

1. Pendahuluan

Latar Belakang

Analisis sentimen adalah teknik komputasi untuk mengenali dan mengklasifikasikan sudut pandang yang direpresentasikan dalam teks, guna menentukan opini penulis apakah positif atau negatif [1]. Teknik ini telah terbukti sangat bermanfaat di hampir semua domain bisnis dan sosial karena sentimen pelanggan tentang produk dan layanan suatu bisnis dapat ditentukan melalui komentar dan ulasan [2]. Kategorisasi sentimen seperti positif dan negatif memungkinkan pengguna untuk memilih opini sesuai dengan preferensi mereka [3]. Mengingat banyaknya postingan di media sosial, analisis sentimen menjadi sangat penting untuk mengklasifikasikan data terstruktur dan tidak terstruktur [4]. E-commerce, yang mengacu pada pertukaran elektronik barang dan jasa melalui alat komunikasi elektronik seperti telepon dan internet [5], telah menjadi kebutuhan yang tak terelakkan bagi masyarakat modern. Perkembangan teknologi informasi dan komunikasi telah mengubah cara orang berbelanja, dengan semakin banyak konsumen beralih ke platform online untuk membeli barang dan jasa. Fenomena ini telah menciptakan pasar e-commerce yang sangat kompetitif, di mana ulasan pelanggan memainkan peran krusial dalam keputusan pembelian dan reputasi merek.

Di Indonesia, salah satu platform e-commerce dengan pengguna terbesar adalah Shopee, yang menjadi fokus penelitian ini. Sebagai pasar yang berkembang pesat, Indonesia menyaksikan pertumbuhan signifikan dalam adopsi e-commerce, dengan Shopee menjadi salah satu pemain utama di industri ini. Popularitas Shopee di kalangan konsumen Indonesia membuat platform ini menjadi sumber yang kaya akan data ulasan pelanggan, menjadikannya objek penelitian yang ideal untuk analisis sentimen dalam konteks e-commerce. Pentingnya analisis sentimen dalam e-commerce tidak dapat dilebih-lebihkan. Bagi perusahaan, memahami sentimen pelanggan memungkinkan mereka untuk meningkatkan produk dan layanan, mengelola reputasi online, dan mengembangkan strategi pemasaran yang lebih efektif. Bagi konsumen, ulasan dan sentimen dari pelanggan lain menjadi panduan berharga dalam membuat keputusan pembelian yang terinformasi. Oleh karena itu, pengembangan metode analisis sentimen yang akurat dan efisien menjadi sangat penting dalam lanskap e-commerce yang terus berkembang.

Beberapa penelitian sebelumnya telah melakukan analisis sentimen menggunakan berbagai metode. Penelitian [6] tentang analisis sentimen ulasan e-commerce yang ditargetkan pada Shopee menggunakan Naïve Bayes Classifier dengan Pembobotan TF-IDF, menunjukkan nilai akurasi 72%, nilai recall 72%, dan nilai presisi 78%. Penelitian [7] menganalisis perbandingan Naïve Bayes dan Regresi Logistik dalam analisis sentimen pada ulasan marketplace menggunakan Pelabelan Berbasis Peringkat, dengan hasil pengujian Naïve Bayes untuk dataset 2 label Tokopedia menunjukkan nilai presisi 81,13%, nilai recall 81,03%, dan nilai akurasi 80,88%. Penelitian [8] menganalisis sentimen ulasan produk di toko online Jiniso Official Shop menggunakan metode Naïve Bayes Classifier (NBC) dengan WordCloud, mencapai tingkat akurasi 94% atau 0,941. Sementara itu, penelitian [9] melakukan studi perbandingan Support Vector Machine dan Naïve Bayes Classifier dengan TF-IDF untuk analisis sentimen pada ulasan produk Amazon, dengan hasil pengujian menunjukkan akurasi Naïve Bayes Classifier sebesar 84%, presisi 82,8%, recall 82,8%, dan f1-score 82,6%. Penelitian [10] menerapkan algoritma Naïve Bayes dengan Pembobotan TF-IDF dan Fitur Unigram pada aplikasi analisis sentimen di toko online Zalora dan Berrybenka, mencapai akurasi 86,66% untuk data pelatihan.

Meskipun penelitian-penelitian tersebut telah menggunakan berbagai platform e-commerce dan mencapai tingkat akurasi yang bervariasi menggunakan Naïve Bayes untuk analisis sentimen, masih ada ruang untuk peningkatan dan eksplorasi lebih lanjut. Khususnya, kombinasi fitur leksikon dan N-gram dengan Naïve Bayes belum banyak dieksplorasi dalam konteks e-commerce. Pendekatan ini berpotensi untuk meningkatkan akurasi klasifikasi sentimen dengan memanfaatkan kekuatan komplementer dari kedua jenis fitur tersebut. Fitur leksikon, yang mengandalkan kamus sentimen yang telah ditentukan sebelumnya, menyediakan dasar yang kuat untuk mengidentifikasi kata-kata yang mengandung sentimen dan menangkap terminologi khusus domain. Di sisi lain, fitur N-gram menangkap konteks lokal dengan mempertimbangkan urutan kata, membuatnya mahir dalam mengidentifikasi pola sentimen dalam kombinasi kata - terutama berharga dalam ulasan e-commerce. Dengan menggabungkan kedua set fitur ini, kami bertujuan untuk memanfaatkan kekuatan masing-masing sambil mengurangi kelemahannya.

Pemilihan Naïve Bayes sebagai algoritma utama dalam penelitian ini didasarkan pada beberapa keunggulan kunci dalam konteks analisis sentimen untuk ulasan e-commerce. Pertama, Naïve Bayes dikenal karena efisiensinya dan kecepatannya, terutama dengan dataset besar, membuatnya sangat cocok untuk memproses volume tinggi ulasan yang umum di platform e-commerce. Kedua, algoritma ini berkinerja baik dengan data berdimensi tinggi, yang merupakan karakteristik data teks ketika menggunakan representasi bag-of-words atau N-gram. Ketiga, Naïve Bayes telah menunjukkan kinerja yang kuat dalam tugas klasifikasi teks, sering mencapai hasil yang kompetitif meskipun asumsinya sederhana. Oleh karena itu, penelitian ini bertujuan untuk melanjutkan dan memperluas pendekatan penelitian sebelumnya dengan menggabungkan fitur leksikon dan N-gram dalam kerangka Naïve Bayes. Harapannya, pendekatan ini dapat memberikan

manfaat signifikan bagi penelitian di masa depan dan memberikan wawasan baru dalam optimalisasi teknik klasifikasi sentimen untuk platform e-commerce.

Topik dan Batasannya

Penelitian ini berfokus pada klasifikasi sentimen untuk platform e-commerce Shopee menggunakan algoritma Naïve Bayes. Topik utama yang akan dieksplorasi adalah penggunaan gabungan fitur Leksikon dan N-gram untuk meningkatkan akurasi klasifikasi sentimen Naïve Bayes. Sistem yang dibangun dalam penelitian ini akan mengambil ulasan pelanggan Shopee sebagai input dan menghasilkan klasifikasi sentimen (positif atau negatif) sebagai output. Pemilihan Shopee sebagai objek penelitian didasarkan pada popularitasnya di Indonesia dan volume besar ulasan pelanggan yang tersedia, menyediakan dataset yang kaya untuk analisis.

Meskipun demikian, penelitian ini memiliki beberapa batasan yang perlu diperhatikan. Pertama, analisis hanya dilakukan pada ulasan berbahasa Indonesia, mengingat konteks lokal dari platform Shopee di Indonesia. Kedua, dataset yang digunakan terbatas pada 10.000 ulasan pelanggan Shopee yang diambil dari Google Play Store. Batasan ini ditetapkan untuk mengelola kompleksitas komputasi dan waktu pengolahan data. Ketiga, klasifikasi sentimen dalam penelitian ini hanya dibagi menjadi dua kategori: positif dan negatif. Meskipun pendekatan ini menyederhanakan analisis, namun tetap mampu memberikan wawasan yang berharga tentang sentimen pelanggan secara umum.

Tujuan

Penelitian ini memiliki beberapa tujuan utama yang saling terkait. Pertama, kami bertujuan untuk mengembangkan model klasifikasi sentimen menggunakan Naïve Bayes dengan fitur Leksikon. Model ini akan memanfaatkan kamus sentimen yang telah ditentukan sebelumnya untuk mengidentifikasi kata-kata yang mengandung sentimen dalam ulasan. Kedua, kami akan mengembangkan model klasifikasi sentimen menggunakan Naïve Bayes dengan fitur N-gram. Pendekatan ini akan mempertimbangkan urutan kata dalam ulasan, yang dapat memberikan konteks yang lebih baik untuk analisis sentimen. Ketiga, kami akan menggabungkan kedua pendekatan tersebut untuk mengembangkan model klasifikasi sentimen menggunakan Naïve Bayes dengan gabungan fitur Leksikon dan N-gram.

Selanjutnya, kami bertujuan untuk membandingkan kinerja ketiga model tersebut dalam hal akurasi, presisi, dan recall. Perbandingan ini akan memberikan wawasan tentang efektivitas relatif dari masing-masing pendekatan dalam konteks ulasan e-commerce. Terakhir, kami akan mengevaluasi efektivitas gabungan fitur Leksikon dan N-gram dalam meningkatkan akurasi klasifikasi sentimen Naïve Bayes untuk ulasan e-commerce. Tujuan ini didasarkan pada hipotesis bahwa kombinasi kedua jenis fitur tersebut dapat menghasilkan model yang lebih robust dan akurat dibandingkan dengan penggunaan fitur tunggal.

Organisasi Tulisan

Selanjutnya, tulisan ini akan diorganisasikan sebagai berikut: Bagian 2 akan membahas studi terkait yang telah dilakukan sebelumnya. Bagian 3 akan menjelaskan metodologi penelitian, termasuk pengumpulan data, preprocessing, ekstraksi fitur, gabungan fitur, dan implementasi model Naïve Bayes. Bagian 4 akan menyajikan hasil eksperimen dan analisis komparatif dari ketiga model yang diusulkan. Terakhir, Bagian 5 akan memberikan kesimpulan dan saran untuk penelitian masa depan.