

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan inovasi digital mengalami peningkatan yang sangat pesat dan telah membawa dampak signifikan pada berbagai aspek kehidupan, seperti transportasi, komunikasi, hingga layanan digital yang mempermudah aktivitas sehari-hari. Salah satu inovasi tersebut adalah teknologi digital ride-hailing, seperti aplikasi Grab. Aplikasi ini tidak hanya menyediakan layanan transportasi, tetapi juga fitur tambahan seperti pemesanan makanan, pengiriman barang, serta layanan pembayaran digital. Di Indonesia, Grab telah berkembang menjadi salah satu platform digital terpopuler dalam memenuhi kebutuhan transportasi dan logistik harian, seiring dengan meningkatnya jumlah pengguna.

Sebagai aplikasi dengan jutaan pengguna, Grab menerima banyak ulasan dari penggunanya di Google Play Store. Ulasan pengguna terdiri dari dua elemen utama, yakni nilai yang berupa rating dan komentar dalam bentuk teks [1]. Ulasan-ulasan ini dapat menggambarkan bagaimana pengalaman pengguna dalam menggunakan layanan Grab, baik dari mereka yang merasa puas maupun yang mengalami kendala. Banyak pengguna memberikan ulasan positif karena merasa puas dengan kecepatan layanan, keramahan pengemudi, atau kemudahan dalam penggunaan aplikasi. Sebaliknya, tidak sedikit juga pengguna yang memberikan ulasan negatif akibat masalah teknis pada aplikasi, tarif yang dianggap tidak sesuai, keterlambatan pengemudi, atau kendala dalam sistem pembayaran.

Dataset yang berupa ulasan pengguna merupakan data yang sangat penting bagi perusahaan sebagai bahan evaluasi untuk meningkatkan kualitas layanan digital bagi platform tersebut. Oleh karena itu, diperlukan metode yang tepat guna untuk menganalisis pola dari ulasan pengguna di Google Play Store secara sistematis, salah satunya melalui analisis sentimen. Analisis sentimen adalah bidang yang luas yang mencakup pemrosesan bahasa alami, linguistik komputasi, dan penambahan

teks. Tujuannya adalah untuk menganalisis komentar, sikap, evaluasi, sentimen, dan penilaian seseorang tentang pembicara atau penulis yang terkait dengan produk, topik, organisasi, individu, layanan, atau aktivitas [2]. Dengan analisis sentimen kita dapat memahami apakah isi suatu kalimat bersifat positif atau negatif.

Kendala yang sering ditemui dalam analisis sentimen yaitu ketidakseimbangan jumlah data dari kelas yang ada pada dataset ulasan. Yang paling sering ditemukan adalah ulasan negatif bisa jauh lebih banyak jika dibandingkan dengan ulasan positif dan bisa sebaliknya ulasan positif bisa lebih banyak daripada ulasan negatif. Ketidakseimbangan ini dapat berdampak pada kinerja algoritma machine learning, karena model cenderung memprediksi kelas mayoritas dan kurang memperhatikan kelas minoritas, meskipun kelas minoritas justru krusial untuk dideteksi dengan akurat. Kondisi ini tentunya akan membuat hasil dari analisis sentimen kurang efektif dan mengurangi akurasi klasifikasi dalam mengidentifikasi berbagai permasalahan yang dialami pengguna [3].

Untuk mengatasi permasalahan ini, maka diterapkan teknik resampling yang berfungsi untuk menyeimbangkan jumlah data antara ulasan positif dan negatif. Penelitian menerapkan beberapa teknik resampling, seperti oversampling menggunakan metode SMOTE (Synthetic Minority Over-sampling Technique) Salah satu pendekatan yang terbukti efektif dalam mengatasi ketidakseimbangan kelas mampu meningkatkan akurasi dan sensitivitas model klasifikasi terhadap kelas minoritas [4]. ADASYN (Adaptive Synthetic Sampling) terbukti mampu meningkatkan performa algoritma klasifikasi pada data tidak seimbang, terutama dari sisi geometric mean. ADASYN bekerja dengan menyesuaikan jumlah data sintetis berdasarkan kompleksitas sampel minoritas yang sulit dipelajari oleh model, sehingga pembelajaran lebih adaptif dan fokus pada area yang sulit dipisahkan [5]. Random Oversampling digunakan untuk menyeimbangkan distribusi kelas pada data latih. Pendekatan ini merujuk pada sampel dari kelas minoritas direplikasi secara acak hingga proporsinya sama dengan kelas mayoritas sebelum dilakukan pelatihan model klasifikasi [6].

Teknik hybrid sampling seperti SMOTEENN merupakan metode hybrid sampling yang menggabungkan teknik oversampling SMOTE dan undersampling ENN. Teknik ini digunakan untuk menangani ketidakseimbangan kelas dengan terlebih dahulu memperbanyak data kelas minoritas menggunakan SMOTE, kemudian membersihkan data dari noise dan overlap menggunakan ENN [7].

Support Vector Machine (SVM) dipilih sebagai algoritma untuk melakukan klasifikasi dalam penelitian ini karena memiliki keunggulan dalam menangani data teks berdimensi tinggi terutama Ketika jumlah fitur (kata-kata atau istilah) sangat besar. SVM efektif dalam menangani masalah ruang dimensi tinggi dan dapat memberikan hasil yang baik bahkan dengan jumlah fitur yang besar [8]. Meskipun demikian, SVM juga memiliki tantangan dalam mengelola dataset yang sangat besar dan tidak seimbang, sehingga diperlukan strategi pendukung untuk mengoptimalkan kinerjanya dalam analisis sentimen. Oleh karena itu, penelitian ini mengombinasikan algoritma Support Vector Machine dengan pendekatan teknik resampling untuk meningkatkan performa klasifikasi yang lebih akurat dan seimbang.

