

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era modern yang selalu berinovasi dan perkembangan yang pesat di berbagai sektor kehidupan menjadikan manusia sangat kompetitif sehingga lowongan pekerjaan formal yang menjamin kesejahteraan sulit didapatkan. Dari fenomena ini banyak orang yang memilih untuk menjalani pekerjaan non formal yaitu berwirausaha sebagai mata pencahariannya, seperti membuat usaha makanan dan minuman karena bidang ini sangat tinggi konsumennya(Arsi & Waluyo, 2021a). Meskipun ramai konsumen, bidang ini juga banyak persaingan terutama untuk pebisnis mikro yang biasa menggunakan resep dari situs maupun aplikasi online. Meskipun demikian tak semua resep yang ditemukan di situs dan aplikasi sesuai dengan apa yang diharapkan meskipun telah mengikuti tiap proses sesuai ketentuan yang ada, disisi lain pengguna sosial media dalam memberi ulasan online memainkan peran penting dalam memengaruhi keputusan individu maupun organisasi(Cahyo Saputra et al., 2025).

Yummy merupakan salah satu aplikasi resep tutorial memasak dan resep yang mengeksplorasi kuliner di Indonesia, aplikasi ini merupakan milik anak perusahaan dari IDN yang menerima ulasan beragam dari para pengguna di Google Play Store(Laila, 2024). dari permasalahan tersebut penelitian ini akan melakukan klasifikasi pada ulasan aplikasi yummy menjadi positif dan negatif lalu melakukan komparasi tiga metode menggunakan naive bayes, Support vector machine (SVM), Bidirectional Encoder Representations from Transformers (BERT).

Naive Bayes dikenal sederhana dan efisien untuk data bersih dan terstruktur, namun sering kali gagal memahami konteks kompleks ini dibuktikan pada penelitian yang dilakukan di balai pelatihan kerja pondok pesantren Al-Ikhwan, metode naive bayes mampu mengklasifikasikan data pendaftar yang lulus dengan akurasi sebesar 90% (Hananto et al., 2024). Pada analisis sentimen yang dilakukan melalui aplikasi twitter untuk menilai komentar positif dan negatif pada proyek kereta cepat pun menghasilkan akurasi sebesar 81% yang menghasilkan

suara terbanyak menyetujui adanya proyek itu (Sanjaya et al., 2024). Sementara hasil dari penelitian (Rian Pratama et al., 2022) mengenai prediksi persediaan stok, metode naive bayes mendapatkan akurasi 91,20% sedangkan C4.5 mendapat akurasi 96,80% hal ini menunjukkan naive bayes sedikit lemah dari metode C4.5 untuk memprediksi stok.

Dari hasil penelitian terdahulu model naïve bayes dan support vector machine menjadi primadona karena keunggulannya dalam melakukan klasifikasi pada sentimen analisis begitu pula pada model baru seperti BERT yang prosesnya menggunakan pemrosesan Bahasa alami sehingga sangat menarik untuk dijadikan penelitian, dengan adanya penelitian komparasi juga diharapkan dapat menambah kajian keilmuan dibidang data mining.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, penelitian ini dirancang untuk menjawab pertanyaan-pertanyaan terkait efektivitas berbagai metode klasifikasi teks dalam konteks ulasan berbahasa Indonesia. Untuk itu, perlu dirumuskan secara sistematis beberapa masalah utama yang menjadi fokus kajian dalam penelitian ini:

1. Bagaimana performa metode Naive Bayes, SVM, dan BERT dalam klasifikasi ulasan negatif dan positif berdasarkan dataset yang tersedia?
2. Bagaimana melakukan komparasi model untuk menentukan metode yang paling optimal dalam hal akurasi, efisiensi, dan kemampuan menangkap konteks teks?

1.3 Tujuan Penelitian

Untuk menjawab rumusan masalah yang telah diajukan, penelitian ini menetapkan sejumlah tujuan yang ingin dicapai. Tujuan-tujuan ini disusun agar penelitian dapat terarah, sistematis, dan memberikan kontribusi nyata baik secara teoretis maupun praktis dalam bidang analisis sentimen berbasis teks:

1. Melakukan Klasifikasi ulasan pada dataset menggunakan model Naïve Bayes, SVM dan BERT
2. Melakukan evaluasi kinerja pada tiap model dan menentukan model paling optimal dalam melakukan klasifikasi

1.4 Manfaat

Penelitian ini diharapkan tidak hanya memberikan hasil secara akademis, tetapi juga memberikan dampak yang lebih luas dalam pengembangan ilmu pengetahuan dan praktik di lapangan. Adapun manfaat yang dapat diperoleh dari penelitian ini dapat ditinjau dari dua sisi, yaitu manfaat teoritis dan manfaat praktis:

1. Teoritis:

Diharapkan penelitian ini dapat dijadikan referensi bagi studi serupa di masa depan, khususnya dalam pengembangan metode analisis teks pada bahasa Indonesia.

2. Praktis:

Membantu memahami opini pengguna dengan lebih akurat, sehingga dapat meningkatkan kualitas layanan atau produk

