

Laporan Search Engine Sentimen untuk Klasifikasi Keluhan dan Saran Pelanggan pada Rumah Makan Bebek Pak Slamet



Disusun oleh :

Nama: Zekki Maulana Rahman

NIM : A11.2023.14989

FAKULTAS ILMU KOMPUTER

**UNIVERITAS DIAN NUSWANTORO
2024**

Latar Belakang

Dalam industri kuliner yang kompetitif, ulasan pelanggan di platform digital menjadi aset krusial. Namun, banyaknya jumlah komentar menyulitkan pemilik usaha untuk memetakan ulasan mana yang bersifat pujian (positif), keluhan (negatif), maupun saran perbaikan. Dibutuhkan sebuah sistem yang tidak hanya dapat mencari kata kunci, tetapi juga mampu mengelompokkan sentimen secara otomatis untuk membantu pengambilan keputusan bisnis.

Sistem ini penting untuk membantu pihak manajemen melakukan *filtering* terhadap ratusan file ulasan (.txt) guna menemukan masalah spesifik (seperti harga atau pelayanan) sekaligus mengetahui apakah masalah tersebut berujung pada sentimen negatif atau sekadar saran perbaikan.

Masalah dan Objektif

Masalah

Masalah utama yang diidentifikasi adalah ketidakmampuan sistem pencarian teks konvensional dalam memahami polaritas emosi di balik sebuah kata kunci, yang sering kali menyebabkan hasil pencarian menjadi kurang relevan secara kontekstual. Selain itu, penggunaan *stopword removal* standar sering kali menghilangkan kata-kata kunci sentimen yang justru membawa informasi krusial. Oleh karena itu, riset ini bertujuan untuk membangun sebuah *pipeline* sistem temu kembali informasi yang mampu melakukan klasifikasi otomatis ke dalam kategori POSITIF, NEGATIF, dan SARAN. Objektif teknisnya mencakup implementasi perhitungan relevansi hibrida yang menggabungkan kemiripan leksikal dan keselarasan sentimen.

Tujuan Riset

- Membangun sistem pencarian (Search Engine) yang mampu mengurutkan hasil berdasarkan relevansi teks dan kesamaan sentimen.
- Mengimplementasikan teknik NLP (Preprocessing) untuk menangkap kata-kata kunci sentimen dalam bahasa Indonesia.
- Mengukur relevansi dokumen menggunakan metode *Hybrid Scoring* (Jaccard + Cosine Similarity).

Dataset

Sumber Data:

Data ulasan pelanggan Rumah Makan Bebek Pak Slamet dilaman media instragram. Saya mengambil data berdasarkan 1 bulan terahir sehingga kritik dan saran bisa berdampak langsung terhadap kemajuan Rumah Makan Bebek Pak Slamet

Format

File .txt yang disimpan dalam direktori Data/Raw/.

Kamus

Sistem menggunakan pendekatan berbasis kamus (lexicon-based) yang terbagi menjadi:

- Kamus_Positif (46 kata): Mencakup aspek rasa (enak, lezat), pelayanan (ramah), dan harga (murah).
- Kamus_Negatif (54 kata): Mencakup aspek waktu (lama, antri), harga (mahal), dan emosi (kecewa).
- Kamus_Saran (46 kata): Mencakup kata kerja usulan (perlu, harus, perbaiki) dan hasil stemming terkait.

Metodologi

Preprocessing

Sistem ini mengadopsi arsitektur Information Retrieval (IR) yang dimulai dengan tahap preprocessing intensif. Proses diawali dengan case folding untuk menyeragamkan teks, diikuti oleh tokenizing menggunakan Regular Expression untuk menghapus karakter non-alfabet. Tahap paling krusial adalah Custom Stopword Removal, di mana sistem secara cerdas mempertahankan kata-kata yang ada dalam kamus sentimen meskipun kata tersebut masuk dalam daftar stopword umum. Selanjutnya, library Sastrawi digunakan untuk proses stemming guna mereduksi kata berimbuhan menjadi kata dasar tanpa menghilangkan makna aslinya.

Perhitungan Skor

Menghitung Term Frequency (TF) untuk menentukan skor sentimen dokumen. Untuk mekanisme pencarian, sistem menerapkan Hybrid Scoring dengan bobot 70% pada Jaccard Similarity untuk mengukur irisan kata secara leksikal, dan 30% pada Sentiment Alignment menggunakan Cosine Similarity untuk mengukur kesamaan profil emosional antara query dan ulasan. Implementasi ini dilakukan menggunakan framework Streamlit untuk antarmuka web, dengan optimasi performa melalui fitur caching (@st.cache_data) agar proses pengolahan data masif tetap responsif.

Search & Ranking

- Jaccard Similarity (Bobot 70%): Mengukur irisan kata antara query dan ulasan.
- Sentiment Alignment (Bobot 30%): Menggunakan *Cosine Similarity* untuk mencocokkan profil sentimen query dengan dokumen.

Hasil dan Evaluasi

Hasil

Hasil pengujian menunjukkan bahwa sistem mampu menampilkan distribusi sentimen secara visual melalui sidebar, memberikan gambaran cepat mengenai kondisi layanan rumah makan. Dalam aspek pencarian, sistem berhasil mengurutkan dokumen berdasarkan Relevance Score tertinggi, di mana ulasan yang memiliki kesesuaian kata kunci sekaligus kesesuaian sentimen akan berada di posisi teratas. Evaluasi performa secara kualitatif menunjukkan bahwa penggabungan Jaccard dan Cosine Similarity memberikan hasil yang lebih presisi dibandingkan hanya mengandalkan pencarian kata kunci tunggal, terutama dalam menyaring ulasan yang bersifat saran teknis.

Evaluasi Performa

- Relevansi: Penggunaan bobot 0.7 Jaccard dan 0.3 Cosine memberikan hasil pencarian yang lebih "peka" terhadap konteks sentimen dibandingkan pencarian kata kunci biasa.
- Kecepatan: Implementasi `@st.cache_data` pada Streamlit mempercepat proses pembacaan ulasan yang berulang.

Diskusi dan Kesimpulan

Diskusi

Sistem memiliki keunggulan pada *Custom Stopword Removal* yang menjaga integritas kata sentimen. Namun, sistem masih memiliki keterbatasan dalam menangani kalimat negasi (seperti "tidak enak") karena masih bergantung pada kemunculan kata kunci individual (Kamus).

Kesimpulan

Search engine ini berhasil mengintegrasikan teknik temu kembali informasi konvensional dengan analisis sentimen sederhana, memberikan hasil pencarian yang terperinci dan terklasifikasi untuk kebutuhan manajemen rumah makan.