



Petualangan
Statistika

Analisis Sentimen

Ulasan Pelanggan terhadap 5 Brands Elektronik Rumah Tangga Lokal di Tokopedia

by : Fachrul Alam, A.Md.



Latar Belakang

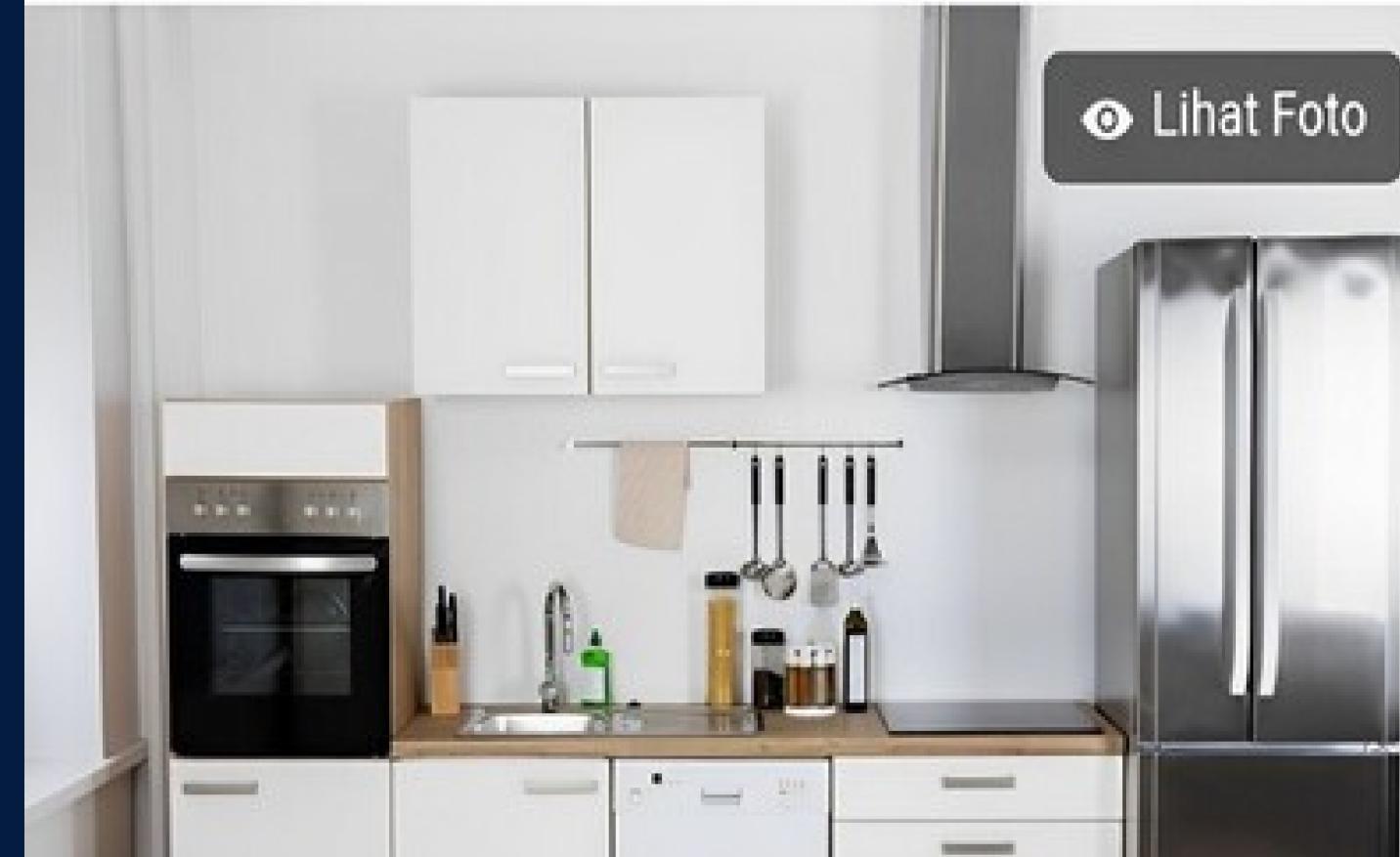
Dengan pertumbuhan teknologi dan peningkatan kebutuhan akan kenyamanan dalam kehidupan sehari-hari, permintaan akan barang elektronik rumah tangga terus meningkat. Di Indonesia, *brand* terkenal, seperti Cosmos, Maspion, Miyako, Polytron, dan Sanken telah menjadi pilihan utama bagi konsumen. Untuk memahami persepsi dan respons konsumen terhadap merek-merek ini, dilakukanlah penelitian analisis sentimen terhadap ulasan produk dari *platform e-commerce*, Tokopedia.

Penelitian ini menggunakan metode ekstraksi fitur *Bag of Words* (BoW) untuk menganalisis ulasan produk, di mana setiap ulasan direpresentasikan sebagai vektor berdasarkan frekuensi kata-kata dalam ulasan tersebut. Selanjutnya, analisis sentimen dilakukan dengan menggunakan algoritma klasifikasi XGBoost dan Naive Bayes. XGBoost sebagai algoritma *ensemble* yang kuat dan Naive Bayes sebagai algoritma klasifikasi yang sederhana namun efektif dalam analisis teks, diharapkan dapat memberikan wawasan berharga bagi produsen dalam meningkatkan kualitas produk dan memenuhi harapan konsumen dengan lebih baik.

Jadi Tuan di Negeri Sendiri, Ini Brand Lokal Elektronik Rumah Tangga yang Wajib Diketahui

Kompas.com, 21 Agustus 2020, 19:31 WIB

Baca di App



Lihat Foto

sumber: kompas.com

Metodologi Penelitian

01

Identifikasi Masalah

02

Persiapan Data

03

Preprocessing

04

Ekstraksi Fitur &
Algoritma Klasifikasi

05

Hasil & Analisis

06

Simpulan & Saran



01/ Identifikasi Masalah

Dalam penelitian ini ingin mengetahui ulasan masyarakat terhadap pembelian produk elektronik lokal melalui Tokopedia dan Bagaimana mengklasifikasikan ulasan tersebut menggunakan metode *Bag of Words* (BoW) dengan algoritma XGBoost dan Naive Bayes?

02/ Persiapan Data

Di bawah ini merupakan data gabungan ulasan produk *brand* (Cosmos, Maspion, Miyako, Polytron, dan Sanken) yang diambil melalui *web scraping* di *platform* Tokopedia. Pengambilan data ulasan terbaru dilaksanakan pada tanggal 19 Februari 2024.

| Nama Pengguna | Produk | Rating | Ulasan | Tanggal |
|---------------|---|-----------|--|-----------------|
| h***o | POLYTRON Mesin Cuci 2 Tabung Giant Series Hijab 8 KG PWM 8076 | bintang 5 | | mantap Hari ini |
| hapid | POLYTRON SMART DIGITAL TV 32" PLD 32CV1869 | bintang 5 | barang original, penerimaan lumbayan cepat. | Hari ini |
| a***l | POLYTRON Kulkas 1 Pintu Belleza 180 Liter PRA 18J | bintang 5 | very good respond, normal delivery, good quality | Hari ini |
| T***a | POLYTRON Professional Active Speaker PAS PRO12F6 | bintang 5 | barang OK, sistem pengirimannya yg perlu di perhatikan | Hari ini |
| A***a | POLYTRON Bluetooth Speaker PMA 9321 /B | bintang 5 | sudah sampai dan packing aman dan cepat . ongkir nya juga gratis. mantap | Hari ini |



Terdiri atas 5 kolom dan 2461 observasi (baris). Dalam hal ini difokuskan kepada pemrosesan kolom ulasan dan rating.

03 / Preprocessing

a. Data Clean & Case Folding

| | |
|----------|--|
| Ulasan | Bisa muat banyak Barangnya original  Gampang dipakai Hemat listrik  Desainnya elegan  Packaging aman  |
| Cleaning | bisa muat banyak barangnya original  gampang dipakai hemat listrik  desainnya elegan  packaging aman |

Dalam tahap ini, data akan dibersihkan dengan menghapus data duplikat dan *null*. Langkah selanjutnya adalah mengubah semua teks menjadi huruf kecil.

b. Remove Punctuation

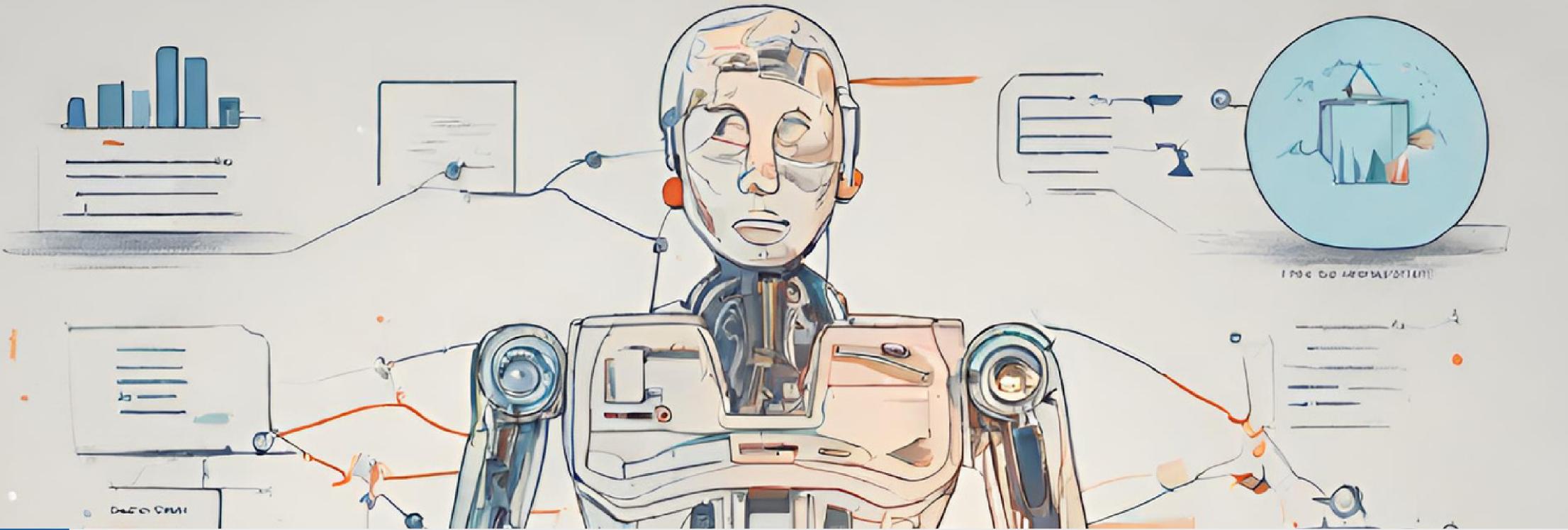
| | |
|----------|--|
| Cleaning | bisa muat banyak barangnya original  gampang dipakai hemat listrik  desainnya elegan  packaging aman |
| Cleaning | bisa muat banyak barangnya original gampang dipakai hemat listrik desainnya elegan packaging aman |

Setelah itu, dilakukan tahap penghapusan tanda baca untuk menghilangkan simbol-simbol yang tidak diperlukan dari teks.

c. Tokenizing

| Cleaning | Tokenizing |
|--|--|
| bisa muat banyak barangnya original gampang dipakai hemat listrik desainnya elegan packaging aman | [bisa, muat, banyak, barangnya, original, gampang, dipakai, hemat, listrik, desainnya, elegan, packaging, aman] |

Tokenisasi teks digunakan untuk memecah teks menjadi token-token yang lebih kecil, seperti kata-kata atau frasa-frasa.



d. Normalisasi

| Tokenizing | Normalisasi |
|--|---|
| <code>['beli', 'speaker', 'aktif', 'ya', 'yg', 'jelas', 'jelas', 'aja', 'lah', 'polytron', 'the', 'winning', 'theme']</code> | <code>['beli', 'speaker', 'aktif', 'ya', 'yang', 'jelas', 'jelas', 'saja', 'lah', 'polytron', 'the', 'winning', 'theme']</code> |

Pada tahap normalisasi, dilakukan perbaikan terhadap kata-kata singkatan agar dapat dipahami dengan lebih baik dalam konteks yang lebih luas.

03 / Preprocessing

f. Stopword

| Normalisasi | Stopword |
|---|--|
| <code>['beli', 'speaker', 'aktif', 'ya', 'yang', 'jelas', 'jelas', 'saja', 'lah', 'polytron', 'the', 'winning', 'theme']</code> | <code>['beli', 'aktif', 'polytron']</code> |

Selanjutnya dilakukan penghapusan kata-kata yang dianggap tidak bermakna dalam analisis teks, seperti kata sambung, kata asing, dan kata-kata umum lainnya.

03 / Preprocessing

| Stopword | Stemming | text |
|--|--|----------------------------------|
| ['respon', 'membatalkan', 'semoga', 'kena', 'tokopedia'] | ['respon', 'batal', 'moga', 'kena', 'tokopedia'] | respon batal mogo kena tokopedia |

Pada tahap *stemming*, dilakukan penyederhanaan kata-kata menjadi bentuk dasarnya dengan menghapus awalan atau akhiran kata sehingga mempermudah analisis teks. Setelah itu gabungkan kembali teks.

g. Stemming

i. Labeling

Setelah teks telah dilakukan preprocessing, diberikan label sesuai dengan ratingnya. Rating 1, 2, dan 3 akan diberi label negatif (0), sedangkan rating 4 dan 5 akan diberi label positif (1).

| text | Rating | Label |
|--|--------|-------|
| barang fungsi barang masak fungsi hasil gantung beras beli bintang 5 | 5 | 1 |
| beli aktif polytron | 5 | 1 |
| barang sesuai deskripsi tidak cacat kirim tes nyala suhu dingin | 5 | 1 |
| barang bagus pakai | 5 | 1 |
| barang sesuai denenggakn barang pes enggak cacat mulus | 5 | 1 |

04/1 Ekstraksi Fitur & Klasifikasi

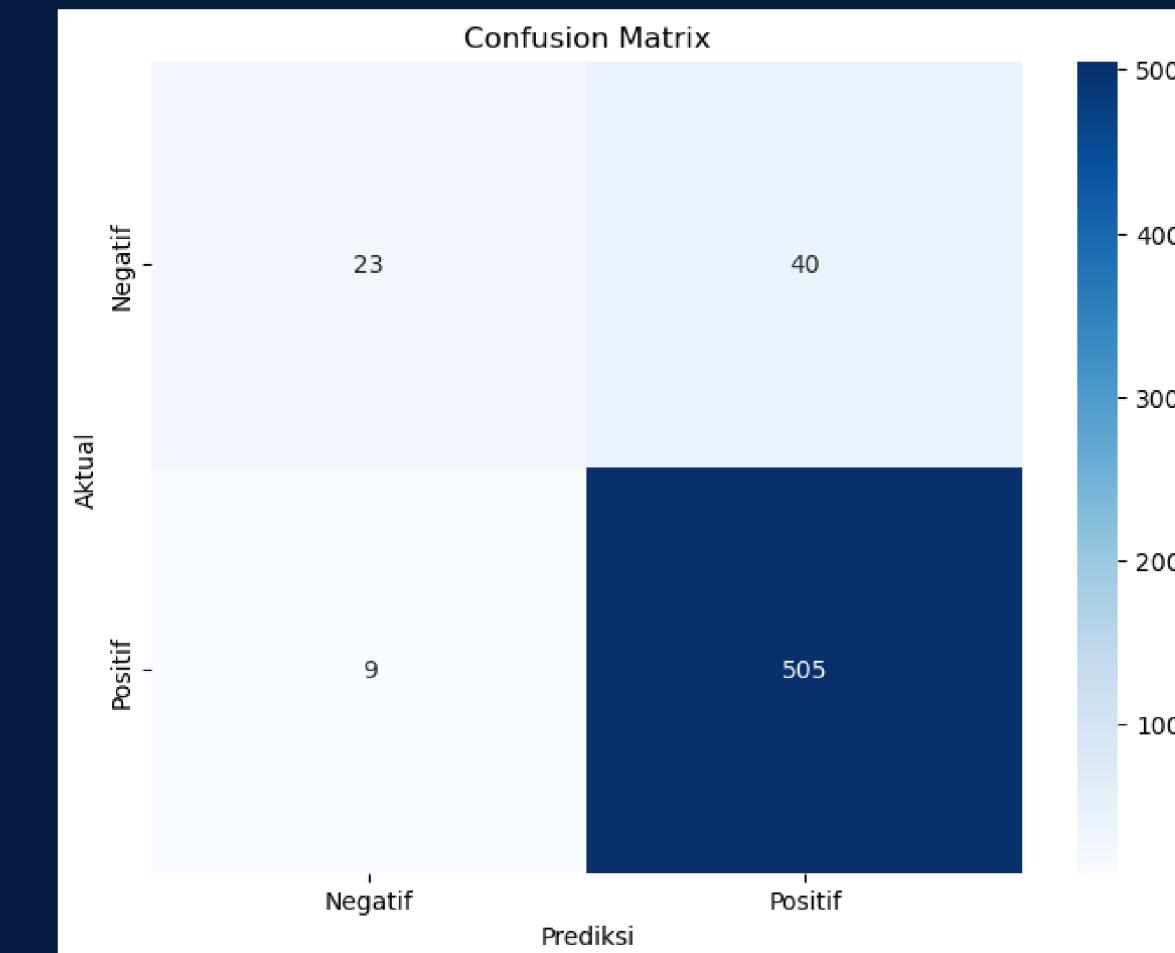
1. BoW & XGBoost

metode yang digunakan untuk ekstraksi fitur '**BoW**' dan algoritma klasifikasi dengan algoritma **XGBoost** serta proses pencarian parameter terbaik menggunakan GridSearch.

| Parameter | max_depth | 5 |
|-----------|--------------|------|
| Terbaik | n_estimators | 50 |
| | max_df | 0.5 |
| | max_features | None |
| | min_df | 1 |
| | stop_words | None |

Hasil Klasifikasi

Confusion Matrix



| Laporan Klasifikasi: | | | | |
|----------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.72 | 0.37 | 0.48 | 63 |
| 1 | 0.93 | 0.98 | 0.95 | 514 |
| accuracy | | | 0.92 | 577 |
| macro avg | 0.82 | 0.67 | 0.72 | 577 |
| weighted avg | 0.90 | 0.92 | 0.90 | 577 |

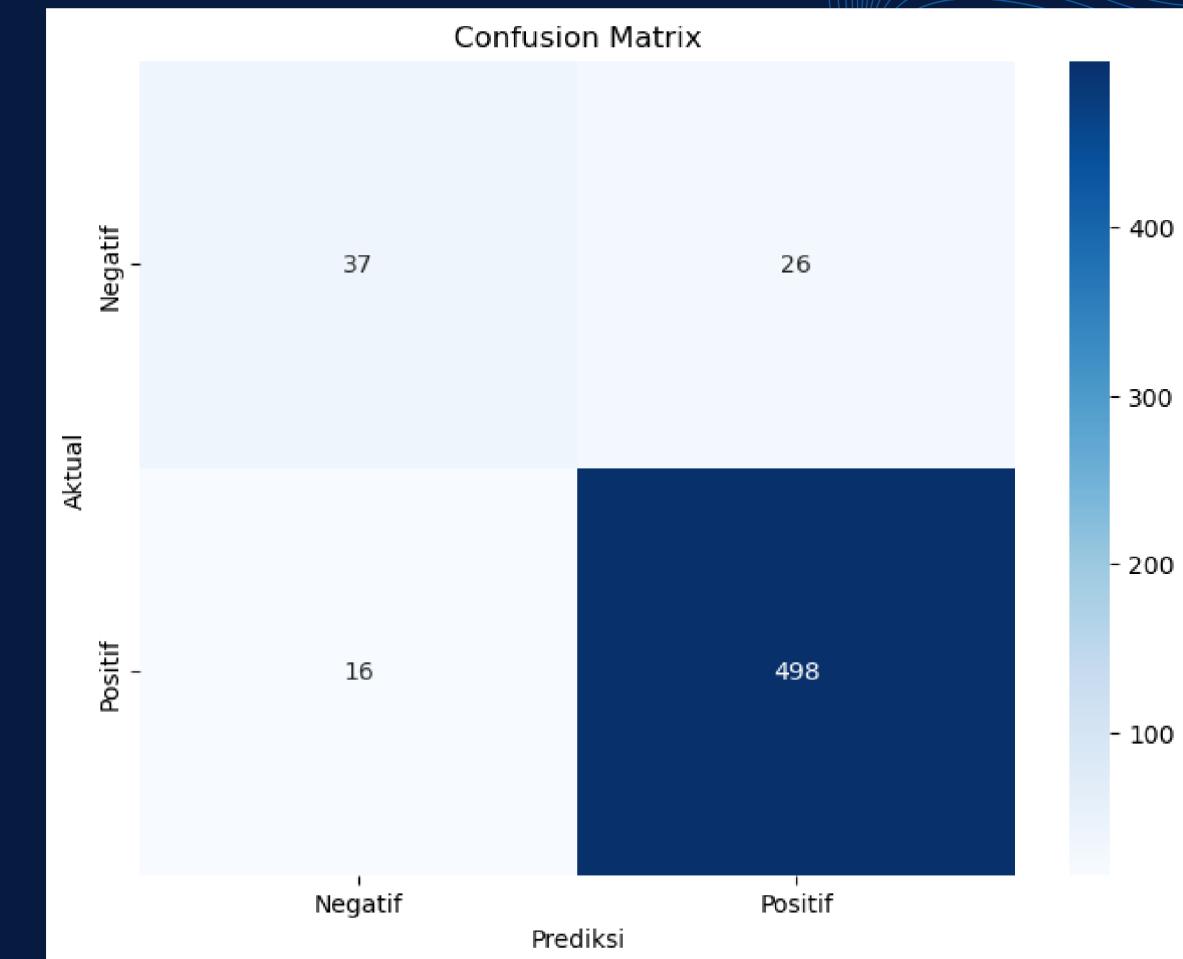
04/1 Ekstraksi Fitur & Klasifikasi

2. BoW & Naive Bayes

Naive Bayes dikenal sebagai salah satu algoritma klasifikasi yang sederhana namun mampu memberikan kinerja yang bagus dalam berbagai kasus.

| Parameter Terbaik | alpha | 0.5 |
|-------------------|--------------|------|
| | max_df | 0.5 |
| | max_features | None |
| | min_df | 2 |
| | stop_words | None |

Confusion Matrix



Hasil Klasifikasi

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.70 | 0.59 | 0.64 | 63 |
| 1 | 0.95 | 0.97 | 0.96 | 514 |
| accuracy | | | 0.93 | 577 |
| macro avg | 0.82 | 0.78 | 0.80 | 577 |
| weighted avg | 0.92 | 0.93 | 0.92 | 577 |

05/1 Hasil dan Analisis

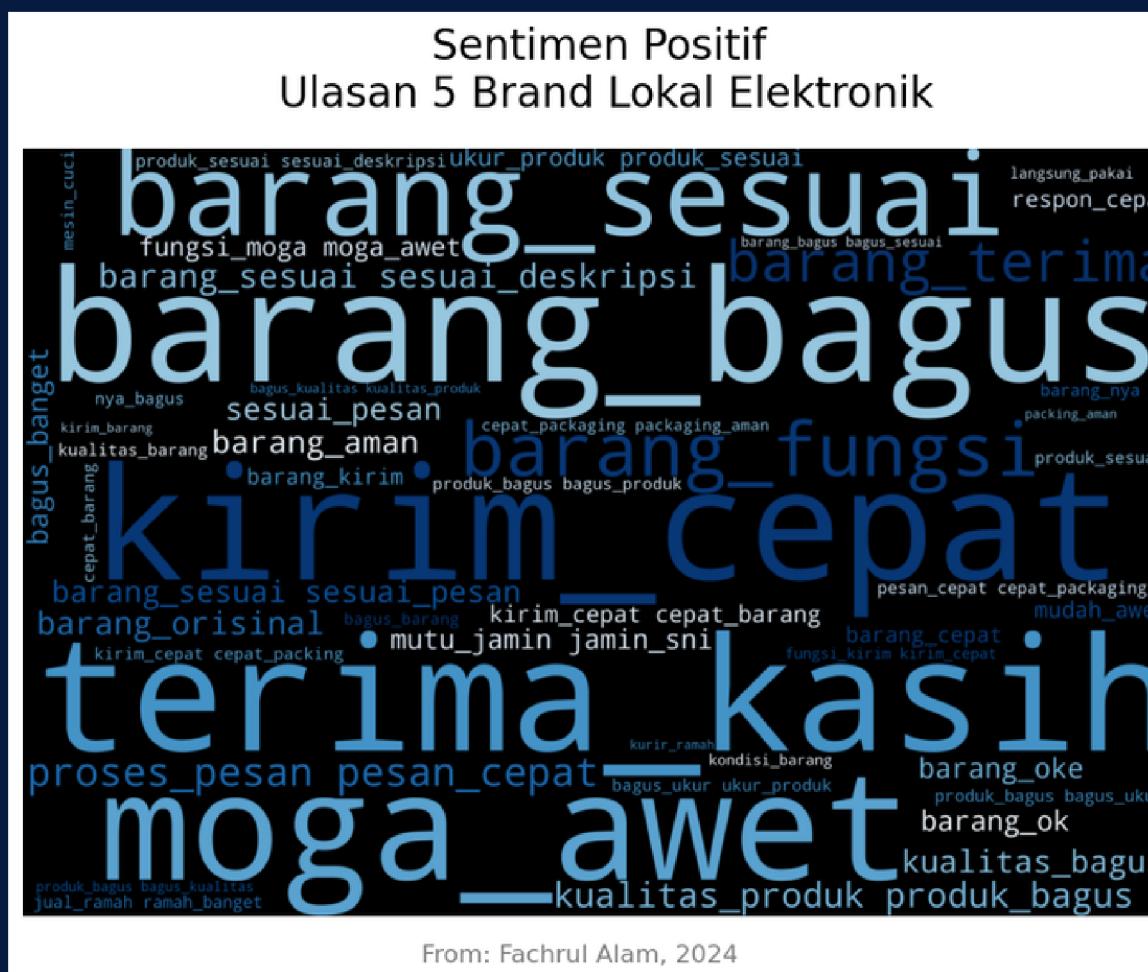
- Dalam kasus ini, meskipun akurasi kedua model (XGBoost dan Naive Bayes) berada di sekitar 92% - 93%, lebih baik fokus pada nilai *F1-score* karena label tidak seimbang.
- Nilai *F1-score* memberikan informasi yang lebih baik tentang keseimbangan antara presisi dan *recall* untuk setiap kelas.
- Perlu diperhatikan bahwa nilai *F1-score* untuk kelas 0 (sentimen negatif) lebih tinggi untuk Naive Bayes (64%) dibandingkan dengan XGBoost (48%).
- Begitupun *F1-score* untuk kelas 1 (sentimen positif) lebih tinggi untuk Naive Bayes (96%) dibandingkan dengan XGBoost (95%).
- *F1-score* label 1 (sentimen positif) tinggi karena mayoritas data termasuk dalam kelas ini, namun penting juga untuk memperhatikan *F1-score* label 0 (sentimen negatif) dalam kasus ketidakseimbangan kelas.

| Laporan Klasifikasi XGBoost: | | | | |
|------------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.72 | 0.37 | 0.48 | 63 |
| 1 | 0.93 | 0.98 | 0.95 | 514 |
| accuracy | | | 0.92 | 577 |
| macro avg | 0.82 | 0.67 | 0.72 | 577 |
| weighted avg | 0.90 | 0.92 | 0.90 | 577 |

| Laporan Klasifikasi Naive Bayes: | | | | |
|----------------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.70 | 0.59 | 0.64 | 63 |
| 1 | 0.95 | 0.97 | 0.96 | 514 |
| accuracy | | | 0.93 | 577 |
| macro avg | 0.82 | 0.78 | 0.80 | 577 |
| weighted avg | 0.92 | 0.93 | 0.92 | 577 |

05/2 Word Cloud

Setelah itu dilakukan visualisasi dengan menggunakan *word cloud* untuk memperlihatkan kata-kata yang sering muncul dalam ulasan, baik sentimen positif maupun negatif.



Setimen Positif

Setimen Positif

Word Cloud menunjukkan bahwa ulasan positif cenderung mencakup kata-kata, seperti "barang bagus", "sesuai", "kirim cepat", dan "terima kasih", menunjukkan kepuasan pelanggan terhadap kualitas produk, kecocokan dengan deskripsi, serta pengiriman yang cepat. Ini menunjukkan bahwa mayoritas pelanggan merasa puas dengan produk dan pengalaman berbelanja mereka.

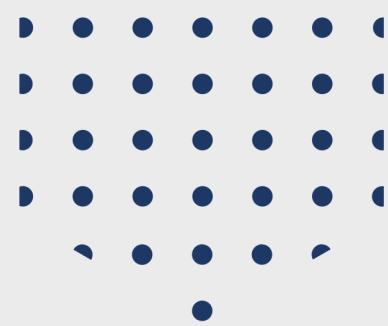
05/2 Word Cloud

Sentimen Negatif

Word Cloud juga menunjukkan bahwa ulasan negatif sering kali mengandung kata-kata, seperti "tidak", "rusak", "pecah", dan lainnya, menunjukkan ketidakpuasan pelanggan terhadap kondisi barang yang diterima, seperti kerusakan atau pecah. Hal ini mengindikasikan bahwa terdapat beberapa masalah dengan kualitas produk yang mempengaruhi pengalaman berbelanja pelanggan secara negatif.



Sentimen Negatif



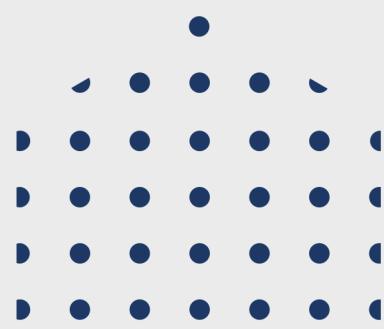
06/ Simpulan & Saran

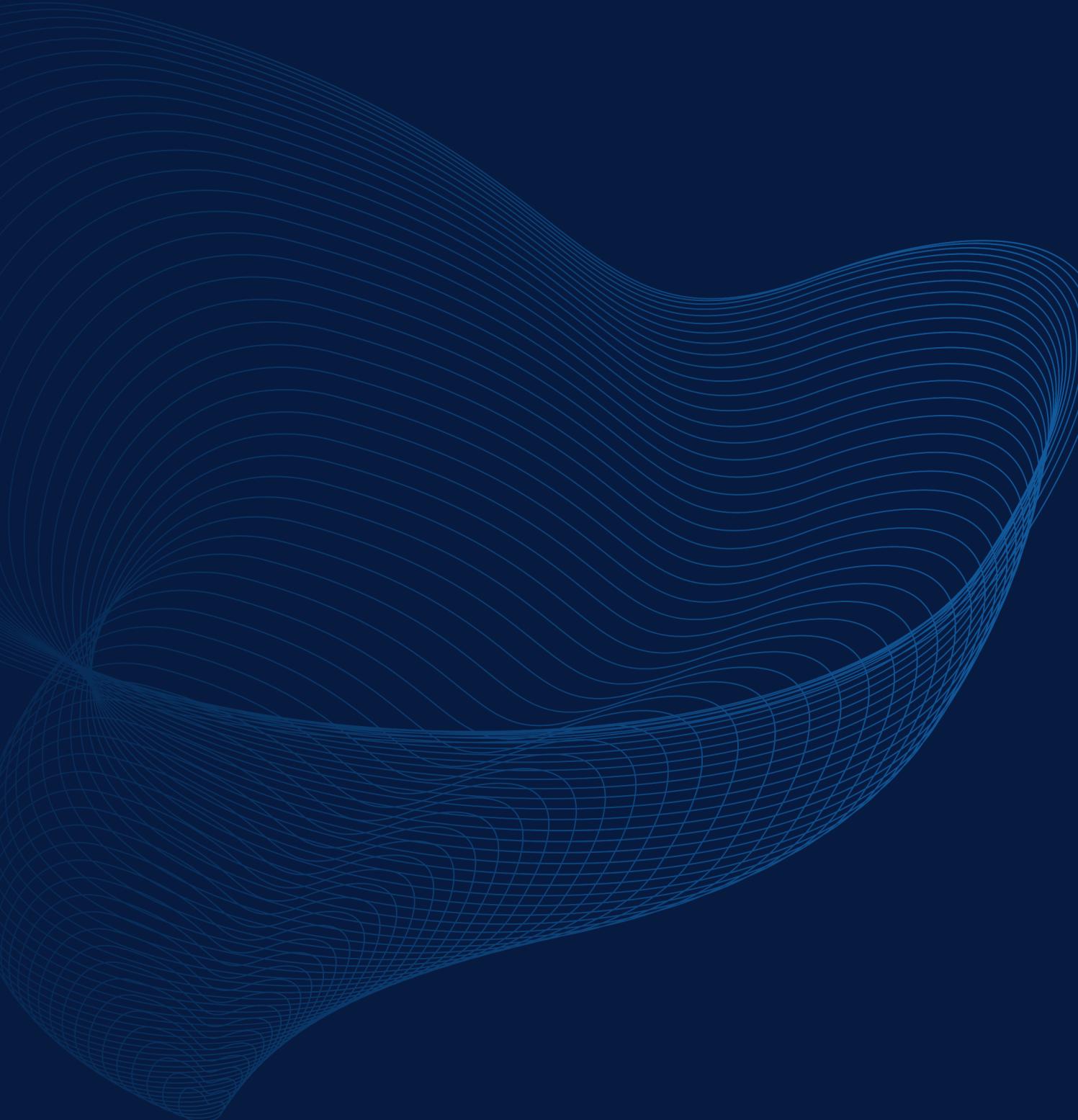
Simpulan

- Data diperoleh dari platform Tokopedia melalui teknik web scraping.
- Ekstraksi fitur dilakukan menggunakan metode Bag of Words (BoW).
- Model dibangun menggunakan algoritma XGBoost dan Naive Bayes.
- Algoritma Naive Bayes menunjukkan performa yang lebih baik dalam memprediksi sentimen produk berdasarkan ulasan pengguna.
- Kedua algoritma masih kurang baik dalam memprediksi kelas 0 (sentimen negatif) karena ketidakseimbangan label.

Saran

- Melakukan peningkatan pada pemrosesan data untuk menangani ketidakseimbangan label pada kelas sentimen.
- Melakukan eksplorasi lebih lanjut terhadap teknik ekstraksi fitur yang dapat meningkatkan performa model.
- Memperluas cakupan data ulasan produk dari berbagai *platform e-commerce*.
- Meskipun mayoritas ulasan positif, perlu perhatikan ulasan negatif. Brand perlu konsisten tingkatkan kualitas produk dan layanan untuk memenuhi harapan pelanggan.





Terima Kasih

Email Adress

fachrulalam68@gmail.com