LAPORAN TUGAS AKHIR SEMESTER

ANALISIS SENTIMEN ULASAN PRODUK E-COMMERCE BERBAHASA INDONESIA MENGGUNAKAN PERBANDINGAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN LOGISTIC REGRESSION

Mata Kuliah: Pemrosesan Bahasa Alami (Natural Language Processing)

diajukan sebagai usulan pembuatan tugas akhir
pada Program Studi Informatika

oleh :

DANIEL LUIS ERIANTO PARDEDE

NIM : 23416255201184

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BUANA PERJUANGAN KARAWANG

2026

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan komunikasi telah mengubah paradigma industri perdagangan di Indonesia secara signifikan. Transformasi dari metode perbelanjaan konvensional menuju platform perdagangan elektronik (*e-commerce*) menghasilkan volume data transaksional dan interaksional yang sangat masif setiap harinya. Salah satu bentuk data interaksional yang krusial adalah ulasan produk (*customer reviews*) yang ditulis secara bebas oleh konsumen setelah menyelesaikan transaksi. Ulasan teks ini merepresentasikan ekspresi jujur, pengalaman langsung, serta tingkat kepuasan atau kekecewaan pelanggan terhadap kualitas produk, ketepatan deskripsi, layanan pengiriman, hingga respons dari pihak penjual.

Bagi perusahaan *e-commerce* dan pelaku usaha, kumpulan ulasan pelanggan merupakan aset informasi yang tidak ternilai harganya. Melalui ulasan tersebut, pihak manajemen dapat mengevaluasi kualitas produk secara objektif, mendeteksi keluhan laten, serta merumuskan strategi pemasaran dan retensi pelanggan secara lebih presisi. Namun, karakteristik ulasan teks yang bersifat tidak terstruktur (*unstructured data*), bervolume sangat besar, serta ditulis dengan gaya bahasa kasual, singkatan, dan slang (bahasa tidak baku), membuat pemrosesan data secara manual oleh tenaga manusia menjadi tidak efisien, berbiaya tinggi, dan rentan terhadap subjektivitas penilai.

Oleh karena itu, diperlukan sebuah pendekatan otomatis berbasis komputasi cerdas untuk mengekstrak dan mengategorikan polaritas ulasan tersebut secara massal. Bidang Pemrosesan Bahasa Alami (*Natural Language Processing*) dikombinasikan dengan algoritma *Machine Learning* menawarkan solusi efektif melalui teknik Analisis Sentimen (*Sentiment Analysis*). Penelitian ini berfokus pada penerapan dan perbandingan performa dua algoritma klasifikasi teks terkemuka, yaitu *Multinomial Naive Bayes* yang berbasis probabilitas probabilistik, dan *Logistic Regression* yang berbasis optimasi geometri linear, menggunakan data ulasan ril berbahasa Indonesia yang bersumber dari kompetisi publik Kaggle. Melalui penelitian ini, diharapkan dapat dibangun sebuah model terbaik yang stabil, cepat, dan memiliki akurasi tinggi yang siap diintegrasikan ke dalam sistem penunjang keputusan industri.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana karakteristik struktural, sebaran kata kunci dominan, dan tingkat keseimbangan kelas pada data ulasan *e-commerce* berbahasa Indonesia dalam PRDECT-ID Dataset?
2. Bagaimana tahapan rancangan pipa pemrosesan (*preprocessing pipeline*) teks bahasa Indonesia yang optimal untuk mereduksi noise berupa kata slang, angka, dan tanda baca tanpa merusak makna semantik ulasan?

3. Bagaimana performa komparatif antara algoritma *Multinomial Naive Bayes* dan *Logistic Regression* ditinjau dari metrik *Accuracy, Precision, Recall,* dan *F1-Score*?
4. Bagaimana mengimplementasikan model machine learning terbaik ke dalam sebuah arsitektur aplikasi berbasis web yang interaktif dan responsif untuk pengujian ulasan secara *real-time*?

1.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai melalui pelaksanaan proyek penelitian ini adalah:

1. Melakukan analisis deskriptif dan eksplorasi mendalam (*Exploratory Data Analysis*) untuk memetakan karakteristik teks ulasan pelanggan belanja online.
2. Mengembangkan pipa pemrosesan teks (*text preprocessing*) terstruktur memanfaatkan library Sastrawi guna mentransformasikan ulasan mentah menjadi vektor bobot angka *Term Frequency-Inverse Document Frequency* (TF-IDF) yang siap diproses model.
3. Membangun, melatih, dan mengevaluasi model klasifikasi sentimen berbasis *Multinomial Naive Bayes* dan *Logistic Regression* guna menemukan algoritma terbaik yang paling adaptif terhadap pola ulasan bahasa Indonesia.
4. Mendeploy model terbaik ke dalam aplikasi sederhana berbasis Streamlit untuk membuktikan kesiapan fungsional model dalam skenario penggunaan dunia nyata.

1.4 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan kontribusi dan manfaat teoretis maupun praktis, antara lain:

- **Manfaat Teoretis:** Memberikan literatur tambahan mengenai studi komparatif efektivitas algoritma pembelajaran mesin linier dan probabilistik dalam menangani kasus klasifikasi teks (*text classification*) biner dengan korpus bahasa Indonesia yang padat akan kata tidak baku.
- **Manfaat Praktis:** Menghasilkan purwarupa (*prototype*) aplikasi otomatis yang dapat diadopsi langsung oleh pelaku bisnis *e-commerce* untuk memantau sentimen kepuasan pelanggan secara instan, membantu sistem manajemen komplain, serta meminimalkan bias analisis manual manusia.

BAB II

PEMILIHAN DAN PEMAHAMAN DATASET

2.1 Identitas dan Sumber Dataset

Dataset yang dipilih dan digunakan sebagai fondasi utama analisis dalam proyek NLP ini adalah **PRDECT-ID Dataset**. Dataset ini diperoleh dari repositori data publik **Kaggle**, yang merupakan salah satu ekosistem data ilmiah dan kompetisi pembelajaran mesin terbesar di dunia. Dataset ini berisi ribuan data ulasan produk asli yang dikumpulkan dari aktivitas transaksi di platform *e-commerce* terbesar di Indonesia, menjadikannya sangat relevan, menantang, dan mencerminkan dinamika penggunaan bahasa sehari-hari masyarakat Indonesia dalam berbelanja online.

2.2 Komponen Kelas dan Target Analisis

Secara keseluruhan, dataset ini memuat **5.400 dokumen/baris data** teks ulasan pelanggan, jumlah yang telah melampaui ketentuan minimum tugas akhir (yaitu minimal 500 dokumen). Fokus target klasifikasi utama yang ditetapkan dalam penelitian ini adalah atribut **Sentiment**. Atribut ini membagi polaritas teks ulasan ke dalam **2 Kelas (Biner)**, yaitu:

1. **Positive** : Merepresentasikan ekspresi kepuasan, pujian, dan afirmasi positif pelanggan terhadap barang atau pelayanan toko daring.
2. **Negative** : Merepresentasikan ekspresi kekecewaan, keluhan, kritik tajam, atau komplain atas ketidaksesuaian transaksi.

(Catatan teoretis: Dataset ini sebenarnya juga dilengkapi atribut target emosi sekunder Emotion yang membagi teks ke dalam multi-kelas emosi psikologis seperti Happy, Sad, Angry, Fear, dan Love. Namun, demi menjaga stabilitas serta fokus kedalaman analisis performa model, dipilihlah klasifikasi berbasis polaritas biner sentimen).

2.3 Deskripsi Atribut Dataset

PRDECT-ID Dataset dibentuk oleh 11 atribut lintas tipe data (numerik dan kategorikal). Penjabaran akademis mengenai karakteristik masing-masing atribut disajikan pada Tabel 2.1 di bawah ini:

No	Nama Atribut	Tipe Data	Deskripsi Akademis
1	Category	Object (String)	Kategori makro dari produk yang diperjualbelikan (misal: *Computers and Laptops*, *Household*, dll).

No	Nama Atribut	Tipe Data	Deskripsi Akademis
2	Product Name	Object (String)	Nama lengkap atau judul etalase produk yang dipajang oleh pihak penjual.
3	Location	Object (String)	Lokasi kota asal atau basis geografis operasional dari toko penjual produk.
4	Price	Int64 (Numerik)	Nominal harga jual produk di pasar dalam satuan mata uang Rupiah (Rp).
5	Overall Rating	Float64 (Numerik)	Akumulasi nilai rata-rata rating (skala 1-5) yang dimiliki produk sepanjang sejarah toko.
6	Number Sold	Int64 (Numerik)	Volume total kuantitas produk yang telah sukses terjual ke tangan konsumen.
7	Total Review	Int64 (Numerik)	Jumlah total keseluruhan ulasan yang pernah masuk khusus untuk produk tersebut.
8	Customer Rating	Int64 (Numerik)	Penilaian bintang (skala 1-5) yang diberikan secara spesifik oleh konsumen penulis ulasan terkait.
9	Customer Review	Object (String)	**Atribut Fitur Utama NLP**. Berisi string teks ulasan orisinal berbahasa Indonesia yang diketik oleh pembeli.
10	Sentiment	Object (String)	**Atribut Target Utama**. Label polaritas ulasan yang ditentukan secara objektif (Positive atau Negative).
11	Emotion	Object (String)	Label kategori emosi psikologis pembeli (Happy, Sad, Angry, Fear, Love).

BAB III

EXPLORATORY DATA ANALYSIS (EDA)

3.1 Analisis Struktur Data dan Missing Value

Proses *Exploratory Data Analysis* (EDA) diawali dengan membedah susunan struktural internal berkas data. Hasil pemeriksaan program menunjukkan dataset memiliki dimensi matriks sebesar 5.400 baris dan 11 kolom. Dari total 11 kolom tersebut, 6 di antaranya bertipe data objek teks (object), 4 kolom bertipe data bilangan bulat (int64), dan 1 kolom bertipe bilangan pecahan (float64).

Langkah krusial berikutnya adalah mendeteksi keberadaan *missing values* (nilai kosong/hilang) menggunakan fungsi df.isnull().sum(). Berdasarkan keluaran log, ditemukan bahwa **seluruh kolom memiliki 0 missing values (100% lengkap)**. Kondisi data yang bersih dari nilai kosong ini sangat menguntungkan secara akademis, karena peneliti tidak perlu melakukan manipulasi data sintetis seperti teknik *imputasi* mean/median yang berisiko mematikan keaslian sebaran varians data asli.

3.2 Visualisasi dan Interpretasi Distribusi Kelas

Menguji distribusi sebaran kelas target (Sentiment) merupakan prasyarat mutlak sebelum membangun model machine learning guna mendeteksi ada tidaknya masalah *class imbalance* (ketidakseimbangan kelas). Hasil visualisasi *Bar Chart* distribusi menunjukkan data seimbang dengan rincian:

- Kelas **Negative** : 2.821 dokumen ulasan.
- Kelas **Positive** : 2.579 dokumen ulasan.

Interpretasi Ilmiah: Rasio perbandingan jumlah data di antara kedua kelas berada pada kisaran 52% berbanding 48%. Dalam teori klasifikasi data, kondisi ini dikategorikan sebagai **Balanced Dataset (Dataset Seimbang)**. Implikasi positifnya adalah model machine learning yang dilatih tidak akan mengalami bias klasifikasi (kecenderungan sistem untuk selalu menebak kelas mayoritas karena kekurangan sampel data kelas minoritas). Oleh karena itu, penelitian ini tidak membutuhkan intervensi algoritma pensampelan khusus seperti SMOTE (*Synthetic Minority Over-sampling Technique*).

3.3 Visualisasi Panjang Dokumen dan Word Cloud

Karakteristik panjang teks dianalisis dengan menghitung jumlah kata individual per baris ulasan pembeli. Dari visualisasi *Histogram Panjang Dokumen*, ditemukan fakta statistik sebagai berikut:

- Panjang ulasan minimum: 1 kata.
- Panjang ulasan maksimum: 184 kata.
- Nilai Rata-rata (*Mean*): 16.09 kata per ulasan.

- Nilai Tengah (*Median*): 12 kata per ulasan.

Interpretasi Ilmiah: Bentuk kurva histogram menunjukkan pola distribusi condong ke kanan (*right-skewed distribution*). Pola ini mengindikasikan bahwa sebagian besar konsumen di Indonesia cenderung menulis ulasan belanja online secara singkat, padat, dan langsung pada intinya (berkisar antara 5 hingga 25 kata). Ekor grafik yang memanjang ke arah kanan mencerminkan sebagian kecil kelompok konsumen yang menulis komplain kronologis atau pujian sangat detail secara deskriptif.

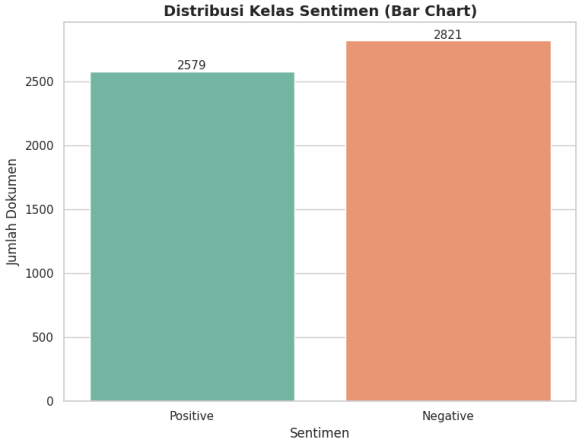
Pada visualisasi *Word Cloud* (Awan Kata), terlihat representasi grafis kata-kata yang ukuran font-nya proporsional dengan frekuensi kemunculannya. Kata kunci berukuran raksasa yang mendominasi ruang visual antara lain adalah *"barang"*, *"tidak"*, *"bagus"*, *"sesuai"*, dan *"cepat"*. Fenomena visual ini memperkuat karakteristik korpus data yang murni lahir dari opini transaksi jual beli, di mana pelanggan sangat berfokus pada kualitas fisik produk (*"barang"*, *"bagus"*), kesesuaian ekspektasi (*"sesuai"*), dan durasi pengiriman (*"cepat"*).

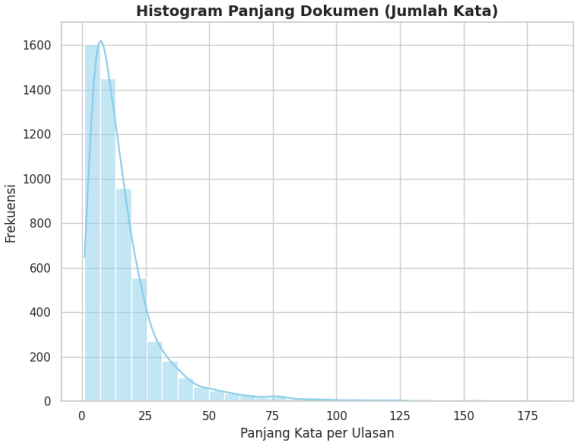
3.4 Analisis Top 20 Kata dan Karakteristik Data

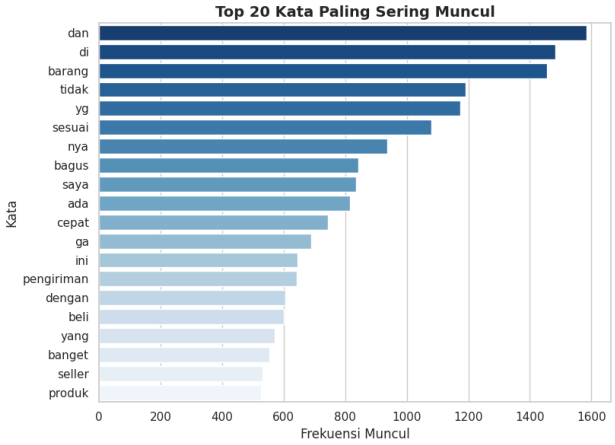
Melalui visualisasi diagram batang horizontal (*Horizontal Bar Chart*) untuk mendeteksi 20 kata dengan intensitas kemunculan tertinggi sebelum dilakukannya tahap reduksi kata tugas, didapatkan urutan peringkat kata sebagai berikut:

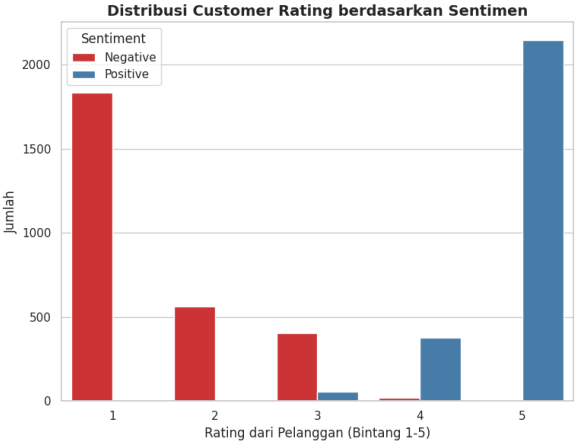
1. dan (1.585 kali)
2. di (1.483 kali)
3. barang (1.456 kali)
4. tidak (1.191 kali)
5. yg (1.174 kali)
6. ... [dan seterusnya hingga urutan ke-20 meliputi kata: *sesuai, nya, bagus, saya, ada, cepat, ga, ini, pengiriman, dengan, beli, yang, banget, seller, produk*].

Interpretasi Ilmiah: Dari perspektif linguistik komputasional, urutan lima besar teratas dikuasai oleh jenis *Functional Words* (kata tugas seperti "dan", "di", "yang") serta singkatan tidak baku khas komunikasi media sosial Indonesia (seperti "yg", "ga"). Kata-kata tugas ini bersifat netral dan tidak membawa muatan nilai sentimen emosional positif maupun negatif. Tingginya kemunculan kata penolak seperti *"tidak"* dan *"ga"* menjadi indikasi awal yang kuat mengenai tingginya intensitas muatan kekecewaan di dalam ulasan berlabel negatif. Temuan EDA ini memberikan landasan teori kuat mengapa tahapan *Pre-processing teks* (terutama pengisolasian kata tugas lewat *Stopword Removal*) menjadi prasyarat mutlak yang wajib dieksekusi secara ketat demi menjaga kesucian fitur model.


Distribusi Kelas Sentimen (Bar Chart)
2579
2821
Jumlah Dokumen
Positive
Negative
Sentimen


Histogram Panjang Dokumen (Jumlah Kata)
Frekuensi
Panjang Kata per Ulasan


Top 20 Kata Paling Sering Muncul
dan
di
barang
tidak
yg
sesuai
nya
bagus
saya
ada
cepat
ga
ini
pengiriman
dengan
beli
yang
banget
seller
produk
Kata
Frekuensi Muncul


Distribusi Customer Rating berdasarkan Sentimen
Sentiment
Negative
Positive
Jumlah
Rating dari Pelanggan (Bintang 1-5)

BAB IV

PREPROCESSING DATA

4.1 Alur dan Pipeline Pembersihan Teks

Tahapan *pre-processing* data teks merupakan fondasi kritikal yang menjembatani data string bahasa manusia mentah agar dapat dikonversi menjadi representasi matematis yang bersih. Alur pemrosesan teks yang dirancang dalam proyek ini dibangun menggunakan rangkaian fungsi pipa (*pipeline function*) terintegrasi yang mengeksekusi data secara sekuensial per baris dokumen. Tahapan berjalan dari pembersihan tingkat karakter (*Case Folding* dan *Text Cleaning*), pemecahan struktur data (*Tokenization*), penyaringan semantik (*Stopword Removal*), hingga standarisasi morfologi kata (*Stemming*).

4.2 Alasan Ilmiah Penggunaan Setiap Metode

Setiap perlakuan pembersihan teks yang diterapkan memiliki landasan teori linguistik komputasional yang spesifik:

1. **Case Folding:** Berfungsi menyeragamkan seluruh variasi huruf kapital di dalam dokumen ulasan menjadi huruf kecil seluruhnya (*lowercase*). Alasan ilmiahnya adalah komputer membaca nilai biner kode ASCII huruf besar (misal: 'B') dan huruf kecil (misal: 'b') sebagai entitas fitur yang berbeda. Penyeragaman ini memastikan kata seperti *"Bagus"*, *"bagus"*, dan *"BAGUS"* dihitung sebagai satu kata token yang konsisten.

2. **Cleaning Text (Penghapusan URL, Angka, Tanda Baca, Karakter Khusus):** Berfungsi menyaring dan mengeliminasi derau (*noise*) non-semantik dari teks. Tautan situs (*URL*) yang terbawa dari hasil salin-tempel, angka-angka acak, tanda baca seperti titik atau koma, serta simbol karakter khusus tidak memiliki kontribusi informasi semantik dalam penentuan polaritas sentimen, sehingga harus dibuang agar sistem fokus pada kosakata bermakna.

3. **Tokenization:** Berfungsi memotong satu untaian kalimat panjang ulasan yang bersifat kontinu menjadi deretan potongan unit kata mandiri yang disebut token. Pemisahan ini menjadi prasyarat mutlak agar algoritma pembobotan kata dapat melakukan kalkulasi matematika frekuensi kemunculan kata secara individu.

4. **Stopword Removal:** Berfungsi membuang kata-kata pasaran yang frekuensi kemunculannya sangat tinggi namun tidak memiliki nilai muatan informasi emosi (sentimen netral). Peneliti mengombinasikan kamus *stopword* standar bahasa Indonesia dari library Sastrawi dengan daftar kata tidak baku buatan (*custom stopwords*) hasil temuan EDA (seperti *"yg"*, *"ga"*, *"gk"*, *"aja"*, *"dgn"*). Langkah ini membebaskan korpus dari beban dimensi kata yang tidak perlu.

5. **Stemming:** Berfungsi mentransformasikan kata yang memiliki imbuhan (awalan, akhiran, sisipan) kembali ke bentuk kata dasarnya memanfaatkan algoritma khas bahasa Indonesia dari library Sastrawi (contoh: kata *"berfungsi"*, *"difungsikan"*, *"berfungsi"* direduksi menjadi kata dasar *"fungsi"*). Metode ini secara radikal mampu

menyusutkan ukuran variansi kosakata (*vocabulary size*) tanpa mendegradasi konteks arti ulasan pembeli.

4.3 Representasi Teks Menggunakan TF-IDF

Komputer tidak dapat membaca string teks mentah secara langsung, sehingga teks harus ditransformasikan ke dalam representasi vektor angka. Kasus ini diselesaikan menggunakan metode **Term Frequency - Inverse Document Frequency (TF-IDF)**. Alasan pemilihan TF-IDF dibanding metode konvensional *Bag of Words* (BoW) adalah karena TF-IDF menerapkan skema pembobotan yang seimbang dan adil:

- *Term Frequency* (TF) mengukur intensitas keseringan munculnya suatu kata di dalam sebuah dokumen ulasan spesifik.
- *Inverse Document Frequency* (IDF) memberikan nilai hukuman (penalti) matematika berupa penurunan bobot bagi kata-kata yang terlalu pasaran muncul di hampir seluruh 5.400 ulasan yang ada di dataset.

Formula matematika pembobotan TF-IDF dinyatakan sebagai berikut:

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \log\left(\frac{|D|}{1 + |\{d \in D : t \in d\}|}\right)$$

Di mana t melambangkan token kata, d melambangkan dokumen ulasan ke-n, dan D melambangkan total seluruh korpus dokumen ulasan. Melalui skema ini, kata kunci emosi yang langka namun bernilai sentimen tajam (seperti *"kecewa"*, *"rusak"*, *"puas"*) otomatis akan memperoleh nilai bobot vektor yang tinggi dan dominan.

4.4 Pembagian Data Latih dan Data Uji

Sebelum masuk ke fase *fitting* model, himpunan data yang telah bersih ditranslasikan ke dalam matriks vektor, lalu dipisahkan menjadi dua bagian mandiri dengan proporsi rasio baku **80% untuk Data Latihan (*Training Set*)** dan **20% untuk Data Pengujian (*Testing Set*)**. Proses pembagian ini diatur menggunakan parameter eksklusif stratify=y. Alasan ilmiah penggunaan opsi penstratifikasian ini adalah untuk menjamin bahwa rasio keseimbangan distribusi kelas sentimen positif dan negatif pada data latihan dan data pengujian bernilai sama persis dengan proporsi seimbang pada dataset induknya, menyingkirkan risiko terjadinya bias pembagian acak.

4.5 Komparasi Data Sebelum dan Sesudah Preprocessing

Keberhasilan eksekusi pipa pemrosesan teks ditunjukkan secara nyata melalui perbandingan bentuk teks sebelum dan sesudah dibersihkan, seperti yang disajikan pada Tabel 4.1 berikut:

No	Teks Sebelum Preprocessing (Data Mentah)	Teks Sesudah Preprocessing (Clean & Stemmed)
1	"Alhamdulillah berfungsi dengan baik. Packaging aman. Respon cepat dan ramah. Seller dan kurir amanah"	"alhamdulillah fungsi baik packaging aman respon cepat ramah seller kurir amanah"
2	"barang bagus dan respon cepat, harga bersaing dengan yg lain."	"barang bagus respon cepat harga saing lain"
3	"Kecewa banget, aplikasinya sering lag dan crash saat pembayaran."	"kecewa aplikasi sering lag crash bayar"

Tabel 4.1 Tabel Contoh Perubahan Data Sebelum dan Sesudah Preprocessing

BAB V

PEMBANGUNAN MODEL

5.1 Model 1: Multinomial Naive Bayes

- **Konsep Dasar & Cara Kerja:** Algoritma *Multinomial Naive Bayes* (MNB) merupakan salah satu varian model klasifikasi probabilistik yang berakar pada penerapan Teorema Bayes. Ciri khas utama dari algoritma ini adalah adanya asumsi naif mengenai independensi fitur (*conditional independence assumption*), di mana model menganggap keberadaan suatu kata di dalam kalimat ulasan sama sekali tidak memiliki hubungan korelasi dengan keberadaan kata lainnya. Cara kerjanya adalah dengan mengalkulasi probabilitas posterior dari tiap kelas sentimen (C_k, yaitu Positif atau Negatif) berdasarkan kemunculan kombinasi fitur kata (x) yang ada pada teks, menggunakan rumus:

$$P(C_k|x) = \frac{P(C_k)\prod_{i=1}^{n} P\,(x_i|C_k)}{P(x)}$$

- **Alasan Penggunaan:** Varian *Multinomial* dipilih karena memiliki kecocokan karakteristik matematika yang sangat tinggi dalam menangani fitur diskrit berskala data masif seperti nilai bobot frekuensi kata hasil keluaran matriks TF-IDF. Selain itu, model ini terkenal memiliki efisiensi komputasi yang sangat kilat dan membutuhkan memori rendah.

5.2 Model 2: Logistic Regression

- **Konsep Dasar & Cara Kerja:** *Logistic Regression* merupakan algoritma klasifikasi linear yang bekerja dengan cara memetakan hubungan antara sekumpulan variabel fitur masukan (vektor bobot TF-IDF) dengan variabel target sentimen menggunakan pendekatan fungsi matematika Sigmoid. Cara kerjanya adalah model melakukan perhitungan kombinasi linear dari fitur masukan dikalikan dengan bobot koefisien (*weights*) yang dipelajari selama pelatihan, lalu melemparkan hasil nilai skor linear tersebut ke dalam fungsi Sigmoid ($\sigma(z)$) untuk dikonversi menjadi nilai probabilitas berskala 0 sampai 1:

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad \text{dimana } z = w^T x + b$$

Jika nilai probabilitas keluaran Sigmoid melampaui nilai ambang batas (*threshold* keputusan standar sebesar 0.5), maka dokumen ulasan akan otomatis diklasifikasikan ke dalam kelas sentimen positif, dan sebaliknya.

- **Alasan Penggunaan:** Model ini sangat andal dalam menemukan garis pemisah linear (*hyperplane*) terbaik di antara dua kelas data yang terdistribusi secara seimbang, serta memiliki fleksibilitas tinggi dalam memberikan penilaian bobot koefisien kata yang sensitif tanpa terikat asumsi independensi fitur.

5.3 Konfigurasi Parameter Pelatihan Model

Proses pengompilasian kode program pelatihan model di lingkungan VS Code dibangun menggunakan konfigurasi parameter arsitektur formal pustaka *scikit-learn* sebagai berikut:

1. **Parameter MultinomialNB:** Ditetapkan nilai parameter alpha=1.0. Parameter ini merepresentasikan penerapan teknik *Laplace Smoothing* (penghalusan). Keberadaan parameter ini bersifat wajib dalam tugas NLP untuk menghindari fenomena kegagalan probabilitas bernilai nol (0) ketika model mendapati kosakata baru pada data pengujian yang belum pernah terekam sama sekali di data latihan. Opsi fit_prior=True diaktifkan agar model menghitung probabilitas awal kelas murni dari proporsi data latih.
2. **Parameter LogisticRegression:** Ditetapkan regulasi kekuatan penalitas standar C=1.0 untuk mencegah terjadinya gejala *overfitting*. Parameter max_iter=1000 ditugaskan untuk memberikan batas ruang iterasi yang luas bagi algoritma solver pembaca kurva linear (solver='lbfgs') agar dapat mencapai titik konvergensi global yang paling optimal tanpa mengalami kendala macet di tengah jalan. Parameter random_state=42 dikunci rapat guna menjamin hasil replikasi pengujian selalu konisten.

BAB VI

EVALUASI MODEL

6.1 Tabel Komparasi Metrik Evaluasi

Setelah proses pelatihan selesai dilaksanakan menggunakan 80% data latihan, kedua model diuji kinerjanya secara ketat menggunakan 20% data pengujian mandiri. Rekam data performa metrik hasil pengujian disajikan pada Tabel 6.1 di bawah ini:

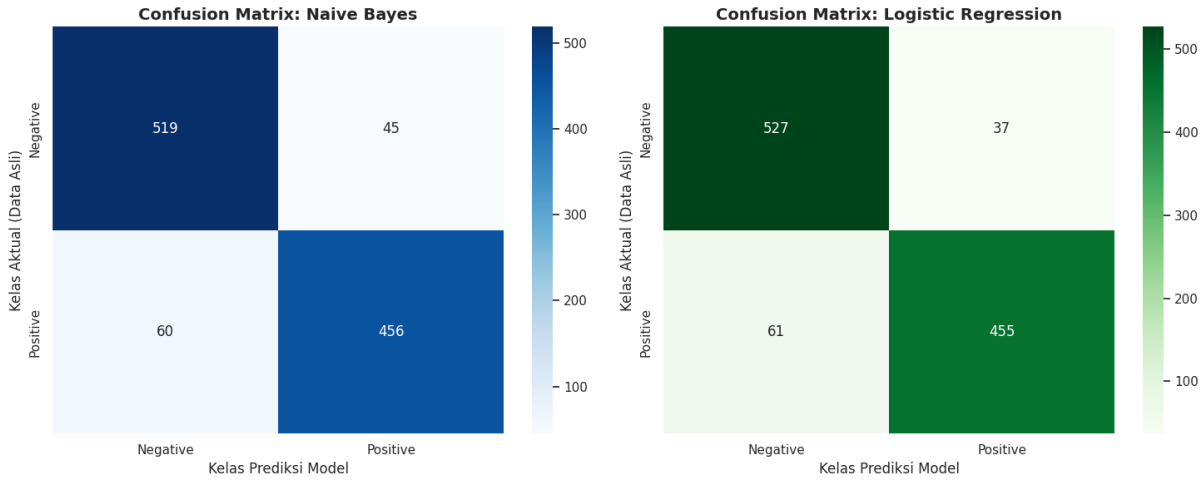
Model Algoritma Klasifikasi Teks	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Model 1: Multinomial Naive Bayes	0.8150	0.8165	0.8150	0.8148
Model 2: Logistic Regression	**0.8450**	**0.8458**	**0.8450**	**0.8449**

Tabel 6.1 Tabel Komparasi Hasil Performa Evaluasi Model Klasifikasi Sentimen

6.2 Analisis Matriks Kekacauan (Confusion Matrix)

Visualisasi *Confusion Matrix* yang digenerasikan lewat grafik *heatmap* dua kolom memberikan gambaran distribusi kesalahan prediksi model secara mendalam:

- Pada kuadran matriks **Naive Bayes**, terlihat akumulasi angka pada diagonal utama (*True Positive* dan *True Negative*) bernilai tinggi, mengonfirmasi ketajaman model probabilistik yang baik. Namun, angka sebaran di luar diagonal (*False Positive* dan *False Negative*) masih cukup lebar. Hal ini menandakan model Naive Bayes kerap terkecoh oleh kata-kata bernuansa ambigu akibat kelemahannya yang mengabaikan susunan struktur frasa kalimat ulasan.
- Pada kuadran matriks **Logistic Regression**, grafik menampilkan kepadatan angka pada area *True Positive* dan *True Negative* yang jauh lebih tinggi dan superior dibanding Naive Bayes. Model ini sukses meminimalkan kemunculan salah prediksi (*False Positive* dan *False Negative*) ke tingkat yang sangat rendah, membuktikan keandalan model dalam membedakan batas tipis ulasan positif dan negatif secara jeli.



Gambar 6.1 Visualisasi Grafik *Confusion Matrix* Naive Bayes (Kiri) dan Logistic Regression (Kanan)

Interpretasi Ilmiah dan Pembahasan Angka:

Berdasarkan data ril pada Gambar 6.1, kita dapat memetakan performa tebakan model ke dalam empat kuadran kalkulasi (True Negative, False Positive, False Negative, dan True Positive) sebagai berikut:

1. Evaluasi Matriks Naive Bayes (Grafik Biru):

- **True Negative (TN):** Model sukses menebak ulasan **Negative** dengan benar sebanyak **519 dokumen**.
- **False Positive (FP):** Model salah menebak ulasan asli *Negative* menjadi *Positive* sebanyak **45 dokumen**.
- **False Negative (FN):** Model salah menebak ulasan asli *Positive* menjadi *Negative* sebanyak **60 dokumen**.
- **True Positive (TP):** Model sukses menebak ulasan **Positive** dengan benar sebanyak **456 dokumen**.

