

1. Pendahuluan

Latar Belakang

Di era digitalisasi, internet menjadi sesuatu yang tidak dapat dipisahkan dari kehidupan manusia. Munculnya berbagai perkembangan teknologi, salah satunya adalah pinjaman online. Pinjaman online adalah konsep peminjaman uang secara online untuk memudahkan transaksi [1]. Pinjaman online juga membantu pembiayaan UMKM untuk mengembangkan usaha. Namun, kemudahan yang diberikan menyebabkan pelanggaran seperti pencurian data pelanggan. Pendapat atau opini yang diungkapkan terkait pinjaman online ini beragam. Opini masyarakat terkait pinjaman online beragam, belum diketahui termasuk opini negatif atau positif. Oleh karena itu, perlu dilakukan analisis sentimen untuk membantu masyarakat mengetahui kelas sentimen dalam data tweet agar masyarakat lebih bijak dalam menggunakan platform pinjaman online. Salah satu cara untuk mengolah ulasan tersebut adalah dengan menggunakan analisis sentimen, yang mengklasifikasikan opini menjadi sentimen positif atau negatif [2][3].

Berbagai penelitian telah dilakukan terkait analisis sentimen menggunakan beragam metode. Sebuah studi tentang analisis sentimen kebijakan PPKM selama pandemi COVID-19 menggunakan algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes Classifier (NBC) menunjukkan akurasi masing-masing 79,67% dan 78,86% [4]. Penelitian lain pada ulasan aplikasi MyPertamina di Google Playstore menggunakan algoritma Naïve Bayes Classifier menunjukkan akurasi 87%, presisi 86%, recall 90%, dan f1 score [5]. Sementara itu, penelitian analisis sentimen terhadap pinjaman online menggunakan algoritma Naïve Bayes Classifier menghasilkan akurasi 75% [6].

Penggunaan Logistic Regression dalam analisis sentimen di Twitter menghasilkan akurasi 78,57%, presisi 76,92%, recall 83,3%, dan F1-score 80% [7]. Perbandingan antara Logistic Regression dan K-Nearest Neighbor dalam analisis sentimen ulasan aplikasi kependudukan digital menunjukkan bahwa K-Nearest Neighbor memiliki

kinerja yang lebih baik dengan akurasi 82,43% dibandingkan Logistic Regression dengan 78,83% [8].

Penelitian lain mengenai analisis sentimen dengan tingkat kepuasan pengguna Penyedia Jasa Telekomunikasi Seluler Indonesia di Twitter menggunakan metode Support Vector Machine dan fitur berbasis leksikon menghasilkan nilai akurasi sebesar 79%, presisi 65%, recall 97%, dan f-measure 78% [9]. Penelitian Lestari dkk. (2019) menggunakan metode Learning Vector Quantization (LVQ) dan fitur berbasis leksikon untuk klasifikasi video clickbait di YouTube menunjukkan bahwa penggunaan fitur berbasis leksikon dapat meningkatkan akurasi sistem secara signifikan, yaitu dari 54.54% menjadi 90.91%, precision dari 1 menjadi 85.71%, recall tetap 1, dan f-measure dari 28.58% menjadi 92.31%. Kesimpulannya, penambahan fitur berbasis leksikon pada algoritma Naïve Bayes dan Regresi Logistik memiliki potensi untuk meningkatkan performa analisis sentimen [10].

Topik dan Batasannya

Topik penelitian ini adalah analisis sentimen terhadap platform aplikasi peminjaman online, khususnya Kredivo, dengan menggunakan kombinasi metode Logistic Regression dan Naïve Bayes serta penerapan fitur leksikon. Penelitian ini dibatasi pada analisis ulasan pengguna Kredivo yang tersedia di Google Play Store. Data yang digunakan terbatas pada ulasan dalam bahasa Indonesia. Analisis sentimen akan fokus pada klasifikasi sentimen positif dan negatif, tanpa mempertimbangkan sentimen netral.

Tujuan

Tujuan dari penelitian ini adalah untuk mengembangkan model analisis sentimen yang mengkombinasikan metode Logistic Regression dan Naïve Bayes dengan fitur leksikon. Penelitian ini juga bertujuan untuk mengevaluasi efektivitas model kombinasi tersebut dalam mengklasifikasikan sentimen ulasan pengguna Kredivo. Selain itu, kinerja model kombinasi ini akan dibandingkan dengan metode individual, yaitu Logistic Regression dan Naïve Bayes, yang tidak menggunakan fitur leksikon.

Organisasi Tulisan

Dalam penelitian ini, bagian 2 membahas struktur algoritma yang digunakan, bagian 3 menjelaskan langkah-langkah penelitian, bagian 4 membahas temuan dari hasil penelitian, dan bagian 5 mengambil kesimpulan dari hasil.