

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
NAÏVE BAYES TERHADAP PENJUALAN PRODUK D'ALBA**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Nafisah Cendekia Ramadhani

00000059869

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

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
NAÏVE BAYES TERHADAP PENJUALAN PRODUK D'ALBA**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Nafisah Cendekia Ramadhani

00000059869

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

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Nafisah Cendekia Ramadhani

Nomor Induk Mahasiswa : 00000059869

Program Studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS SENTIMEN KOMENTAR INSTAGRAM MENGGUNAKAN
SUPPORT VECTOR MACHINE DAN NAÏVE BAYES TERHADAP
PENJUALAN PRODUK D'ALBA

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, 16 Mei 2025



Nafisah Cendekia Ramadhani

HALAMAN PERSETUJUAN

Skripsi dengan judul

Analisis Sentimen Komentar Instagram Menggunakan Support Vector Machine
dan Naïve Bayes terhadap Penjualan Produk D'Alba

Oleh

Nama : Nafisah Cendekia Ramadhani

NIM : 00000059869

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara
Tangerang, 16 Mei 2025

Pembimbing

Prof. Friska Natalia, Sk.Kom., M.T.
0306128307

Ketua Program Studi Sistem Informasi

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

HALAMAN PENGESAHAN

Skripsi dengan judul

Analisis Sentimen Komentar Instagram Menggunakan Support Vector Machine
dan Naïve Bayes terhadap Penjualan Produk D'Alba

Oleh

Nama : Nafisah Cendekia Ramadhani
NIM : 00000059869
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Kamis, 05 Juni 2025

Pukul 10.00 s.d 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

Wella, S.Kom., M.M.S.I.
0305119101

Pembimbing

Prof. Friska Natalia, Sk.Kom., M.T
0306128307

Ketua Program Studi Sistem Informasi

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

Penguji

Dr. Irmawati, S.Kom., M.M.S.I.
081431

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

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

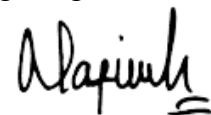
Yang bertanda tangan di bawah ini:

Nama : Nafisah Cendekia Ramadhani
NIM : 00000059869
Program Studi : Sistem Informasi
Jenjang : S1
Judul Karya Ilmiah : Analisis Sentimen Komentar Instagram
Menggunakan Support Vector Machine dan
Naïve Bayes terhadap Penjualan Produk
D'Alba

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, 16 Mei 2025



Nafisah Cendekia Ramadhani

* 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 Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, karena-Nya dapat diselesaikan Skripsi dengan judul: Analisis Sentimen Komentar Instagram Menggunakan Support Vector Machine dan Naïve Bayes terhadap Penjualan Produk D'Alba. Laporan Skripsi ini dilaksanakan sebagai salah satu syarat wajib kelulusan pada jurusan Sistem Informasi Fakultas Teknik dan Informatika di Universitas Multimedia Nusantara. Penulis paham dan menyadari bahwa penyusunan Skripsi ini tidak akan selesai tanpa bantuan dan bimbingan dari berbagai pihak, oleh karena itu penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik Informatika Universitas Multimedia Nusantara.
3. Bu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bu Prof. Dr. Friska Natalia, S.Kom., M.T, sebagai Dosen Pembimbing 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.
6. Teman-teman serta pasangan saya yang telah memberikan semangat, dukungan, dan kebersamaan selama proses penyusunan tugas akhir ini.

Semoga Skripsi ini dapat bermanfaat dan menjadi sumber informasi bagi para pembaca.

Tangerang, 16 Mei 2025



Nafisah Cendekia Ramadhani

ANALISIS SENTIMEN KOMENTAR INSTAGRAM MENGUNAKAN SUPPORT VECTOR MACHINE DAN NAÏVE BAYES TERHADAP PENJUALAN PRODUK D'ALBA

Nafisah Cendekia Ramadhani

ABSTRAK

Penelitian ini dilatarbelakangi oleh tingginya popularitas produk *skincare* D'Alba di Indonesia serta meningkatnya aktivitas konsumen di media sosial, khususnya Instagram. Komentar pengguna dinilai memiliki potensi untuk mencerminkan opini publik dan memengaruhi keputusan pembelian di *marketplace* seperti Shopee dan Tokopedia.

Penelitian menggunakan algoritma Support Vector Machine (SVM) dan Naïve Bayes untuk mengklasifikasikan sentimen komentar menjadi positif, negatif, dan netral. Metode yang digunakan mengacu pada kerangka CRISP-DM, dengan penerapan TF-IDF dalam tahap pemodelan.

