

Implementasi Metode Leksikon dengan Stemming Sastrawi pada Python untuk ekstraksi opini pelanggan warung sate di Purbalingga

Aldit Pratyasto

Program Studi Bisnis Digital, Universitas Tangerang Raya, Kota Tangerang

Penulis Korespondensi : Aldit Pratyasto (e-mail: alditpra@gmail.com)

ABSTRAK

Ulasan pelanggan pada platform digital merupakan aset krusial bagi pengembangan bisnis kuliner, namun analisisnya secara manual tidak efisien seiring bertambahnya volume data. Penelitian ini bertujuan merancang dan mengimplementasikan sistem klasifikasi sentimen otomatis terhadap ulasan pelanggan warung sate di Kabupaten Purbalingga, sebuah domain kuliner lokal yang memiliki karakteristik bahasa unik dan jarang diteliti. Penelitian ini menerapkan metode Pemrosesan Bahasa Alami (NLP) dengan pendekatan berbasis leksikon terhadap 1000 ulasan yang dikumpulkan dari Google Maps. Setelah melalui tahap pembersihan data, sebanyak 915 ulasan valid dianalisis lebih lanjut. Proses NLP meliputi case folding, tokenizing, stopword removal, dan stemming menggunakan library Sastrawi, dengan kamus leksikon terdiri dari 61 kata positif dan 61 kata negatif. Evaluasi model terhadap rating bintang asli menunjukkan akurasi keseluruhan sebesar 63%. Meskipun model menunjukkan performa yang baik dalam mengidentifikasi sentimen positif (F1-score 0.77), performanya terbatas pada kelas negatif (F1-score 0.28) dan netral (F1-score 0.20). Hasil ini mengindikasikan bahwa pendekatan leksikon efektif sebagai dasar analisis opini, namun pengembangan lebih lanjut diperlukan untuk mengatasi tantangan seperti ketidakseimbangan data dan kompleksitas bahasa pada ulasan.

KATA KUNCI Klasifikasi Sentimen, Pemrosesan Bahasa Alami, Sate Purbalingga, Python, Sastrawi, Leksikon.

1. PENDAHULUAN

Di era digital, reputasi bisnis, khususnya di sektor kuliner, semakin bergantung pada opini pelanggan yang dibagikan secara daring. Platform seperti Google Maps telah menjadi referensi utama bagi calon konsumen sekaligus sumber masukan vital bagi pemilik usaha. Secara global, analisis terhadap ulasan daring telah diakui sebagai komponen penting untuk mendapatkan keunggulan kompetitif di industri restoran [1]. Volume data yang besar ini menyimpan wawasan berharga, namun analisisnya secara manual menjadi tidak praktis dan subjektif.

Oleh karena itu, diperlukan sebuah metode otomatis untuk mengekstrak inti dari opini pelanggan secara cepat dan objektif. Pemrosesan Bahasa Alami (Natural Language Processing - NLP) menawarkan solusi melalui teknik klasifikasi sentimen. Penelitian sejenis telah banyak dilakukan pada berbagai domain di Indonesia, seperti pada ulasan layanan pesan-antar makanan [2], aplikasi e-commerce [3], hingga ulasan restoran di kota-kota besar [4]. Namun, penelitian-penelitian tersebut seringkali berfokus pada layanan

berskala nasional dan belum banyak yang menyentuh ranah kuliner lokal spesifik, yang berpotensi memiliki tantangan linguistik tersendiri seperti penggunaan dialek atau istilah informal yang khas.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan menerapkan klasifikasi sentimen pada ulasan pelanggan warung sate di Purbalingga. Dengan memanfaatkan bahasa pemrograman Python dan library Sastrawi, sistem ini dirancang untuk melakukan pra-pemrosesan teks dan mengklasifikasikan sentimen menggunakan pendekatan berbasis leksikon. Hasil dari analisis ini diharapkan dapat memberikan wawasan terukur mengenai aspek yang disukai dan tidak disukai pelanggan, yang dapat menjadi landasan strategis bagi pemilik usaha untuk meningkatkan kualitas produk dan layanan.

2. BAHAN DAN METODE

Metodologi penelitian ini dirancang secara sistematis, mulai dari pengumpulan data hingga evaluasi model, yang diimplementasikan sepenuhnya melalui skrip Python.

