

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang semakin pesat telah mendorong perubahan besar dalam berbagai bidang kehidupan, salah satunya dalam sektor layanan publik dan bisnis. Dengan meningkatnya penggunaan internet dan media digital, masyarakat kini lebih mudah dalam memberikan tanggapan atau ulasan terhadap suatu produk maupun layanan yang telah mereka gunakan. Salah satu platform yang banyak digunakan oleh masyarakat untuk menyampaikan opini dan pengalaman mereka adalah Google Maps, di mana pengguna dapat memberikan ulasan (review) terhadap tempat seperti restoran, hotel, dan berbagai jenis layanan lainnya [1].

Ulasan pelanggan yang terdapat pada Google Maps memiliki nilai informasi yang sangat penting, terutama bagi pelaku usaha di bidang kuliner. Melalui analisis terhadap ulasan tersebut, pemilik restoran dapat memahami tingkat kepuasan pelanggan, kualitas pelayanan, cita rasa makanan, serta menemukan aspek yang perlu ditingkatkan. Namun, jumlah ulasan yang besar dan beragam menjadikan proses analisis secara manual menjadi sulit dan memakan waktu. Oleh karena itu, diperlukan suatu pendekatan otomatis yang dapat mengklasifikasikan ulasan pelanggan secara efisien berdasarkan sentimen positif dan negatif [2].

Salah satu pendekatan yang banyak digunakan dalam analisis sentimen (sentiment analysis) adalah machine learning, yang memanfaatkan algoritma pembelajaran untuk mengenali pola dalam teks dan menentukan maknanya. Di antara berbagai algoritma yang tersedia, Logistic Regression menjadi salah satu metode yang sederhana namun efektif untuk menyelesaikan permasalahan klasifikasi biner, seperti membedakan antara sentimen positif dan negatif [3]. Metode ini bekerja dengan menghitung probabilitas dari suatu data termasuk ke dalam kelas tertentu, berdasarkan fitur-fitur yang diekstraksi dari teks ulasan.

Berbagai penelitian sebelumnya menunjukkan bahwa Logistic Regression

mampu memberikan hasil yang akurat dalam analisis sentimen teks, terutama ketika dikombinasikan dengan metode representasi teks seperti TF-IDF (Term Frequency–Inverse Document Frequency) [4]. Selain itu, penggunaan teknik resampling, seperti SMOTE (Synthetic Minority Over-sampling Technique), dapat membantu meningkatkan kinerja model dengan menyeimbangkan distribusi data yang tidak seimbang antara ulasan positif dan negatif [5].

Dengan demikian, penelitian ini dilakukan untuk menganalisis dan mengimplementasikan algoritma Logistic Regression dalam mengklasifikasikan sentimen pelanggan terhadap ulasan restoran di Google Maps. Melalui penelitian ini, diharapkan dapat diperoleh model klasifikasi yang akurat, yang nantinya dapat membantu pelaku usaha memahami persepsi pelanggan secara lebih efektif dan mengambil keputusan bisnis yang lebih tepat.

1.2 Rumusan Masalah

Berdasarkan latar belakang muncul rumusan masalah d bawah ini:

1. Bagaimana kinerja algoritma Logistic Regression dalam mengklasifikasikan sentimen pelanggan pada ulasan restoran di Google Maps?
2. Bagaimana pengaruh penerapan teknik SMOTE (Over-sampling) terhadap peningkatan akurasi model klasifikasi sentimen?
3. Bagaimana perbandingan hasil kinerja Logistic Regression dengan model lain seperti Support Vector Machine (SVM) dan Neural Network (MLP) dalam hal akurasi, presisi, recall, dan f1-score?
4. Bagaimana penerapan GridSearchCV dapat mengoptimalkan parameter pada model Logistic Regression dan meningkatkan performa klasifikasi sentimen?

1.3 Batasan Masalah

Berdasarkan penelitian, muncul Batasan masalah d bawah ini:

1. Dataset yang digunakan merupakan ulasan restoran pada platform Google Maps yang berbahasa Indonesia, dengan label sentimen positif dan negatif.
2. Penelitian ini hanya menggunakan algoritma Logistic Regression, Support Vector Machine (SVM), dan Neural Network (MLP) sebagai pembandingan.

3. Representasi teks yang digunakan adalah metode TF-IDF (Term Frequency–Inverse Document Frequency).
4. Teknik SMOTE (Synthetic Minority Over-sampling Technique) digunakan untuk mengatasi ketidakseimbangan data antara kelas positif dan negatif.
5. Evaluasi model dilakukan dengan menggunakan metrik Accuracy, Precision, Recall, dan F1-Score.
6. Data ini Saya ambil dari Kaggle.com di link d bawah ini.
<https://www.kaggle.com/code/saiashwin2001/notebook15dc78c50f/input>

1.4 Tujuan Penelitian

Adapun tujuan penelitian , tujuan penelitian ini adalah:

1. Menganalisis performa algoritma Logistic Regression dalam mengklasifikasikan sentimen pelanggan pada ulasan restoran di Google Maps.
2. Mengevaluasi pengaruh teknik SMOTE (Over-sampling) terhadap peningkatan kinerja model klasifikasi.
3. Membandingkan hasil klasifikasi antara Logistic Regression dengan algoritma lain seperti SVM dan Neural Network (MLP).
4. Mengoptimalkan performa model dengan menggunakan GridSearchCV untuk mendapatkan parameter terbaik dari masing-masing algoritma.

1.5 Manfaat Penelitian

Manfaat penelitian yang ingin saya capai adalah:

1. Bagi Akademisi, penelitian ini dapat menjadi referensi dalam pengembangan penelitian di bidang Natural Language Processing (NLP) dan Machine Learning khususnya pada analisis sentimen teks.
2. Bagi Pelaku Bisnis Restoran, hasil penelitian ini dapat membantu dalam memahami persepsi pelanggan terhadap layanan yang diberikan sehingga dapat meningkatkan kualitas pelayanan dan strategi pemasaran.
3. Bagi Pengembang Sistem, penelitian ini dapat menjadi dasar untuk membangun sistem analisis sentimen otomatis berbasis data ulasan

pelanggan.

1.6 Sistematika Penulisan

Berisi sistematika penulisan skripsi yang memuat uraian secara garis besar isi skripsi untuk tiap-tiap bab. Peneliti harus dapat mendeskripsikan (menggambarkan) apa saja isi masing-masing Bab yang akan disusun. Jelaskan secara singkat isi dari bab I, bab II, bab III, bab IV, dan bab V.

BAB I PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Membahas teori-teori yang mendukung penelitian seperti analisis sentimen, representasi teks (TF-IDF), dan algoritma klasifikasi (Logistic Regression, SVM, MLP), serta penelitian terdahulu yang relevan.

BAB III METODE PENELITIAN

Menjelaskan alur penelitian, pengumpulan data, preprocessing, pembagian dataset, implementasi model, dan metode evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil implementasi, pengujian, dan analisis kinerja model klasifikasi sentimen.

BAB V PENUTUP

Menyimpulkan hasil penelitian serta memberikan saran untuk penelitian selanjutnya.