Berdasarkan hasil pengujian yang sudah dilakukan, algoritma SVM menunjukkan akurasi yang lebih tinggi sebesar 90,05% dibandingkan dengan Naïve Bayes yang hanya mencapai 83,89%. Namun, korelasi antara sentimen media sosial dan penjualan tergolong rendah. Temuan ini menunjukkan perlunya pendekatan statistik lanjutan seperti uji statistik lanjutan dengan p-value untuk memahami lebih dalam pengaruh opini digital terhadap pembelian.

Kata kunci: analisis sentimen, d'alba, instagram, naïve bayes, support vector machine, *marketplace*.

SENTIMENT ANALYSIS OF INSTAGRAM COMMENTS USING SUPPORT VECTOR MACHINE AND NAÏVE BAYES ON D'ALBA PRODUCT SALES

Nafisah Cendekia Ramadhani

ABSTRACT (English)

This research was motivated by the high popularity of D'Alba skincare products in Indonesia and the increasing consumer activity on social media, especially Instagram. User comments are considered to have the potential to reflect public opinion and influence purchasing decisions on marketplaces such as Shopee and Tokopedia.

The study used the Support Vector Machine (SVM) and Naïve Bayes algorithms to classify comment sentiment into positive, negative, and neutral. The method used refers to the CRISP-DM framework, with the application of TF-IDF in the modeling stage.

Based on the results of the tests that have been carried out, the SVM algorithm shows a higher accuracy of 90.05% compared to Naïve Bayes which only reached 83.89%. However, the correlation between social media sentiment and sales is relatively low. These findings suggest the need for advanced statistical approaches such as advanced statistical tests with p -values to better understand the influence of digital opinions on purchases.

Keywords: *d'alba, instagram, naïve bayes, sentiment analysis, support vector machine, marketplace.*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR	vi
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Batasan Masalah	7
1.4 Tujuan dan Manfaat Penelitian	8
1.4.1 Tujuan Penelitian	8
1.4.2 Manfaat Penelitian	9
1.5 Sistematika Penulisan	9
BAB II LANDASAN TEORI	10
2.1 Penelitian Terdahulu	10
2.2 Teori Penelitian	17
2.2.1 Skincare	17
2.2.2 D'Alba	17
2.2.3 Instagram	18
2.2.4 Marketplace (Shopee dan Tokopedia)	19
2.2.5 Analisis Sentimen	19
2.3 Framework dan Algoritma Penelitian	20
2.3.1 Support Vector Machine (SVM)	20
2.3.2 Naïve Bayes	22
2.3.3 CRISP-DM	23

2.3.4	<i>Text-Mining</i>	24
2.3.5	Vektorisasi	26
2.3.6	<i>Confusion Matrix</i>	27
2.4	<i>Tools dan Software Penelitian</i>	28
2.4.1	Microsoft Excel.....	28
2.4.2	<i>Anaconda</i>	28
2.4.3	<i>Jupyter notebook</i>	29
2.4.4	<i>Python</i>	29
BAB III METODOLOGI PENELITIAN		30
3.1	Gambaran Umum Objek Penelitian.....	30
3.2	Metode Penelitian.....	31
3.2.1	Perbandingan Metode.....	32
3.2.2	Perbandingan Algoritma	33
3.3	Teknik Pengumpulan Data	34
3.3.1	Periode Pengambilan Data	34
3.4	Teknik Analisis Data.....	34
3.4.1	Rumusan Hipotesis	36
3.4.2	Variabel Data.....	37
3.4.2.1	Variabel Independen.....	37
3.4.2.2	Variabel Dependen.....	37
3.4.3	Teknik Pengambilan Sampel	37
3.5	Teknik Pengujian	38
3.6	Alur Penelitian.....	39
3.6.1	Identifikasi Masalah (<i>Business Understanding</i>)	40
3.6.2	Pengumpulan Data (<i>Data Understanding</i>)	40
3.6.3	Pra-proses Data (<i>Data Preparation</i>).....	41
3.6.4	Penerapan Algoritma (<i>Modeling</i>)	44
3.6.5	Evaluasi Model (<i>Evaluation</i>)	45
3.6.6	Interpretasi Hasil dan Kesimpulan (<i>Deployment</i>).....	45
BAB IV ANALISIS DAN HASIL PENELITIAN		47
4.1	<i>Business Understanding</i>	47
4.2	<i>Data Understanding</i>	47