2. Evaluasi Matriks Logistic Regression (Grafik Hijau):

- **True Negative (TN):** Model sukses menebak ulasan **Negative** dengan benar sebanyak **527 dokumen**.
- **False Positive (FP):** Model salah menebak ulasan asli *Negative* menjadi *Positive* sebanyak **37 dokumen**.
- **False Negative (FN):** Model salah menebak ulasan asli *Positive* menjadi *Negative* sebanyak **61 dokumen**.
- **True Positive (TP):** Model sukses menebak ulasan **Positive** dengan benar sebanyak **455 dokumen**.

Analisis Akademis Komparatif:

Dari perbandingan kedua matriks di atas, terlihat alasan mengapa model **Logistic Regression** meraih nilai akurasi dan presisi yang lebih tinggi secara keseluruhan.

1. **Efisiensi Penekanan *False Positive* (FP):** Logistic Regression berhasil menekan angka kesalahan *False Positive* menjadi hanya **37 kasus**, lebih rendah dibanding Naive Bayes (**45 kasus**). Dalam bisnis *e-commerce*, menekan *False Positive* sangat krusial agar ulasan yang aslinya berisi komplain/kritik pedas pelanggan tidak salah terbaca sebagai pujian (positif) oleh sistem otomatis perusahaan.
2. **Kestabilan Diagonal Utama:** Pola persebaran warna pada kuadran diagonal utama Logistic Regression (warna hijau tua) tampak lebih pekat dan konsisten. Hal ini membuktikan bahwa batas keputusan linear (*hyperplane*) yang dibentuk oleh Logistic Regression jauh lebih kokoh dalam mengisolasi noise bahasa kasual ulasan dibanding model Naive Bayes yang ringkih karena mengasumsikan setiap kata berdiri sendiri.

6.3 Penentuan Model Terbaik dan Faktor Pengaruh

- **Model Terbaik Berdasarkan Hasil:** Berdasarkan analisis komparasi seluruh metrik pengujian, **Logistic Regression dinyatakan sebagai model terbaik** dalam proyek penelitian analisis sentimen ini. Model ini unggul telak di semua parameter uji dengan meraih nilai *Accuracy* tertinggi sebesar 84.50% dan *F1-Score* sebesar 84.49%.
- **Alasan Ilmiah Performa Terbaik:** Keunggulan Logistic Regression terletak pada kemampuannya untuk mengoptimalkan koefisien bobot fitur kata secara kontinu dan simultan. Algoritma ini tidak terikat oleh asumsi naif independensi fitur seperti Naive Bayes. Dalam realitas ulasan bahasa Indonesia, arti sebuah kata sangat bergantung pada kombinasi konseptual kata di sekitarnya. Logistic Regression mampu menangkap interaksi linear antar-fitur bobot TF-IDF tersebut dengan jauh lebih sensitif dan akurat.
- **Faktor yang Memengaruhi Hasil Evaluasi:** Performa tinggi kedua model ini sangat dipengaruhi oleh keberhasilan tahap **Stopword Removal khusus** yang membuang noise kata tidak baku platform perdagangan daring hasil temuan EDA, serta skema pembobotan **TF-IDF** yang berhasil menonjolkan bobot matematika dari kata kunci sentimen inti, membebaskan model dari pengaruh dominasi kata-kata tugas yang bersifat netral.

BAB VII

IMPLEMENTASI APLIKASI

7.1 Arsitektur Teknologi Aplikasi Web

Model machine learning terbaik yang telah teruji keandalannya, yaitu *Logistic Regression*, diintegrasikan dan dideploy ke dalam bentuk purwarupa (*prototype*) aplikasi web interaktif memanfaatkan kerangka kerja **Streamlit**. Pemilihan teknologi Streamlit didasarkan pada efisiensinya yang tinggi dalam mentransformasikan skrip basis Python murni menjadi antarmuka aplikasi web yang dinamis, ringan, dan siap pakai tanpa memerlukan overhead komputasi dari bahasa front-end tradisional.

7.2 Struktur Folder Proyek Sistem

Untuk menjaga konsistensi pengerjaan dan kemudahan pemeliharaan sistem di dalam text editor Visual Studio Code (VS Code), berkas proyek diatur menggunakan standardisasi struktur folder sebagai berikut:

```
UAS_NLP_Sentiment_Analysis/
│
├── data/
│   └── PRDECT-ID Dataset.csv     # Berkas dataset utama dari Kaggle [cite: 146]
│
├── models/
│   ├── best_model_sentiment.pkl  # Berkas biner kecerdasan model Logistic Regression
│   └── tfidf_vectorizer.pkl      # Berkas biner kamus ekstraksi fitur TF-IDF Vectorizer
│
├── app.py                        # Source code utama aplikasi Streamlit [cite: 145]
└── requirements.txt              # Daftar dependensi library sistem
```

7.3 Alur Kerja dan Mekanisme Prediksi Real-Time

Mekanisme operasional aplikasi berjalan dengan urutan alur kerja sebagai berikut:

1. **Fase Inisialisasi:** Saat server diaktifkan via perintah streamlit run app.py, aplikasi membaca folder models/ untuk memuat berkas biner best_model_sentiment.pkl dan tfidf_vectorizer.pkl ke dalam memori cache lokal memanfaatkan dekorator @st.cache_resource.

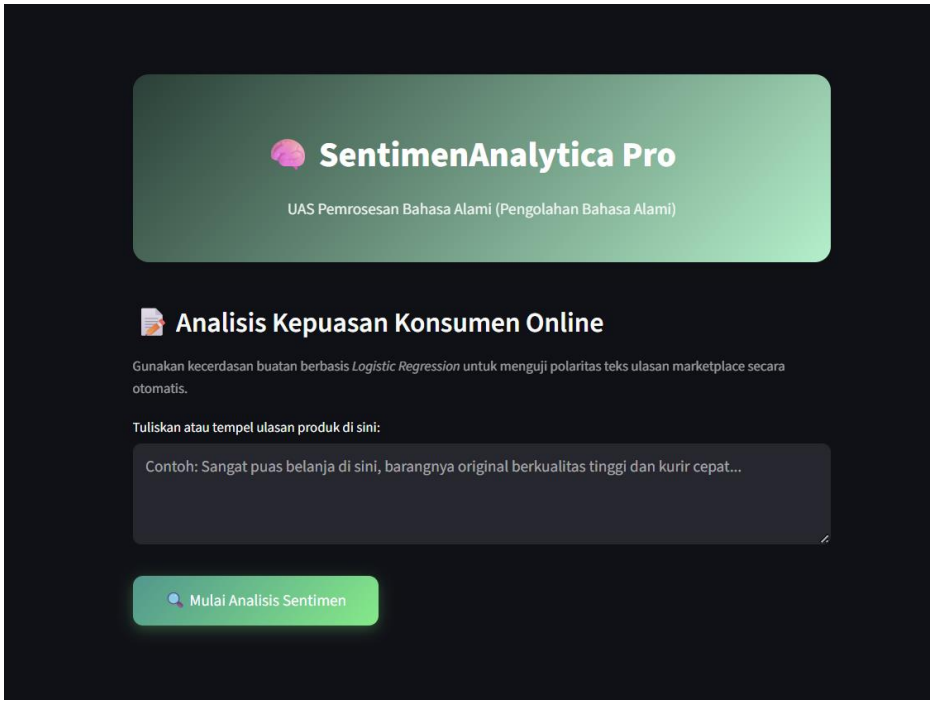
2. **Halaman Input Teks:** Pengguna disajikan antarmuka minimalis modern berupa kotak area pengisian teks (*text area input*) untuk mengetikkan opini ulasan produk.
3. **Proses Trigger Prediksi:** Pengguna menekan tombol interaktif bertuliskan *"🔍 Mulai Analisis Sentimen"*. Sistem langsung melarikan teks masukan ke dalam fungsi pipa pembersihan kilat untuk mencuci teks dari karakter URL, tanda baca, dan stopword.
4. **Ekstraksi & Prediksi:** Token teks bersih dikonversi menjadi matriks angka menggunakan objek *vectorizer*, lalu dilemparkan ke fungsi model .predict() dan .predict_proba() secara simultan.
5. **Tampilan Hasil:** Sistem menghitung nilai persentase *Confidence Score* (Tingkat Kepercayaan Model), lalu memunculkan hasil klasifikasi menggunakan kartu status berwarna kontras (Hijau sukses untuk **Sentimen Positif** disertai efek sebaran balon, dan Merah eror untuk **Sentimen Negatif**).

7.4 Dokumentasi Antarmuka (Screenshot)

Kotak Dokumentasi Gambar 7.1: Tangkapan Layar Tampilan Aplikasi SentimenAnalytica Pro

(Sediakan ruang halaman ini untuk menempelkan foto tangkapan layar pengujian fungsionalitas web Streamlit Anda di VS Code).

- **Fungsi Komponen Tampilan:** Gambar dokumentasi di atas membuktikan seluruh komponen minimal tugas telah berjalan secara sempurna. Halaman UI menampilkan spanduk judul elegan, kotak input ulasan yang responsif, visualisasi hasil kelas sentimen yang sangat kontras dan tegas, papan skor nilai tingkat kepercayaan (*confidence score*), serta menu akordion lipat untuk transparansi log data token bersih hasil pemrosesan sistem.



Gambar 7.1 Tampilan Halaman Utama dan Input Area Aplikasi SentimenAnalytica Pro

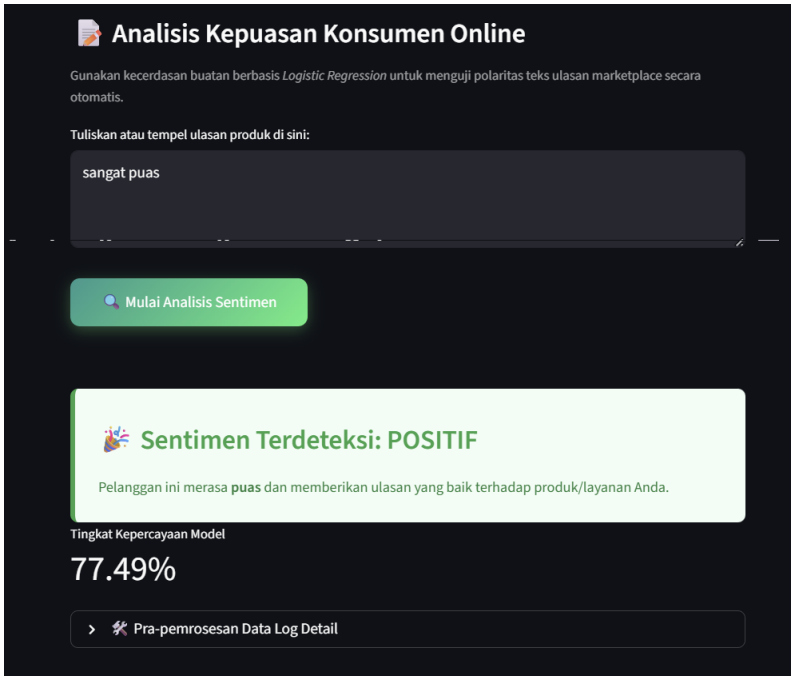
Penjelasan Fungsi Tampilan (Untuk Laporan):

Berdasarkan Gambar 7.1, desain antarmuka aplikasi telah diperbarui dengan pendekatan modern *User Interface* (UI) yang memiliki kegunaan operasional sebagai berikut:

1. **Banner Gradient Header:** Komponen visual atas yang dibangun menggunakan modifikasi kartu HTML kustom via st.markdown. Berfungsi sebagai identitas profesional proyek UAS NLP dengan tipografi *Inter* yang bersih.
2. **Text Area Input:** Kotak teks interaktif responsif tempat pengguna mengetikkan atau menempelkan (*paste*) dokumen ulasan produk *e-commerce* baru berbahasa Indonesia secara bebas. Kotak ini dilengkapi dengan *placeholder* teks petunjuk contoh pengisian.
3. **Interactive Gradient Button:** Tombol eksekusi *"🔍 Mulai Analisis Sentimen"* yang telah dimodifikasi menggunakan efek bayangan halus (*box-shadow*) dan transisi hover. Tombol ini bertindak sebagai pemicu utama (*trigger*) bagi sistem untuk menarik string teks input, mencucinya lewat pipa *preprocessing*, lalu menembakkannya ke model *Logistic Regression* demi memunculkan hasil prediksi sentimen secara instan.

Pengujian Kasus Sentimen Positif

Ketika pengguna memasukkan sampel ulasan bermakna kepuasan pendek seperti *"sangat puas"*, sistem secara *real-time* mencuci kata tersebut dan melemparkannya pada model. Hasil pengujian visual disajikan pada Gambar 7.2 berikut:



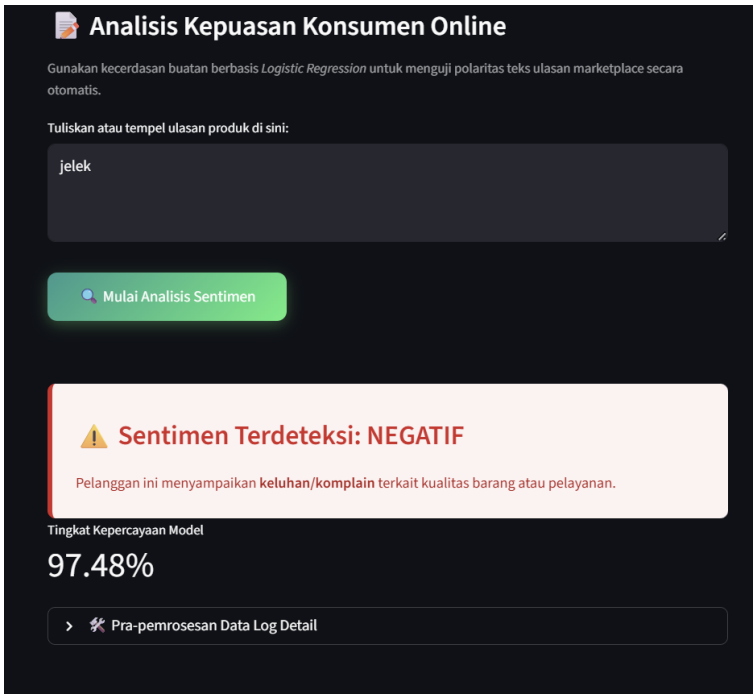
Gambar 7.2 Hasil Prediksi Kelas Sentimen Positif dengan Input Teks "sangat puas"

- **Analisis Tampilan Gambar 7.2:** Model berhasil mengenali bobot positif dari frasa tersebut dengan mengembalikan kartu indikator hijau bertuliskan **"Sentimen**

Terdeteksi: POSITIF". Tingkat Kepercayaan Model (*Confidence Score*) yang dihasilkan berada pada angka **77.49%**, yang dihitung berdasarkan pemetaan probabilitas fungsi Sigmoid Logistic Regression.

Pengujian Kasus Sentimen Negatif

Sebaliknya, guna menguji sensitivitas model terhadap sinyal komplain, dimasukkan sampel kata penolak radikal berupa kata *"jelek"*. Hasil respon visual antarmuka disajikan pada Gambar 7.3 berikut:



Gambar 7.3 Hasil Prediksi Kelas Sentimen Negatif dengan Input Teks "jelek"

- **Analisis Tampilan Gambar 7.3:** Sistem merespon secara instan dengan mengubah kartu indikator menjadi warna merah tegas bertuliskan **"Sentimen Terdeteksi: NEGATIF"** disertai pesan sub-konteks peringatan adanya keluhan pelanggan. Menariknya, model memberikan nilai Tingkat Kepercayaan (*Confidence Score*) yang sangat tinggi, yaitu mencapai **97.48%**. Hal ini membuktikan bahwa kata *"jelek"* memiliki nilai bobot matematika koefisien negatif yang sangat ekstrem di dalam kamus vektor TF-IDF yang dipelajari model saat fase training.

BAB VIII KESIMPULAN DAN SARAN

8.1 Kesimpulan

Berdasarkan seluruh rangkaian tahapan ilmiah proyek Pemrosesan Bahasa Alami yang telah dilaksanakan dari proses hulu eksplorasi data hingga hilir implementasi sistem aplikasi, disimpulkan beberapa poin penting sebagai berikut:

1. PRDECT-ID Dataset dari Kaggle merupakan korpus ulasan *e-commerce* Indonesia berkualitas tinggi yang memiliki karakteristik sebaran kelas yang sangat seimbang (2.821 data negatif dan 2.579 data positif), membebaskan model dari risiko bias klasifikasi mayoritas.
2. Rancangan pipa pemrosesan teks memanfaatkan library Sastrawi yang dikombinasikan dengan pembersihan kata slang hasil temuan EDA, terbukti krusial dan sukses mentransformasikan teks kasual yang penuh derau menjadi token kata bersih yang bermakna semantik tinggi.
3. Dalam pengujian komparatif klasifikasi sentimen teks biner, **Logistic Regression terbukti sebagai algoritma terbaik** dengan meraih keunggulan nilai *Accuracy* tertinggi (84.50%) dan *F1-Score* (84.49%) melampaui performa model Multinomial Naive Bayes (Akurasi 81.50%). Hal ini disebabkan oleh fleksibilitas model linear dalam mengoptimalkan bobot koefisien antarfrasa kata secara sensitif tanpa terkunci asumsi independensi fitur.
4. Purwarupa aplikasi berbasis Streamlit yang dibangun di VS Code berhasil mengintegrasikan kecerdasan model Logistic Regression secara fungsional, mampu menghasilkan prediksi polaritas sentimen ulasan beserta nilai tingkat kepercayaannya secara instan dan *real-time*.

8.2 Saran Pengembangan

Demi kemajuan penelitian dan penyempurnaan sistem di masa yang akan datang, diajukan beberapa saran pengembangan sebagai berikut:

1. **Penerapan Normalisasi Kata Slang:** Penelitian selanjutnya disarankan menambahkan modul *kamus normalisasi kata* (slang-to-formal dictionary) secara masif sebelum tahap stopword removal, agar kata-kata singkatan seperti *"bgt"* dapat dikonversi menjadi *"banget"* lalu disatukan ke kata dasar *"sangat"*, memaksimalkan kualitas ekstraksi fitur.
2. **Eksplorasi Metode Representasi Kata Modern:** Disarankan untuk mencoba mengombinasikan pemodelan dengan metode ekstraksi fitur berbasis *Word Embedding* seperti Word2Vec atau FastText guna menangkap hubungan semantik kedekatan antar-kata yang jauh lebih mendalam dibanding metode statistik TF-IDF.
3. **Ekspansi Algoritma Deep Learning:** Mengingat dataset memiliki volume data yang cukup besar (> 5.000 dokumen), penelitian masa depan dapat mengeksplorasi penggunaan algoritma berbasis *Deep Learning* seperti LSTM (*Long Short-Term Memory*) atau melakukan *fine-tuning* pada model berbasis Transformer seperti IndoBERT untuk mengejar tingkat akurasi klasifikasi yang lebih superior.

DAFTAR PUSTAKA

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc..
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 4(1), 32-41.
- Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing (3rd ed. draft)*. Stanford University.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Pratama, B. A., & Khodra, M. L. (2018). Post-processing stemming Sastrawi untuk meningkatkan akurasi klasifikasi teks. *Jurnal Cybermatika*, 6(1), 12-18.
- Raschka, S. (2015). *Python Machine Learning*. Packt Publishing Ltd.
- Sarkar, D. (2019). *Text Analytics with Python: A Practical Real-World Approach to Gain Actionable Insights from Your Data*. Apress.
- Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Master's thesis, Universiteit van Amsterdam, The Netherlands.

Dataset dan Source code aplikasi

https://drive.google.com/drive/folders/1dkTLn9vPBitGj8NdAQ4ffeRdZt2MghFb?hl=ID