

BAB I

PENDAHULUAN

1.1 Latar Belakang

Aplikasi transportasi online merupakan modifikasi dari ojek konvensional yang biasanya berada di suatu pangkalan. Tidak hanya layanan pemesanan transportasi motor atau mobil saja yang ditawarkan kepada pengguna aplikasi, tetapi juga layanan lainnya seperti, pemesanan makanan, pengiriman barang, belanja, dan masih banyak lagi (Que et al., 2020). Salah satu aplikasi transportasi online yaitu, aplikasi Gojek. Pada laman databoks aplikasi Gojek mendapatkan jumlah unduhan terbanyak dari tahun 2022-2023 dengan rata-rata unduhan per bulan sebesar 957 ribu pada tahun 2023 (Santika, 2024). Pengguna yang telah menggunakan aplikasi Gojek dapat berbagi pengalaman selama menggunakan aplikasi tersebut, baik itu yang positif atau negatif di dalam kolom komentar di situs Google Play yang terdiri dari rating dan ulasan (Amalia et al., 2023). Namun, pengguna sering kali memberikan penilaian yang tidak sejalan dengan komentar yang diberikan. Sehingga hal tersebut tidak sepenuhnya mencerminkan kualitas aplikasi (Suryati et al., 2023).

Oleh karena itu, diperlukan adanya analisis sentimen, dimana tugas utamanya untuk dapat mengidentifikasi polaritas yang terkandung dalam teks pada dokumen atau kalimat apakah mengandung makna positif atau negatif (Maarif & Setiyawati, 2024). Sehingga ulasan-ulasan yang digunakan pada penelitian ini dengan adanya analisis sentimen dapat dijadikan penilaian seberapa baik atau buruknya layanan aplikasi Gojek. Penerapan analisis sentimen tersebut diperlukan adanya algoritma khusus yang dapat membantu mengategorikan ulasan-ulasan tersebut baik positif atau negatif (Rokhman et al., 2021). Penggunaan algoritma *Random Forest* dipilih karena merujuk pada jurnal (Ratnawati & Sulistyaningrum, 2020) dengan judul 'Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel' yang menyatakan bahwa algoritma *Random Forest* memiliki tingkat akurasi yang bagus, dapat meminimalisir *outliers* dan *noise*, dan lebih cepat dari pada *bagging* dan *boosting*. Kemudian penggunaan algoritma *K-Means* menurut penelitian (Iparraguirre-Villanueva et al., 2022) dengan judul 'Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm'

dapat meningkatkan tingkat akurasi dalam klasifikasi analisis sentimen. Penggabungan kedua algoritma ini cukup strategis, karena algoritma *K-Means* dapat mengelompokkan data ke dalam *cluster* yang lebih homogen. Sehingga *Random Forest* dapat lebih fokus pada pola yang ada dalam setiap *cluster*. Hal tersebut pula dapat membantu mempercepat proses pelatihan model.

Menurut (Cindo & Rini, 2019) analisis sentimen dilakukan agar dapat mengidentifikasi opini, emosi, dan sikap yang tergambar dalam sebuah teks, sehingga terbagi menjadi opini positif dan negatif. Proses analisis tersebut akan melakukan klasifikasi data dengan memanfaatkan algoritma *machine learning* (Manik et al., 2021). Sehingga bila sudah dilakukan analisis lebih lanjut dari data ulasan yang sudah didapatkan, hasilnya dapat digunakan oleh perusahaan untuk meningkatkan kinerja dari aplikasinya.

Penelitian lain yang dilakukan oleh (Amalia et al., 2023) dengan judul “Analisa Sentimen Pelanggan pada Review Online Berbasis Text Mining Menggunakan Algoritma K-Means” memperlihatkan bahwa penggunaan algoritma K-Means efektif dalam melakukan pengelompokan data yang beragam sebelum melakukan proses analisis sentimen. Penelitian lain yang dilakukan oleh (Huda et al., 2023) dengan judul “Analisis Sentimen Perbandingan Layanan Jasa Pengiriman Kurir Pada Ulasan Play Store Menggunakan Algoritma *Random Forest* dan *Decision Tree*” memberikan hasil bahwa penggunaan algoritma *Random Forest* memiliki akurasi yang lebih unggul dibandingkan algoritma *Decision Tree*. Pada penelitian (Rokhman et al., 2021) dengan judul “Perbandingan Metode *Support Vector Machine* dan *Decision Tree* Untuk Analisis Sentimen *Review* Komentar Pada Aplikasi Transportasi Online” memberikan hasil bahwa penggunaan algoritma *Support Vector Machine* memiliki nilai akurasi yang lebih tinggi dengan nilai 90.20%. Pada jurnal tersebut juga terdapat saran untuk menggabungkan algoritma klasifikasi dengan algoritma clustering atau algoritma asosiasi.