4.3	<i>Data Preparation</i>	57
4.3.1	<i>Case Folding</i>	57
4.3.2	<i>Tokenization</i>	58
4.3.3	<i>Cleaning dan Stopword removal</i>	59
4.3.4	<i>Stemming</i>	60
4.3.5	<i>Labeling</i>	61
4.4	<i>Modeling</i>	64
4.4.1	<i>Splitting Data</i>	64
4.4.2	<i>Vektorisasi TF-IDF</i>	66
4.4.3	<i>Penerapan Algoritma SVM</i>	67
4.4.4	<i>Penerapan Algoritma Naïve Bayes</i>	69
4.5	<i>Evaluation</i>	70
4.5.1	<i>Confusion Matrix</i>	70
4.5.2	<i>Perbandingan Akurasi Model</i>	72
4.5.3	<i>Visualisasi Hasil Sentimen</i>	74
4.5.4	<i>Korelasi dengan Penjualan di Marketplace</i>	77
4.5.4.1	<i>Korelasi Spearman Komentar vs Quantity Sold</i>	77
4.5.4.2	<i>Korelasi Spearman Komentar vs Rating Produk</i>	79
4.5.4.3	<i>Korelasi Spearman Harga vs Quantity Sold</i>	79
4.5.4.4	<i>Hasil Uji P-Value</i>	80
4.5.4.5	<i>Heatmap Komentar vs Quantity Sold</i>	81
4.5.4.6	<i>Heatmap Komentar vs Rating</i>	82
4.5.4.7	<i>Visualisasi Bubble Chart Rating, Penjualan, dan Harga</i>	83
4.5.4.8	<i>Visualisasi Dominan Sentimen</i>	84
4.5.4.9	<i>Visualisasi Harga Produk berdasarkan Sentimen</i>	85
4.6	<i>Deployment</i>	85
4.7	<i>Hasil Penelitian</i>	87
4.8	<i>Pembahasan Hasil Penelitian</i>	90
BAB V SIMPULAN DAN SARAN		93
5.1	<i>Simpulan</i>	93
5.2	<i>Saran</i>	94
DAFTAR PUSTAKA		95

LAMPIRAN.....	101
----------------------	------------



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	10
Tabel 3.1 Perbandingan Teknik Data mining.....	32
Tabel 3.2 Perbandingan Algoritma.....	33
Tabel 3.3 Contoh Hasil Scraping.....	34
Tabel 3.4 Perbandingan Tools Analisis Data	35
Tabel 3.5 Hipotesis Penelitian.....	36
Tabel 3.6 Contoh Case Folding	41
Tabel 3.7 Contoh Tokenization	42
Tabel 3.8 Contoh Stopword removal	42
Tabel 3.9 Contoh Stemming.....	43
Tabel 3.10 Contoh Vektorisasi dengan TF-IDF	43
Tabel 4.1 Rangkuman Hasil Korelasi.....	88
Tabel 4.2 Perbandingan Hasil Akurasi.....	89
Tabel 4.3 Komparasi Performa Model	90