Bahan penelitian utama adalah dataset berisi 1000 ulasan pelanggan yang dikumpulkan dari 10 warung sate populer di Purbalingga melalui platform Google Maps. Setiap data ulasan mencakup nama lokasi, rating bintang, dan teks ulasan. Setelah melalui proses pembersihan data, sebanyak 915 ulasan digunakan untuk analisis. Selain itu, digunakan pula sebuah kamus sentimen (leksikon) internal yang dibangun secara manual agar relevan dengan konteks kuliner. Langkah ini penting karena efektivitas metode leksikon sangat bergantung pada kualitas dan kekhususan domain dari kamus yang digunakan, di mana leksikon umum seringkali tidak mampu menangkap istilah spesifik dalam ulasan kuliner [5]. Leksikon ini terdiri dari 61 kata berpolaritas positif (contoh: "enak", "mantap", "empuk") dan 61 kata berpolaritas negatif (contoh: "mahal", "lama", "alot").

Peralatan penelitian ini menggunakan bahasa pemrograman Python 3.13.5 dengan beberapa library utama, yaitu Pandas untuk manipulasi data, Sastrawi untuk stemming dan stopwords removal, Scikit-learn untuk evaluasi model, serta Matplotlib, Seaborn, dan WordCloud untuk visualisasi data.

Metode penelitian ini mengikuti alur yang diilustrasikan pada Gambar 1. Tahapan dimulai dengan akuisisi dan pembersihan data, di mana dataset .xlsx dimuat dan kolom Rating dibersihkan untuk mengekstrak nilai numeriknya. Sebanyak 85 baris data tidak valid dihapus. Tahap selanjutnya adalah pra-pemrosesan teks yang krusial untuk meningkatkan kualitas analisis [6]. Proses ini meliputi case folding, cleaning (menghapus karakter non-alfabet), stopwords removal, dan stemming menggunakan Sastrawi.



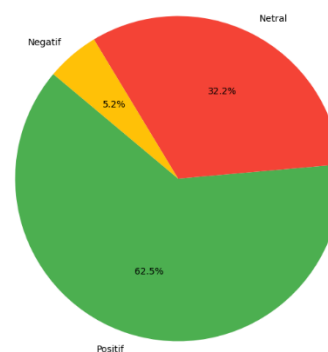
Gambar 1. Diagram Alir Penelitian

Tahap inti adalah klasifikasi sentimen berbasis leksikon, di mana setiap kata dalam ulasan dicocokkan dengan leksikon untuk dihitung skor sentimennya (+1 untuk positif, -1 untuk negatif). Agregasi skor menentukan polaritas akhir ulasan (Positif, Negatif, atau Netral). Terakhir, dilakukan evaluasi model dengan membandingkan hasil prediksi terhadap label asli yang dibuat dari rating bintang. Metrik yang digunakan adalah Accuracy, Precision, Recall, dan F1-Score.

3. HASIL DAN PEMBAHASAN

Dari 1000 ulasan awal, 915 ulasan berhasil diproses setelah melalui tahap pembersihan. Hasil klasifikasi menunjukkan bahwa mayoritas ulasan pelanggan bersifat positif. Gambar 2 secara visual menyajikan distribusi ini, di mana terlihat bahwa 62.5% ulasan diklasifikasikan sebagai positif, 32.2% sebagai netral, dan hanya 5.2% sebagai negatif.

Distribusi Sentimen Ulasan Keseluruhan



Gambar 2. Distribusi Sentimen Ulasan Keseluruhan

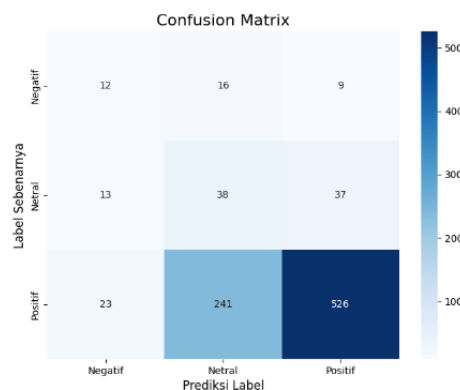
Evaluasi performa model menghasilkan akurasi keseluruhan sebesar 63%. Hasil rinci dari laporan klasifikasi disajikan pada Tabel 1. Sementara itu, Tabel 2 dan Gambar 3 menyajikan confusion matrix yang memetakan performa klasifikasi secara lebih detail. Pada visualisasi heatmap di Gambar 3, sumbu vertikal merepresentasikan label data yang sebenarnya, sedangkan sumbu horizontal merepresentasikan label hasil prediksi model.