Penelitian oleh Sujjada et al. (2023) menunjukkan bahwa algoritma Support Vector Machine (SVM) mampu memberikan performa klasifikasi yang sangat baik dalam analisis sentimen terhadap ulasan aplikasi digital di Google Play Store. Dalam klasifikasi ulasan terhadap tiga aplikasi bank digital (Bank Jago, Neobank, dan Seabank), SVM menghasilkan akurasi rata-rata hingga 91%, dengan nilai F1-score tertinggi mencapai 0,94 [9]. Pada penelitian selanjutnya Teknik oversampling menggunakan SMOTE terbukti efektif dalam meningkatkan performa model klasifikasi pada data tidak seimbang. Dalam penelitian Wulandari dan Badieah (2025), penerapan SMOTE dengan pendekatan One-vs-One (OvO) pada algoritma SVM menghasilkan performa terbaik dengan nilai akurasi sebesar 85%, precision 87%, recall 90%, dan F1-score 87%. Hasil ini lebih tinggi dibandingkan baseline model tanpa resampling, yang hanya mencapai akurasi 82%, dan jauh di atas metode undersampling yang hanya menghasilkan akurasi 59% [10]. Diharapkan dengan penelitian ini dapat meningkatkan pemahaman mengenai penerapan

algoritma Support Vector Machine (SVM) yang dikombinasikan dengan teknik resampling dalam analisis sentimen data tekstual. Hasil klasifikasi yang diperoleh akan memberikan gambaran mengenai efektivitas pengolahan data tidak seimbang dan akurasi model, yang dapat menjadi dasar pengembangan metode klasifikasi sentimen yang lebih optimal di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana sentimen pengguna terhadap aplikasi Grab di Google Play Store?
2. Bagaimana pengaruh penerapan teknik resampling, yaitu oversampling (SMOTE, ADASYN, Random Oversampling) dan hybrid sampling (SMOTEENN) terhadap distribusi dan keseimbangan kelas data ulasan Grab?
3. Bagaimana efektivitas penerapan teknik resampling dapat meningkatkan performa algoritma Support Vector Machine (SVM) untuk klasifikasi sentimen ulasan pengguna aplikasi Grab?

1.3 Batasan Masalah

Batasan masalah dalam penelitian tentang penerapan metode Support Vector Machine untuk analisis sentimen pada ulasan pengguna aplikasi Grab di Google Play Store sebagai berikut:

1. Data ulasan pengguna yang digunakan dalam penelitian ini hanya berasal dari Google Play Store dan tidak mencakup ulasan dari platform lain seperti App Store atau media social.
2. Dataset yang digunakan dikumpulkan dengan metode web scraping dengan jumlah sekitar 15.000 ulasan dari aplikasi.

3. Dataset mencakup informasi penting seperti teks ulasan, tanggal unggahan dan rating.
4. Algoritma klasifikasi yang digunakan adalah Support Vector Machine (SVM) karena keunggulannya dalam menangani data teks
5. Analisis sentimen akan diklasifikasikan ke dalam dua kategori, yaitu positif dan negatif, berdasarkan ulasan pengguna aplikasi Grab di Google Play Store
6. Evaluasi performa model klasifikasi dilakukan berdasarkan metrik evaluasi standar, seperti akurasi, precision, recall, dan F1-score.
7. Menggunakan teknik resampling dataset yaitu oversampling (SMOTE, ADASYN, Random Oversampling) dan hybrid sampling (SMOTEENN).

1.4 Tujuan Penelitian

Tujuan Penelitian dari penerapan metode Support Vector Machine untuk analisis sentimen pada ulasan pengguna aplikasi Grab di Google Play Store sebagai berikut:

1. Melakukan analisis sentimen ulasan pengguna aplikasi Grab di Google Play Store menggunakan metode Support Vector Machine (SVM).
2. Menguji efektivitas teknik resampling, yaitu oversampling (SMOTE, ADASYN, Random Oversampling) dan hybrid sampling (SMOTEENN).
3. Memeriksa tingkat akurasi dan melakukan evaluasi terhadap metode Support Vector Machine (SVM) dalam klasifikasi sentimen ulasan pengguna setelah dilakukan teknik resampling.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Memberikan wawasan mengenai kepuasan serta aspek layanan yang perlu diperbaiki untuk meningkatkan pengalaman pengguna.
2. Memberikan referensi mengenai implementasi metode Support Vector

Machine (SVM) dalam analisis ulasan aplikasi transportasi online.

3. Membantu pengguna dalam memahami tren dan kualitas layanan berdasarkan ulasan pengguna lainnya di Google Play Store.

1.6 Sistematika Penulisan

Berikut adalah sistematika penulisan dalam penelitian ini :

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang studi literatur yaitu penjelasan penelitian sebelumnya dan juga mengulas teori-teori yang relevan dengan penelitian ini.

BAB III METODE PENELITIAN

Menjelaskan metode dan langkah-langkah yang digunakan dalam penelitian, mulai dari pengumpulan data hingga evaluasi hasil.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini menyajikan hasil analisis sentimen dan pembahasan mengenai temuan penelitian.

BAB V PENUTUP

Bab ini berisi mengenai kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan dan saran yang dapat peneliti sampaikan untuk penelitian selanjutnya.