

Analisis Sentimen Ulasan Produk Tokopedia dengan IBM Granite

Capstone Project

oleh Raihan Andi Saungnaga

Latar Belakang

E-commerce seperti Tokopedia menghasilkan ribuan ulasan dari pelanggan setiap hari. Ulasan ini penting karena bisa menjadi indikator kualitas produk dan layanan, tetapi jumlahnya terlalu banyak untuk dibaca manual. Seller kesulitan memahami pola keluhan maupun pujian dari konsumen. Dengan bantuan kecerdasan buatan, proses analisis ini dapat diotomatisasi sehingga insight dapat diperoleh dengan cepat.

Tujuan

Proyek ini bertujuan untuk membangun sistem analisis sentimen otomatis. Ulasan pelanggan diklasifikasikan menjadi tiga kategori: positif, netral, dan negatif. Selain itu, proyek ini membandingkan performa model tradisional (TF-IDF + Logistic Regression) dengan model bahasa besar IBM Granite yang diakses melalui Replicate API. Tujuan akhirnya adalah menghasilkan insight dan rekomendasi praktis yang bisa membantu seller meningkatkan layanan.

Dataset

Dataset yang digunakan berasal dari Kaggle dengan judul Tokopedia Product Reviews 2019. Total terdapat 40.607 ulasan.

Kolom yang tersedia antara lain teks ulasan, rating bintang, kategori produk, dan nama produk. Label sentimen dibuat dari rating: bintang 1–2 sebagai negatif, bintang 3 sebagai netral, dan bintang 4–5 sebagai positif. Dataset ini cukup besar untuk digunakan melatih model serta menguji performa Granite.

Metodologi

Langkah analisis dimulai dengan EDA (Exploratory Data Analysis) untuk memahami karakteristik data. Selanjutnya dibuat model baseline menggunakan TF-IDF dan Logistic Regression untuk klasifikasi sentimen. Setelah itu dilakukan integrasi dengan IBM Granite melalui Replicate API untuk melakukan klasifikasi zero-shot dan menghasilkan ringkasan. Evaluasi dilakukan dengan menghitung F1-score serta melihat confusion matrix.

Hasil EDA

Distribusi rating menunjukkan mayoritas review memiliki nilai lima bintang, sehingga dataset sangat condong ke sentimen positif. Panjang rata-rata teks ulasan adalah 55 karakter dengan variasi antara 5 sampai lebih dari 1000 karakter. Kategori produk yang paling banyak muncul adalah elektronik, fashion, dan olahraga. Hal ini menunjukkan bahwa mayoritas ulasan Tokopedia pada tahun tersebut berasal dari kategori produk populer.

Baseline Model

Model baseline menggunakan TF-IDF untuk mengubah teks menjadi vektor, kemudian Logistic Regression untuk melakukan klasifikasi. Evaluasi menunjukkan nilai Macro-F1 sekitar 0.72, yang berarti performanya cukup seimbang di tiga kelas. Confusion matrix memperlihatkan model dapat mengenali ulasan positif dengan baik, serta masih cukup mampu membedakan ulasan netral dan negatif meski jumlah datanya lebih sedikit.

Granite API

IBM Granite diakses melalui Replicate API dengan teknik prompt few-shot. Model diarahkan untuk hanya mengeluarkan label “neg”, “neu”, atau “pos”. Hasil pengujian pada 1000 sampel review menunjukkan nilai Macro-F1 sekitar 0.44. Model ini cenderung over-predict kelas netral dan sering salah menebak ulasan positif. Meski begitu, Granite memberikan kelebihan berupa fleksibilitas dan kemampuan menghasilkan insight yang lebih kaya dibanding baseline.

Summary Insight

Selain klasifikasi, Granite digunakan untuk membuat ringkasan dari kumpulan ulasan. Pada kategori elektronik, ringkasan menunjukkan bahwa keluhan utama adalah pengiriman lambat, kemasan rusak, dan produk tidak sesuai deskripsi. Sementara itu, hal positif yang paling sering disebut adalah kualitas produk yang baik, harga sesuai, dan pengiriman cepat. Rekomendasi yang dihasilkan adalah memperbaiki sistem logistik dan packaging agar kepuasan pelanggan meningkat.

Perbandingan

Model baseline lebih unggul dalam akurasi karena dilatih khusus untuk dataset Tokopedia. Sebaliknya, Granite lebih lemah dalam klasifikasi, tetapi punya kelebihan karena mampu melakukan summarization dan memberi insight tambahan yang tidak bisa dilakukan baseline. Dengan kata lain, baseline cocok untuk tugas klasifikasi cepat, sedangkan Granite cocok untuk memberikan analisis kualitatif yang lebih mendalam.

Kesimpulan

Ulasan Tokopedia didominasi oleh sentimen positif. Model baseline memberikan hasil yang lebih stabil dan akurat untuk klasifikasi. IBM Granite, walaupun kurang akurat, menambah nilai karena dapat menghasilkan insight ringkas mengenai keluhan, pujian, dan rekomendasi. Kombinasi keduanya dapat menjadi solusi yang ideal: baseline untuk klasifikasi massal, Granite untuk ringkasan dan insight strategis.

Penutup

Terima kasih.

Analisis ini membuktikan bahwa integrasi machine learning tradisional dengan Large Language Model seperti IBM Granite dapat membantu memahami ulasan konsumen secara lebih cepat dan mendalam.