Tabel 1. Hasil Laporan Klasifikasi Model

Kategori	Precision	Recall	F1-Score	Support
Negatif	0.25	0.32	0.28	37
Netral	0.13	0.43	0.20	88
Positif	0.92	0.67	0.77	790
Accuracy			0.63	915
Macro Avg	0.43	0.47	0.42	915
Weighted Avg	0.82	0.63	0.70	915

Tabel 2. Confusion Matrix

Aktual / Prediksi	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Aktual Negatif	12	16	9
Aktual Netral	13	38	37
Aktual Positif	23	241	526



Gambar 3. Heatmap Confusion Matrix

Akurasi keseluruhan sebesar 63% menunjukkan

Analisis kualitatif pada visualisasi berikutnya memberikan wawasan bisnis yang lebih mendalam. Gambar 4 menyajikan perbandingan volume sentimen untuk setiap warung sate yang dianalisis. Selanjutnya, Gambar 5 dan Gambar 6 menampilkan word cloud yang menyoroti kata-kata yang paling sering muncul pada ulasan positif dan negatif. Pada Gambar 5, kata-kata seperti "enak", "sate", "mantap", "tempat", dan "rasa" mendominasi, menunjukkan kepuasan pelanggan berpusat pada kualitas produk dan suasana. Sebaliknya, pada Gambar 6, kata "rasa", "kurang", "mahal", "lama", dan "parkir" menjadi sorotan, mengindikasikan area perbaikan pada aspek rasa, harga, waktu layanan, dan fasilitas. Temuan ini menggarisbawahi bahwa kepuasan pelanggan tidak hanya bergantung pada produk inti (rasa), tetapi juga sangat dipengaruhi oleh pengalaman layanan (service experience).



Penelitian ini berhasil mengimplementasikan sebuah sistem klasifikasi opini untuk ulasan pelanggan warung sate di Purbalingga menggunakan pendekatan berbasis leksikon dengan Python. Sistem ini mampu memproses 915 ulasan secara otomatis dan mencapai akurasi keseluruhan sebesar 63%. Model terbukti efektif dalam mengenali ulasan positif, namun performanya masih kurang memuaskan untuk ulasan negatif dan netral, terutama disebabkan oleh ketidakseimbangan data dan keterbatasan metode leksikon dalam memahami konteks kalimat yang kompleks.

6. DAFTAR PUSTAKA

- 219

- [3] Cahyaningtyas, A., Wibawa, A. P., & Mokodenseho, U. P. (2022). Analisis Sentimen Ulasan Pengguna Aplikasi Shopee Menggunakan Metode Lexicon Based dan Support Vector Machine. *Jurnal Teknik Informatika dan Sistem Informasi (JATISI)*, 9(1), 543-554.
- [4] Simarmata, N., Tulus, & Zarlis, M. (2020). Analisis Sentimen Review Restoran Menggunakan Google Maps API dan Metode Naive Bayes. *Jurnal Media Informatika Budidarma*, 4(4), 1133-1139.
- [5] Basari, A. S. H., Hussin, B., & Anuar, F. I. (2021). Aspect-based sentiment analysis on restaurant reviews: A comparative study. *Indonesian Journal of Electrical Engineering and Computer Science*, 21(1), 488-496.
- [6] Purwitasari, F. A., Sumari, A. D. W., & Santosa, A. W. B. (2020). The Effect of Pre-processing on Sentiment Analysis of Indonesian Text using the Sastrawi Library. *IOP Conference Series: Materials Science and Engineering*, 821(1), 012033.
- [7] Hidayatullah, A. F., & Ma'arif, M. (2021). Perbandingan Metode Naive Bayes dan Lexicon Based untuk Analisis Sentimen Ulasan Aplikasi Ruangguru. *Jurnal Repositor*, 3(5), 625-632.
- [8] Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., & Pokorný, J. (2021). A comprehensive survey on sentiment analysis for restaurant reviews. *IEEE Access*, 9, 79933-79959.
- [9] Cambria, E. (2022). Sentic computing: A new paradigm for AI. *Artificial Intelligence Review*, 55(2), 1483-1503.
- [10] Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2020). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 607-612).
- [11] Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., ... & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing* (pp. 843-857).
- [12] Kurniawan, K., Putri, S. A. W. S., & Sari, D. A. P. (2023). Aspect-Based Sentiment Analysis for Indonesian Hotel Reviews Using IndoBERT. *Jurnal Ilmu Komputer dan Informasi (JIKI)*, 16(1), 69-78.
- [13] Koto, F., & Koto, I. Y. (2020). InSet: A new dataset for Indonesian sentiment analysis. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 5865-5871).