Penggunaan *K-Means* dalam analisis sentimen memiliki potensi untuk melakukan pengelompokan terhadap data ulasan beragam dan jumlahnya sangat banyak ke dalam beberapa cluster. Setiap cluster nantinya akan berisi data yang memiliki jarak minimum antara titik data dan cluster, yang dihitung menggunakan rumus *euclidian distance* (Safitri et al., 2021). *K-Means* merupakan algoritma yang

cukup cepat melakukan clustering (Hidayatullah, 2018). Setelah data dikelompokkan, dilakukan klasifikasi menggunakan algoritma *Random Forest* karena dapat melakukan proses klasifikasi dengan jumlah data yang banyak dan juga kompleks, dapat mengatasi keberadaan *noise* dan *missing value*, dan memiliki hasil akurasi yang baik (Indrayanto et al., 2023). Berdasarkan latar belakang tersebut, penelitian ini diberi judul “Analisis Sentimen Ulasan Aplikasi Gojek Menggunakan Algoritma Kombinasi *K-Means* dan *Random Forest*”. Penggabungan kedua algoritma ini diharapkan dapat meningkatkan nilai *accuracy* klasifikasi data. Pada penelitian ini studi kasus yang akan digunakan adalah ulasan aplikasi Gojek yang diambil dari situs Kaggle, kemudian akan dilakukan evaluasi terhadap hasil analisis sentimen yang sudah didapatkan menggunakan *confusion matrix*.

1.2 Rumusan Masalah Penelitian

Berdasarkan latar belakang tersebut, rumusan masalah dari penelitian ini antara lain:

1. Bagaimana efektivitas kombinasi algoritma *K-Means* dan *Random Forest* dalam analisis sentimen ulasan aplikasi Gojek?
2. Bagaimana hasil evaluasi terhadap analisis sentimen ulasan aplikasi Gojek?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini antara lain:

1. Mengetahui hasil analisis sentimen menggunakan kombinasi algoritma *K-Means* dan *Random Forest*.
2. Mengetahui hasil evaluasi analisis sentimen dengan studi kasus data ulasan aplikasi Gojek.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Memberikan pengetahuan baru terkait penggunaan kombinasi algoritma *K-Means* dan *Random Forest* dalam melakukan analisis sentimen.
2. Membantu perusahaan dalam studi kasus ini adalah Gojek untuk dapat meningkatkan kualitas aplikasinya.

1.5 Batasan Penelitian

Batasan dari penelitian ini antara lain:

1. Sumber data ulasan aplikasi Gojek yang digunakan diambil dari situs Kaggle.
2. Data diambil dari mulai November tahun 2021 hingga Februari tahun 2024.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian yang akan dilakukan seperti permasalahan yang ada dan yang akan diselesaikan, rumusan masalah, tujuan penelitian, manfaat penelitian, dan struktur organisasi.

BAB II KAJIAN PUSTAKA

Pada bab ini berisikan tentang teori-teori yang berkaitan dengan penelitian yang akan dilakukan dan terdapat penelitian terkait yang sudah dilakukan sebelumnya.

BAB III METODELOGI PENELITIAN

Pada bab ini berisikan tentang algoritma yang digunakan dalam penelitian yang dilakukan, terdapat populasi dan sample, instrumen penelitian, prosedur yang dilakukan pada penelitian, dan juga analisis data yang dibutuhkan.

BAB IV TEMUAN DAN PEMBAHASAN

Pada bab ini berisikan temuan dan pembahasan terkait penelitian yang seusi dengan apa saja yang sudah dirumuskan pada rumusan masalah penelitian.

BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI

Pada bab ini berisikan simpulan yang didapatkan setelah melakukan penelitian, implikasi penelitian, dan rekomendasi untuk penelitian selanjutnya.