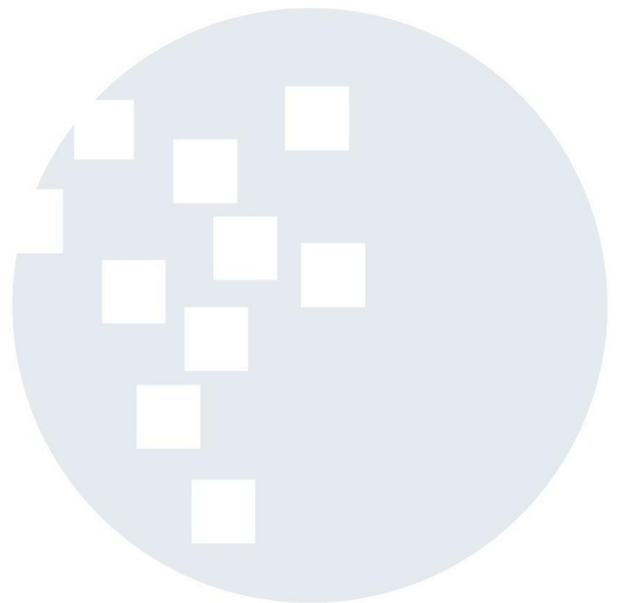
DAFTAR GAMBAR

Gambar 1.1 Pendapatan Produk Skincare di Indonesia.....	1
Gambar 1.2 Jumlah Transaksi Produk Saat Pandemi.....	2
Gambar 1.3 Penjualan Sektor FMCG di E-Commerce Indonesia 2023	2
Gambar 1.4 Sosial Media Ter-Populer 2024	3
Gambar 1.5 E-Commerce dengan Pengunjung Terbanyak	5
Gambar 1.6 Insight Instagram D'Alba.....	6
Gambar 2.1 White Truffle First Spray Serum	18
Gambar 2.2 Hyperplane Negatif dan Positif	21
Gambar 2.3 CRISP-DM Diagram Proses	23
Gambar 2.4 Proses Text mining	24
Gambar 3.1 Flowchart Alur Penelitian.....	39
Gambar 3.2 Alur Preprocessing Data	41
Gambar 4.1 Collection Unggahan D'Alba.....	48
Gambar 4.2 Proses Scraping Data Instagram (1)	48
Gambar 4.3 Proses Scraping Data Instagram (2)	49
Gambar 4.4 Hasil Scraping Data Instagram (Testing)	49
Gambar 4.5 Hasil Scraping Data Instagram (Training).....	50
Gambar 4.6 Hasil Scraping Data Shopee	51
Gambar 4.7 Hasil Scraping Data Tokopedia	52
Gambar 4.8 Hasil Scraping Data Tiktok Shop (Tokopedia)	53
Gambar 4.9 Tampilan Data Merged Marketplace	54
Gambar 4.10 Import dan Load Dataset	54
Gambar 4.11 Info Dataset.....	55
Gambar 4.12 Cek Nilai Kosong	55
Gambar 4.13 Periksa Data Duplikat	56
Gambar 4.14 Hapus Baris Data Duplikat	56
Gambar 4.15 Proses Case Folding.....	57
Gambar 4.16 Hasil Case Folding.....	57
Gambar 4.17 Proses Tokenization.....	58
Gambar 4.18 Hasil Tokenization.....	58
Gambar 4.19 Proses Cleaning	59
Gambar 4.20 Hasil Cleaning	59
Gambar 4.21 Proses Stopword removal	59
Gambar 4.22 Hasil Stopword removal	60
Gambar 4.23 Proses Stemming	60
Gambar 4.24 Hasil Stemming	61
Gambar 4.25 Import Library dan Load Data.....	62
Gambar 4.26 Pembuatan Kamus Lexicon	62
Gambar 4.27 Proses Pelabelan Data Training.....	63
Gambar 4.28 Simpan Hasil Labeling	63
Gambar 4.29 Hasil Data Sesudah di Labeling.....	64

Gambar 4.30 Hasil Pelabelan Data Testing.....	64
Gambar 4.31 Import dan Load Dataset untuk Data Splitting	65
Gambar 4.32 Cek Missing Values	65
Gambar 4.33 Proses Data Splitting	66
Gambar 4.34 Proses Vektorisasi dengan TF-IDF	66
Gambar 4.35 Hasil Analisis TF-IDF	67
Gambar 4.36 Melatih Model SVM.....	68
Gambar 4.37 Prediksi Data Testing dengan Algoritma SVM.....	68
Gambar 4.38 Hasil Akurasi SVM.....	68
Gambar 4.39 Melatih Model Naive Bayes	69
Gambar 4.40 Prediksi Data Testing dengan Naive Bayes.....	69
Gambar 4.41 Hasil Akurasi Naive Bayes	70
Gambar 4.42 Kode untuk Visualisasi Confusion Matrix	71
Gambar 4.43 Hasil Confusion Matrix SVM.....	71
Gambar 4.44 Hasil Confusion Matrix Naive Bayes	72
Gambar 4.45 Hitung Akurasi Kedua Model.....	73
Gambar 4.46 Perbandingan Akurasi SVM dan Naive Bayes	73
Gambar 4.47 Distribusi Sentimen Data Testing	74
Gambar 4.48 Distribusi Sentimen Data Training	75
Gambar 4.49 Pie Chart Distribusi Sentimen Training.....	75
Gambar 4.50 Pembuatan Visualisasi WordCloud	76
Gambar 4.51 Hasil Word Cloud	76
Gambar 4.52 Kode Perhitungan Komentar berdasarkan Label.....	77
Gambar 4.53 Proses Perhitungan Jumlah Komentar Sentimen per Produk	78
Gambar 4.54 Hasil Korelasi antara Sentimen dengan Quantity Sold.....	78
Gambar 4.55 Hasil Korelasi antara Sentimen dengan Rating Produk.....	79
Gambar 4.56 Hasil Korelasi antara Harga dengan Quantity Sold.....	80
Gambar 4.57 Hasil Uji p-value	80
Gambar 4.58 Kode Heatmap Korelasi Komentar dengan Penjualan	81
Gambar 4.59 Hasil Heatmap Korelasi Komentar dengan Penjualan.....	82
Gambar 4.60 Kode Heatmap Korelasi Komentar dengan Rating	82
Gambar 4.61 Hasil Heatmap Korelasi Komentar dengan Rating.....	83
Gambar 4.62 Bubble Chart Rating, Penjualan, dan Harga.....	83
Gambar 4.63 Visualisasi Dominan Sentimen.....	84
Gambar 4.64 Visualisasi Distribusi Harga dengan Sentimen.....	85
Gambar 4.65 Tampilan Awal	86
Gambar 4.66 Tampilan Pilih Jenis Klasifikasi	86
Gambar 4.67 Klasifikasi Distribusi Sentimen	87

DAFTAR LAMPIRAN

Lampiran A Turnitin Similiraty Report	101
Lampiran B Form Konsultasi Bimbingan	102



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA