

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah mendorong peningkatan penggunaan layanan pesan-antar makanan berbasis aplikasi di Indonesia. Kemudahan akses, efisiensi waktu, serta perubahan pola konsumsi masyarakat menyebabkan layanan pesan-antar makanan semakin banyak dimanfaatkan dalam kehidupan sehari-hari. Persaingan platform digital seperti *GoFood*, *GrabFood*, dan *ShopeeFood* juga menunjukkan pertumbuhan industri layanan pesan-antar makanan yang semakin pesat di Indonesia [1]. Salah satu layanan yang berkembang pesat adalah *ShopeeFood Driver* yang digunakan mitra pengemudi untuk membantu proses operasional pengantaran pesanan. Dalam penggunaannya, kualitas performa aplikasi menjadi faktor penting karena dapat memengaruhi efektivitas kerja dan kepuasan mitra pengemudi. Kualitas sistem dan layanan aplikasi yang baik dapat meningkatkan pengalaman pengguna serta mendukung aktivitas operasional secara lebih efektif [2].

Namun, berdasarkan ulasan pengguna pada platform *Google Play Store* masih ditemukan berbagai keluhan terkait performa aplikasi *ShopeeFood Driver*, seperti aplikasi *error*, sistem lambat, kendala navigasi, dan gangguan penerimaan pesanan. Ulasan pengguna pada *Google Play Store* dapat mencerminkan pengalaman penggunaan serta berbagai permasalahan layanan aplikasi yang dirasakan oleh pengguna [3]. Ulasan pengguna tersebut mengandung opini, kritik, serta pengalaman pengguna yang dapat dimanfaatkan sebagai bahan evaluasi layanan aplikasi. Analisis ulasan pengguna sangat berguna dalam pengembangan fitur layanan dan peningkatan kualitas aplikasi karena mencerminkan pengalaman serta kepuasan pengguna terhadap sistem yang digunakan. Jumlah ulasan yang terus bertambah menyebabkan proses analisis secara manual menjadi kurang efektif karena membutuhkan waktu yang lama dan berpotensi menghasilkan penilaian yang subjektif. Oleh karena itu, diperlukan metode otomatis yang mampu

mengidentifikasi dan mengklasifikasikan sentimen pengguna berdasarkan data teks ulasan [4].

Salah satu pendekatan yang banyak digunakan untuk mengolah data teks adalah analisis sentimen. Analisis sentiment merupakan bagian dari *text mining* dan *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi opini positif maupun negative dari suatu data teks [5]. Dalam penerapannya, metode *machine learning* yang sering digunakan dalam penelitian analisis sentimen adalah Random Forest dan XGBoost [6].

Random Forest Merupakan algoritma *ensemble learning* yang bekerja dengan membangun banyak *decision tree* untuk meningkatkan akurasi klasifikasi dan mengurangi risiko *overfitting*. Algoritma ini diketahui memiliki performa yang baik dalam menangani data berdimensi tinggi seperti data teks [7]. Sementara itu, XGBoost merupakan algoritma *boosting* yang mampu meningkatkan performa model melalui proses optimasi *gradient boosting* secara bertahap sehingga menghasilkan performa klasifikasi yang tinggi[8]. Kedua algoritma tersebut dipilih karena sama-sama merupakan metode *ensemble learning* yang umum digunakan dalam klasifikasi data teks, namun memiliki pendekatan yang berbeda, yaitu *bagging* pada Random Forest dan *boosting* pada XGBoost, sehingga memungkinkan adanya perbedaan performa pada dataset tertentu.

Beberapa penelitian sebelumnya menunjukkan adanya perbedaan performa antara Random Forest dan XGBoost pada analisis sentimen. Penelitian oleh Hayyun Lestia Wati dkk [9] menunjukkan bahwa kedua algoritma memiliki performa yang berbeda pada analisis sentimen publik terhadap Rumah Makan Payakumbuh, di mana XGBoost memperoleh hasil evaluasi yang lebih baik dibandingkan Random Forest. Selain itu, penelitian oleh N. Epriyanti, A. Meiriza, dan D. Y. Hardiyanti [6] juga menunjukkan bahwa XGBoost menghasilkan performa klasifikasi yang lebih tinggi dibandingkan Random Forest pada analisis sentimen aplikasi *Access by KAL*. Hasil penelitian tersebut menunjukkan bahwa algoritma XGBoost memiliki performa yang baik dalam klasifikasi sentimen, namun performa model tetap dapat dipengaruhi oleh karakteristik dataset dan metode preprocessing yang digunakan.

