF-IDF	35
4.2.6	Multinomial Naive Bayes	36
4.2.7	Complement Naive Bayes	36
4.3	Uji Coba dan Evaluasi	37
4.3.1	Uji Coba	37
4.3.2	Evaluasi	39
BAB 5	SIMPULAN DAN SARAN	41
5.1	Simpulan	41
5.2	Saran	41
	DAFTAR PUSTAKA	42



DAFTAR TABEL

Tabel 2.1	Contoh Case Folding	7
Tabel 2.2	Contoh Cleansing	7
Tabel 2.3	Contoh Tokenization	7
Tabel 2.4	Contoh Stopword Removal	8
Tabel 2.5	Contoh Normalization	8
Tabel 2.6	Contoh Stemming	9
Tabel 2.7	Contoh Skor Positif dan Negatif dalam Lexicon Bahasa Indonesia	10
Tabel 2.8	Confusion Matrix	13
Tabel 4.1	Perbandingan sebelum dan sesudah cleansing teks	26
Tabel 4.2	Perbandingan sebelum dan sesudah Tokenizing teks	28
Tabel 4.3	Perbandingan sebelum dan sesudah StopWords	30
Tabel 4.4	Perbandingan sebelum dan sesudah Normalization	31
Tabel 4.5	Perbandingan sebelum dan sesudah Stemming	32
Tabel 4.6	Perbandingan Performa Multinomial dan Complement Naive Bayes	39



DAFTAR GAMBAR

Gambar 3.1	Gamabran Umum	18
Gambar 3.2	Flowchart Data Preprocessing	20
Gambar 4.1	Potongan kode crawl data	23
Gambar 4.2	Potongan kode translate	24
Gambar 4.3	Potongan kode case folding	25
Gambar 4.4	Potongan kode cleansing data	26
Gambar 4.5	Potongan kode cleansing data	27
Gambar 4.6	Potongan kode tokenizing	28
Gambar 4.7	Potongan kode remove stop words	29
Gambar 4.8	Potongan kode normalization	31
Gambar 4.9	Potongan kode stemming	32
Gambar 4.10	Potongan kode labeling	34
Gambar 4.11	Hasil dari labeling	34
Gambar 4.12	Potongan kode data splitting	35
Gambar 4.13	Potongan kode TF-IDF	35
Gambar 4.14	Potongan kode Multinomial Naive Bayes	36
Gambar 4.15	Potongan kode Complement Naive Bayes	37
Gambar 4.16	Hasil Pengujian Multinomial Naive Bayes	38
Gambar 4.17	Hasil Pengujian Complement Naive Bayes	39



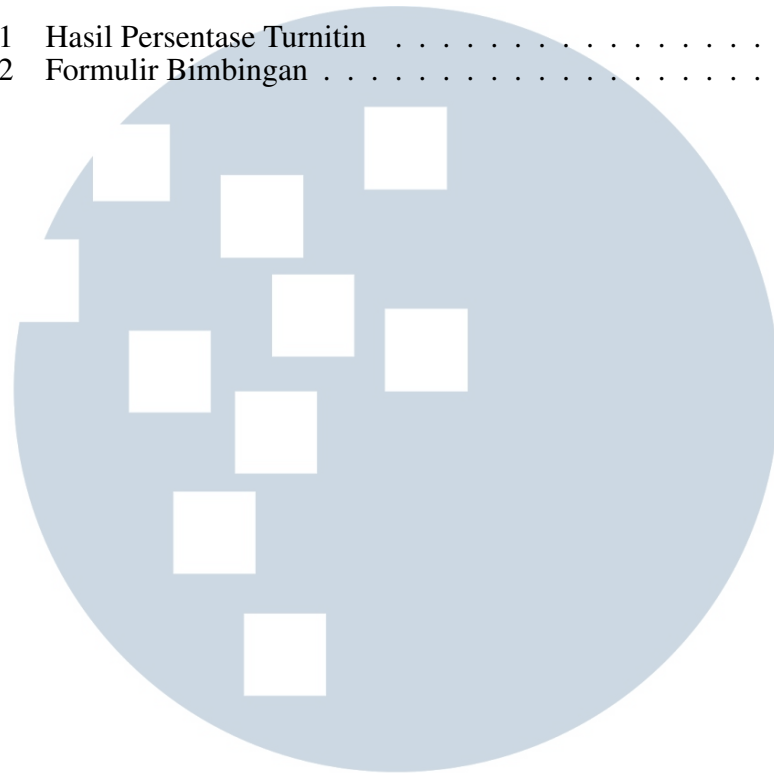
DAFTAR RUMUS

Rumus 2.1	<i>TF-IDF (Term Frequency - Inverse Document Frequency)</i> . . .	11
Rumus 2.2	<i>Term Frequency (TF)</i>	11
Rumus 2.3	<i>Inverse Document Frequency (IDF)</i>	11
Rumus 2.4	<i>Teorema Bayes</i>	12
Rumus 2.5	<i>Confusion Matrix Accuracy</i>	14
Rumus 2.6	<i>Confusion Matrix Precision</i>	14
Rumus 2.7	<i>Confusion Matrix Recall</i>	15
Rumus 2.8	<i>Confusion Matrix F1-Score</i>	15



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	45
Lampiran 2	Formulir Bimbingan	46



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA