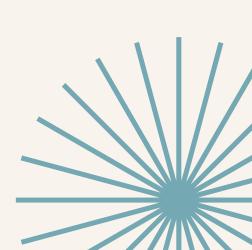
Sentiment Manysis AMAZON PRODUCT REVIEW

Presentation by abde

Alia - Bagus - Ceiba - Givari - Kosmas - Wibi



DATASET

Kaggle

Amazon Product Reviews

568 Ribu Ulasan Konsumen pada Berbagai Produk Amazon

amazon

Jumlah Baris: 568.454

Jumlah Kolom: 10

detail kolom:



- ProductId
- UserId
- ProfileName
- Helpfullness Numerator

- Helpfullness Denominator
- Score
- Time
- Text



missing values: 26 ProfileName kosong, 27 Summary kosong

PRE-PROCESSING



Penghapusan Tanda Baca

Penghapusan Angka

Penghapusan Stopwords

the

Mengeliminasi kata umum *in* bahasa Inggris yang tidak bermakna

and



Stemming

Mengubah kata menjadi bentuk dasarnya menggunakan Porter Stemmer



Mengubah karakter khusus menjadi format ASCII



Menggunakan pola regex untuk pembersihan teks

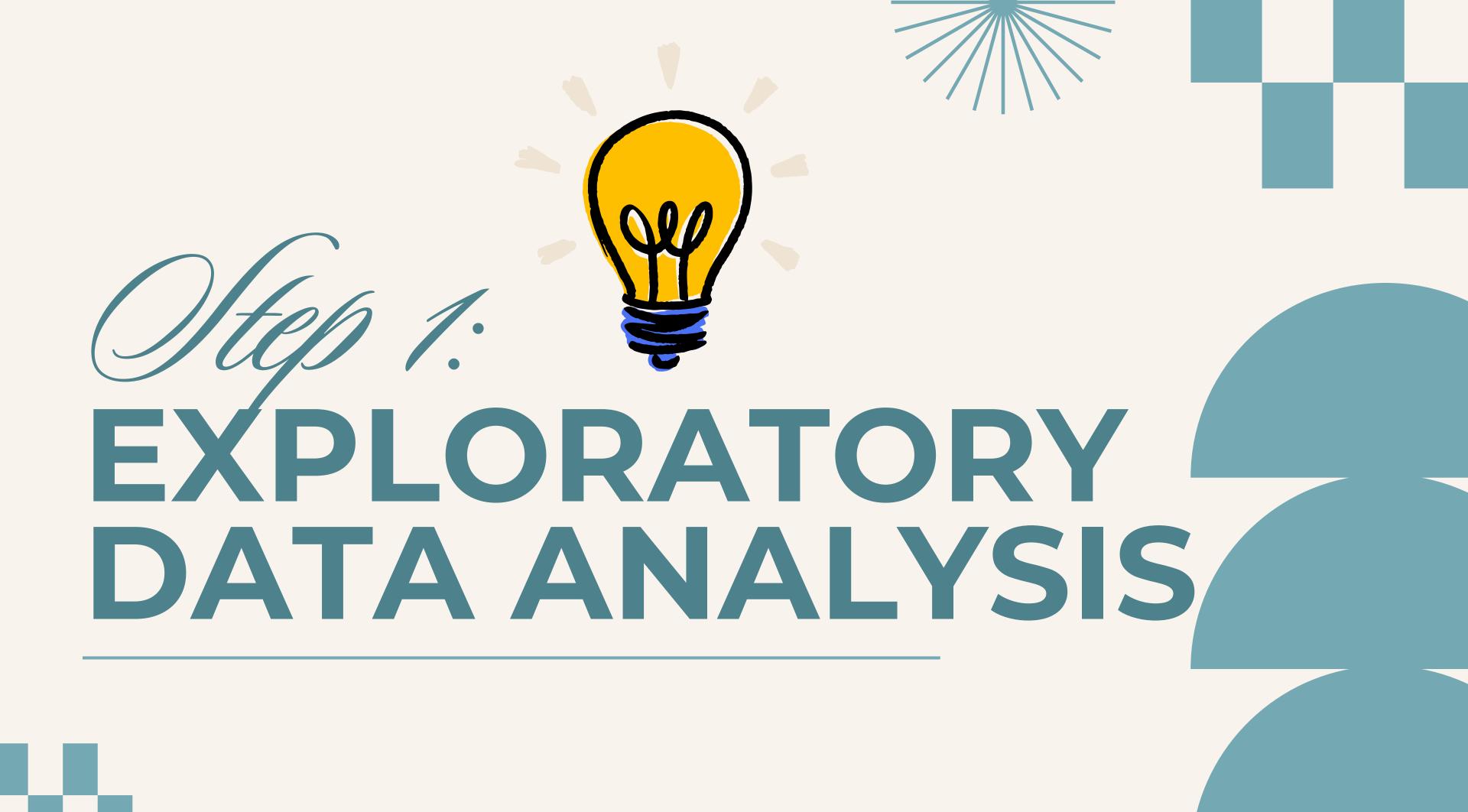
Pembersihan HTML

Menghapus tag dan elemen HTML



Tokenisasi

Memisahkan teks menjadi kata-kata individu



EXPLORATORY DATA ANALYSIS

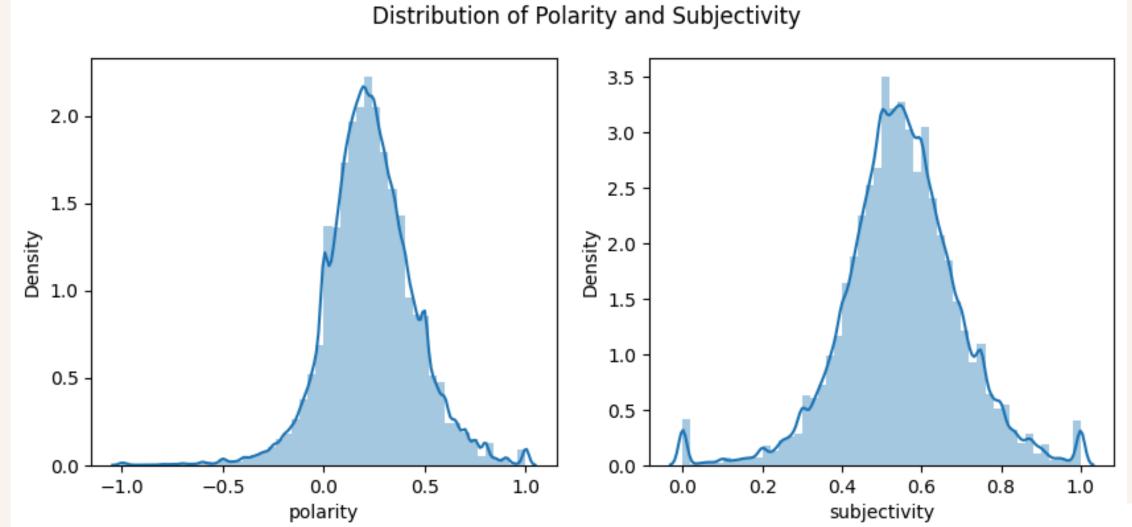
Feature Engineering

- 1. Panjang Teks Review (char length): Rata-rata: 417,1 dan Maksimal: 20.560
- 2. Jumlah Kata (word count): Rata-rata 79,2 kata per ulasan
- 3. **Kepadatan Kata (word density):** Rata-rata 5,14 karakter per kata
- 4. **Skor Polarisasi (text polarity):** Sentimen polaritas menggunakan TextBlob (rentang -1 hingga 1)
- 5. Skor Subjektivitas (text subjectivity): Ukuran objektivitas (rentang 0 hingga 1)



- 1. Rata-rata Polarisasi: 0,233 (sedikit positif)
- 2. Rata-rata Subjektivitas: 0,549 (cukup subjektif)
- 3. Rata-rata Panjang Ulasan: 417 karakter





Kategori Polarisasi

- Positif Lemah: 303.857 ulasan (53,5%)
- Positif: 170.262 ulasan (29,9%)
- Negatif Lemah: 52.595 ulasan (9,3%)
- Positif Kuat: 26.576 ulasan (4,7%)
- Netral: 8.382 ulasan (1,5%)
- Negatif: 5.389 ulasan (0,9%)
- Negatif Kuat: 1.144 ulasan (0,2%)

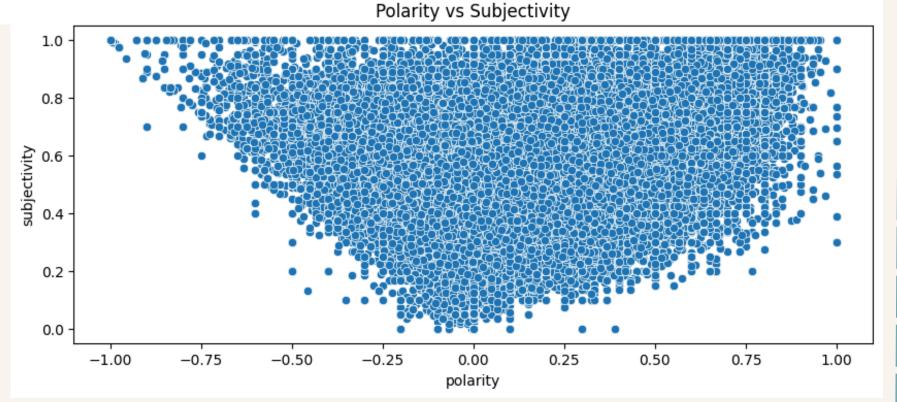
POLARITY VS

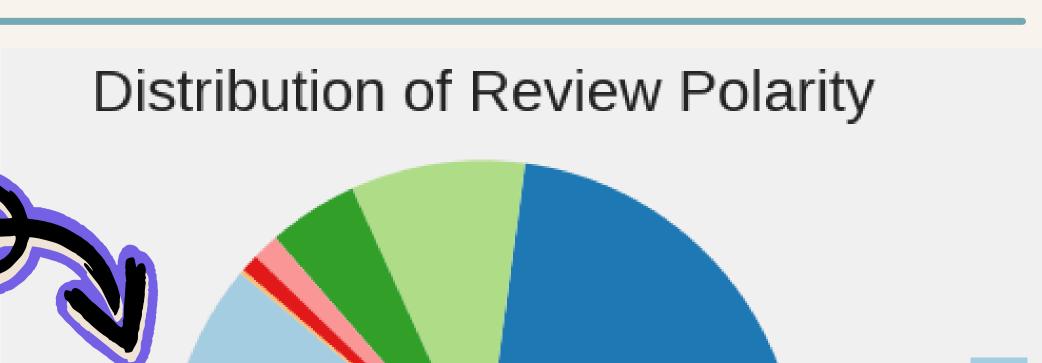


SUBJECTIVITY

Key Findings:

- 87,1% ulasan mengandung sentimen positif
- 10,7% mengandung sentimen negatif
- 1,5% bersifat netral
- Terjadi kecenderungan bias positif yang kuat pada ulasan produk Amazon







Polarity

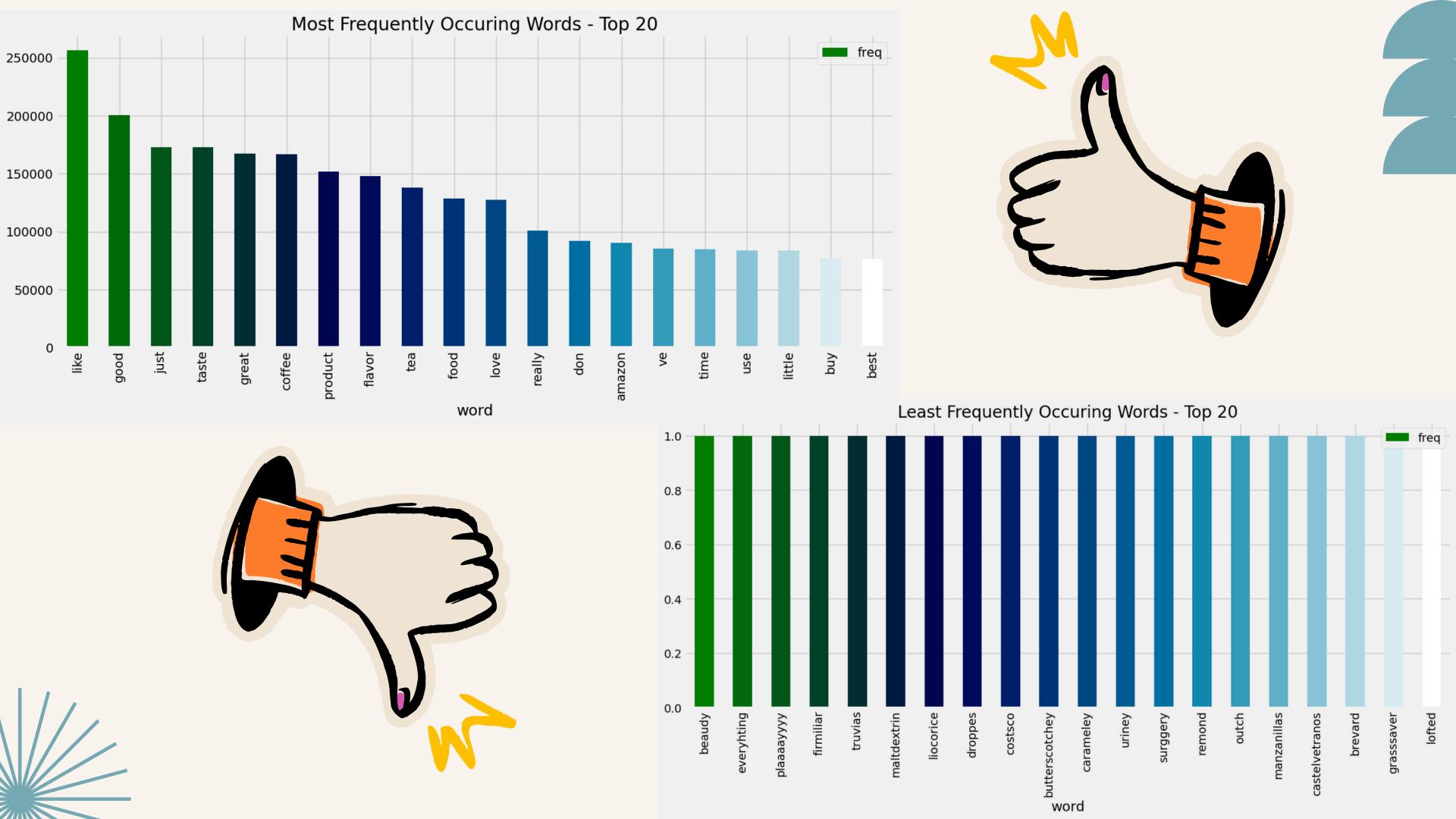
- Weakly Positive: 304220 reviews (53.54%)
- Positive: 172781 reviews (30.41%)
- Weakly Negative: 50800 reviews (8.94%)
- Strongly Positive: 26201 reviews (4.61%)
- Neutral: 7797 reviews (1.37%)
- Negative: 5305 reviews (0.93%)
 - Strongly Negative: 1106 reviews (0.19%)

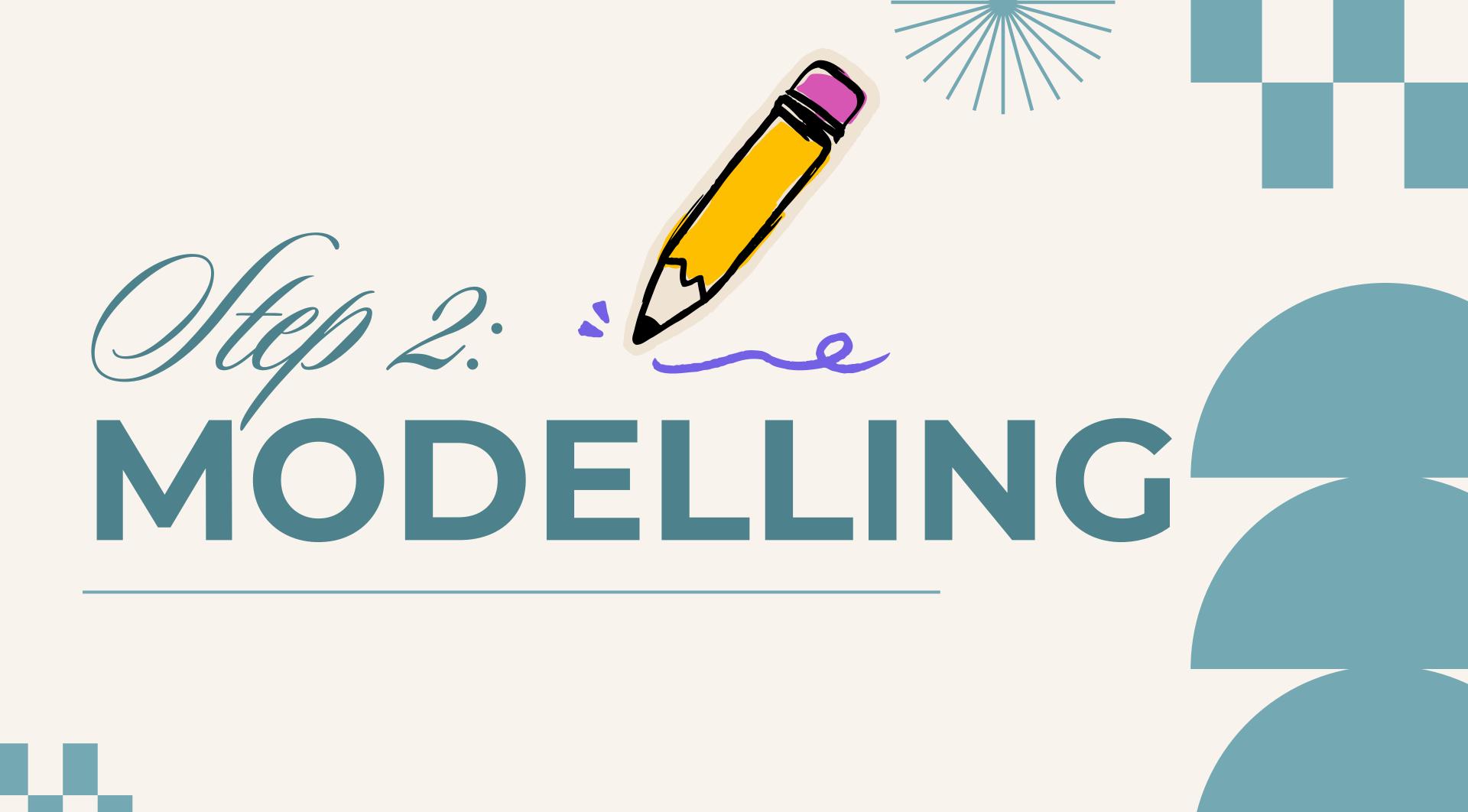


Vocabulary from Reviews loves different treats want candy dogs ereal butter doesn natural

Vocabulary from Reviews







Proses dan Metode

Langkah-langkah

Data preprocessing meliputi:

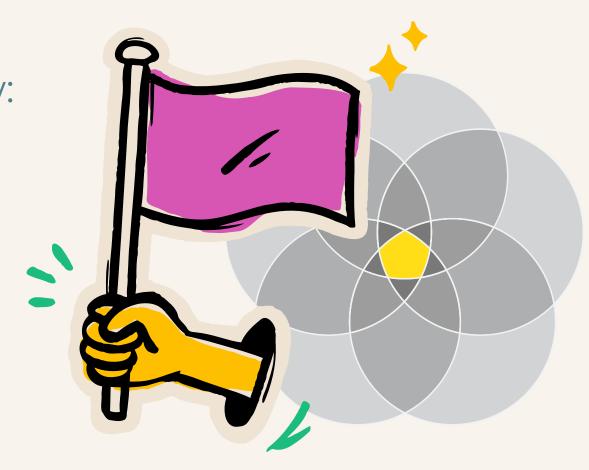
- 1. Teks diubah menjadi huruf kecil
- 2. Pembersihan HTML, penghapusan tanda baca dan stopwords
- 3. Tokenisasi dan stemming pada teks review

Target variabel dibuat biner berdasarkan skor review:

- 1. Score ≥ 3 → Sentimen Positif (1)
- 2. Score < 3 → Sentimen Negatif (0)

Dataset dibagi menjadi:

- Training set: 80% data
- Testing set: 20% data



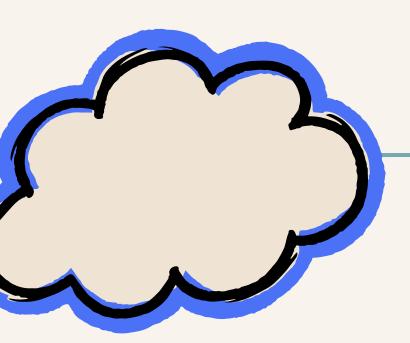
Pipeline dan algoritma

Pipeline yang digunakan:

- CountVectorizer dengan max fitur
 10.000 dan stopwords bahasa Inggris
- TF-IDF Transformer
- Logistic Regression (max_iter=1000)

Model dilatih pada data training dan diuji pada data testing





Result

- Training accuracy: 0.9171

- Test accuracy: 0.9126

- F1 Score: 0.9501 - AUC Score: 0.9349

- Classification Report:

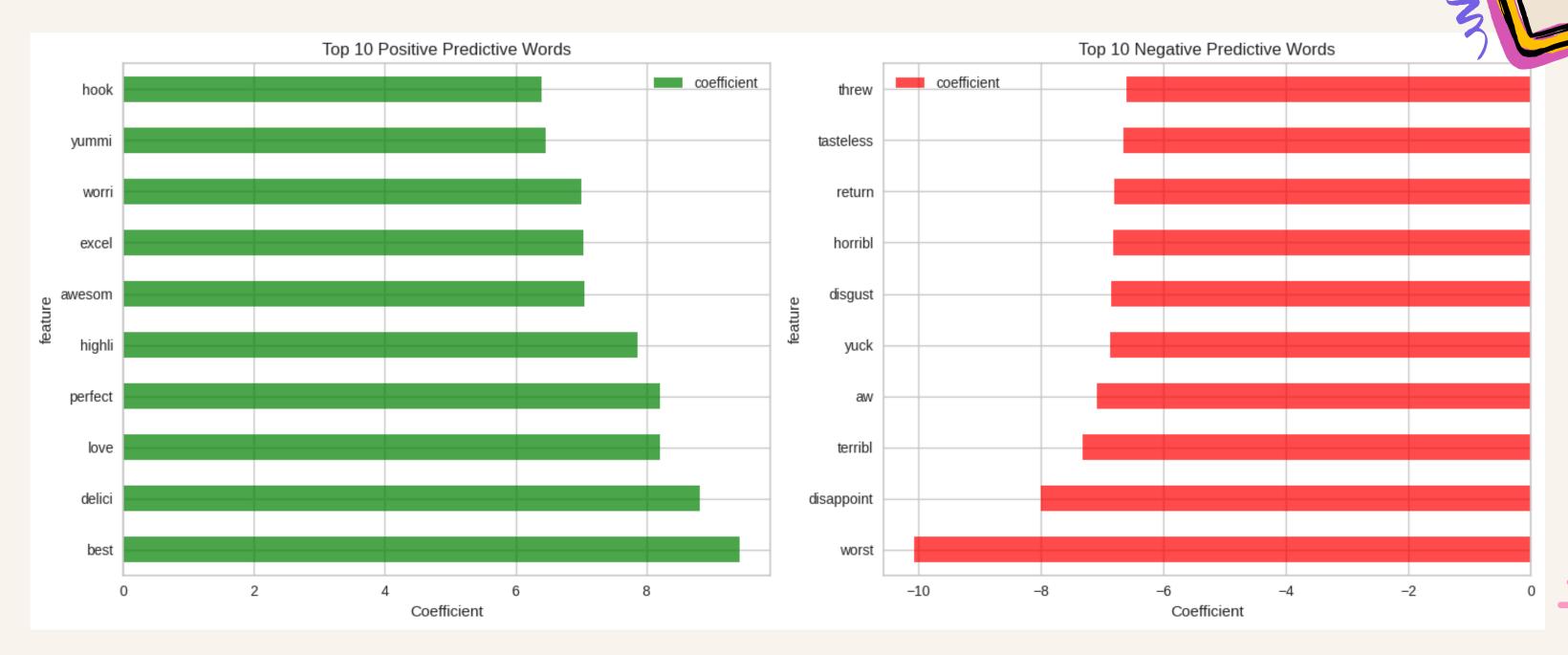
	precision	recall	f1-score	support
0 1	0.78 0.93	0.56 0.97	0.65 0.95	16407 97284
accuracy macro avg weighted avg	0.85 0.91	0.76 0.91	0.91 0.80 0.91	113691 113691 113691





Feature importance

Top fitur yang paling berpengaruh dalam memprediksi sentimen





DASHBOARD: MICROSOFT BI