Selain itu, ulasan pengguna aplikasi berbahasa Indonesia memiliki tantangan tersendiri karena mengandung bahasa tidak baku, singkatan, *typo*, dan *slang word* yang dapat memengaruhi hasil klasifikasi sentimen. Penelitian oleh Ahmad Fauzi dkk [5] menunjukkan bahwa data teks berbahasa Indonesia pada media sosial memiliki karakteristik data yang tidak terstruktur sehingga memerlukan tahapan *preprocessing*, penyeimbangan data menggunakan SMOTE, serta optimasi model untuk meningkatkan performa klasifikasi sentimen. Berdasarkan penelitian terdahulu, penelitian yang secara khusus mengevaluasi dan membandingkan performa Random Forest dan XGBoost pada analisis sentimen ulasan aplikasi *ShopeeFood Driver* berbahasa Indonesia masih terbatas. Oleh karena itu, diperlukan penelitian untuk mengevaluasi performa kedua algoritma tersebut pada karakteristik data ulasan *ShopeeFood Driver* yang bersifat tidak terstruktur dan menggunakan bahasa tidak baku [10].

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk membandingkan performa algoritma Random Forest dan XGBoost menggunakan metode TF-IDF pada klasifikasi sentimen ulasan pengguna aplikasi *ShopeeFood Driver* berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-Score*. Hasil penelitian diharapkan dapat memberikan kontribusi dalam bidang analisis sentimen, khususnya terkait evaluasi performa algoritma *machine learning* pada data teks berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang penelitian yang telah dipaparkan sebelumnya, maka permasalahan dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana performa algoritma Random Forest dan XGBoost berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* pada klasifikasi sentimen ulasan pengguna aplikasi *ShopeeFood Driver*?
2. Algoritma manakah yang memiliki performa terbaik dalam klasifikasi sentimen ulasan pengguna aplikasi *ShopeeFood Driver* berdasarkan hasil evaluasi *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Batasan Masalah

Batasan masalah ditetapkan dalam penelitian ini dengan tujuan untuk memperjelas ruang lingkup kajian sehingga pembahasan dapat dilakukan secara terfokus. Adapun ruang lingkup dan batasan penelitian yang digunakan adalah sebagai berikut:

1. Penelitian ini hanya difokuskan pada aplikasi ShopeeFood Driver sebagai objek kajian, sehingga data ulasan yang dianalisis berasal dari pengguna aplikasi tersebut.
2. Data penelitian yang digunakan berupa 10.000 baris data ulasan pengguna ShopeeFood Driver yang diperoleh melalui proses web scrapping dari Google Play Store pada periode pengambilan data tertentu.
3. Klasifikasi sentimen dalam penelitian ini dibatasi pada 2 kelas (binary classification), yaitu sentiment positif dan negative.
4. Pendekatan yang digunakan dalam penelitian ini hanya mencakup dua algoritma machine learning, yakni Random Forest dan XGBoost, tanpa melakukan perbandingan dengan metode klasifikasi lainnya.
5. Proses pengolahan data teks dibatasi pada tahapan preprocessing dasar, yang meliputi basic cleaning seperti case folding dan penghapusan karakter yang tidak diperlukan, tokenisasi, normalisasi kata tidak baku (*slang word*), normalisasi kata berulang, *stopword removal*, stemming, serta ekstraksi fitur menggunakan metode TF-IDF.
6. Pengukuran performa model hanya dilakukan berdasarkan metrik evaluasi tertentu, yaitu akurasi, precision, recall, dan F1-score, tanpa membahas secara rinci aspek lain seperti efisiensi waktu komputasi.
7. Ulasan yang dianalisis dibatasi pada teks berbahasa Indonesia, sehingga ulasan dengan bahasa selain Indonesia tidak termasuk dalam cakupan penelitian.

1.4 Tujuan Penelitian

Penelitian ini dilaksanakan sebagai upaya memberikan pemahaman yang lebih mendalam terhadap permasalahan yang telah diuraikan sebelumnya, khususnya yang berkaitan dengan analisis sentiment ulasan pengguna aplikasi ShopeeFood Driver. Oleh karena itu, tujuan yang ingin dicapai sebagai berikut:

1. Mengukur performa algoritma Random Forest dan XGBoost pada klasifikasi sentimen ulasan pengguna aplikasi ShopeeFood Driver menggunakan metrik accuracy, precision, recall, dan F1-score.
2. Membandingkan performa algoritma Random Forest dan XGBoost berdasarkan hasil confusion matrix, accuracy, precision, recall, dan F1-score untuk menentukan algoritma dengan performa klasifikasi terbaik.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari pelaksanaan penelitian ini tidak hanya terbatas pada aspek akademik, tetapi juga mencakup manfaat yang dapat dirasakan secara langsung oleh pihak-pihak yang berkaitan dengan objek penelitian. Secara umum, manfaat penelitian ini dapat ditinjau dari sisi teoritis serta manfaat praktis yang ditujukan bagi penggunaan organisasi yang terkait dengan aplikasi ShopeFood Driver

1.5.1 Manfaat Teoritis

Secara teoritis, penelitian ini diharapkan dapat memperkaya kajian ilmiah dalam bidang analisis sentiment dan pengolahan data teks, khususnya yang berkaitan dengan penerapan dan perbandingan algoritma Random Forest dan XGBoost dalam proses klasifikasi sentiment ulasan pada aplikasi layanan pesan-antar makanan. Selain itu, hasil penelitian ini diharapkan dapat menjadi sumber referensi dan bahan pembanding bagi penelitian selanjutnya yang membahas penggunaan metode machine learning dalam permasalahan analisis sentiment dengan objek penelitian yang sejenis.

1.5.2 Manfaat Praktis

Secara praktis, hasil penelitian ini diharapkan dapat memberikan Gambaran

terkait penerapan metode machine learning dalam analisis sentiment ulasan aplikasi yang bersumber dari Google Play Store, sehingga dapat dijadikan sebagai bahan pertimbangan dalam pengembangan model analisis sentiment pada penelitian selanjutnya yang menggunakan dataset ulasan aplikasi digital.

1.6 Sistematika Penulisan

Penyusunan sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum mengenai alur pembahasan dan cakupan materi yang dibahas pada setiap bab dalam penelitian ini. Adapun susunan penulisan skripsi dijelaskan sebagai berikut:

BAB I PENDAHULUAN,

Bab ini menguraikan Gambaran awal penelitian yang mencakup latar belakang permasalahan, perumusan masalah, batasan penelitian, tujuan penelitian, manfaat penelitian, serta penjelasan mengenai sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA,

Bab ini membahas kajian literatur yang mendukung penelitian, meliputi penelitian-penelitian terdahulu yang relevan serta landasan teori yang berkaitan dengan analisis sentiment, konsep machine learning, algoritma ERandom Forest dan XGBoost, serta teori pendukung lainnya.

BAB III METODE PENELITIAN,

Bab ini menjelaskan rancangan dan metode penelitian yang diterapkan, termasuk deskripsi objek penelitian, sumber dan metode pengumpulan data, tahapan praproses data teks, Teknik ekstraksi fitur, algoritma klasifikasi yang digunakan, serta metode evaluasi kinerja model

BAB IV HASIL DAN PEMBAHASAN,

Bab ini menyajikan hasil penerapan dan pengujian model analisis sentiment menggunakan algoritma Random Forest dan Xgboost. Selain itu, bab ini juga memuat analisis dan pembahasan mengenai hasil evaluasi serta perbandingan kinerja kedua algoritma berdasarkan metrik yang digunakan.

BAB V PENUTUP,

Bab ini berisi Kesimpulan yang diperoleh dari keseluruhan hasil penelitian serta saran yang dapat dijadikan sebagai bahan pertimbangan untuk penelitian dan kajian lanjutan di masa mendatang.

