

BAB 1

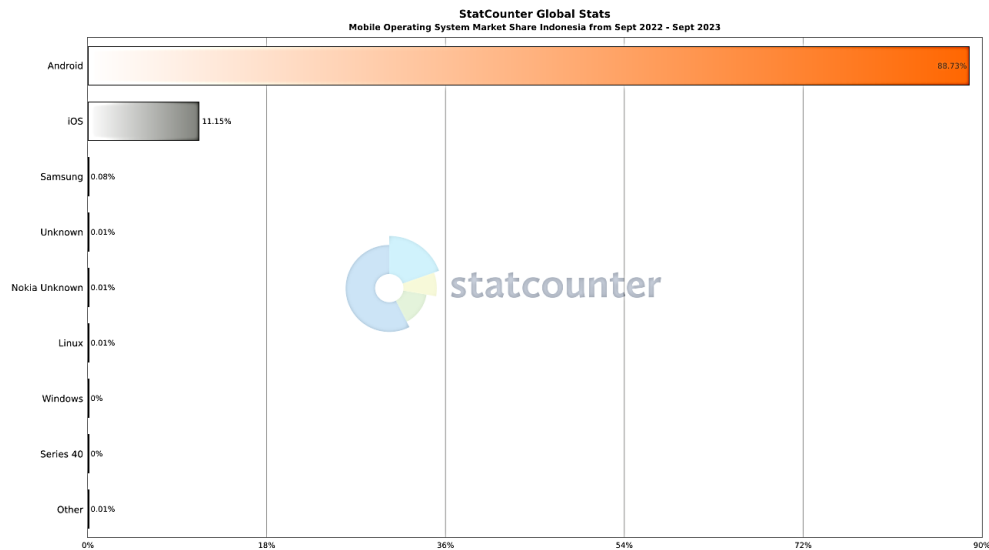
PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi memiliki pengaruh signifikan bagi kehidupan manusia[1]. Hasil nyata dari perkembangan teknologi ini salah satunya adalah peningkatan yang cukup signifikan dalam popularitas penggunaan layanan transportasi *online* melalui internet[1] dengan proses pemesanan yang mudah melalui aplikasi atau situs *web* khusus. Selain itu, karena penggunaan layanan transportasi *online* cenderung mengikuti konsep bisnis berdasarkan permintaan, layanan tersebut didasarkan pada adanya permintaan konsumen, sehingga penyedia layanan akan memberikan kenyamanan yang lebih baik kepada konsumen.

Terdapat sejumlah layanan transportasi *online* yang ada di Indonesia, di antaranya Gojek, *InDrive*, *Maxim* dan Grab. Keberagaman ini memberikan berbagai opsi yang memungkinkan pengguna untuk menentukan layanan yang paling sesuai dengan kebutuhan pelanggan. Grab merupakan satu di antara *startup* yang paling umum digunakan di kawasan Asia Tenggara tahun 2022 untuk pengantar makanan[2] dan termasuk *superapp* terbaik di Asia Tenggara tahun 2022[3]. Grab sebagai pemimpin dalam layanan pengantaran makanan dan jasa transportasi *online* di wilayah Asia Tenggara. Oleh karena itu, Grab telah merilis aplikasi *superapp* yang tersedia untuk berbagai sistem operasi *mobile*, termasuk *Android* dan *iOS*.

Dalam Gambar 1.1 tergambar jelas dominasi sistem operasi *mobile* di Indonesia, di mana *Android* menduduki posisi puncak sebagai sistem operasi *mobile* yang paling populer dengan pangsa pasar sekitar 88,73%. Sementara itu, *iOS* yang digunakan oleh perangkat *Apple*, memiliki pangsa pasar sebesar 11,15%[4]. *Google Play Store*, sebagai salah satu *Apps Market* yang disediakan oleh sistem operasi *mobile Android* memberikan fasilitas bagi pengguna untuk memberikan ulasan dan penilaian terhadap aplikasi yang tersedia di dalamnya menggunakan *rating* 1 sampai 5.



Gambar 1. 1 Pengguna *mobile operating system* di Indonesia (Sumber: *gs.statcounter.com*)

Terkadang penilaian menggunakan rating 1 sampai 5 dari pengguna tidak selalu mencerminkan isi ulasan pengguna. Situasi ini dapat mengakibatkan peringkat tidak sepenuhnya mencerminkan kualitas aplikasi tersebut. Ulasan dalam bentuk kalimat lebih rinci dalam menggambarkan respon pengguna terhadap aplikasi. Ulasan semacam itu mungkin berdampak besar pada calon pengunjung atau pengguna baru yang sedang mempertimbangkan untuk mengunduh aplikasi Grab[5].

Pengolahan data ulasan pada penelitian ini dapat dilakukan dengan analisis sentimen atau penambangan opini yaitu bidang penelitian yang mempelajari sentimen, pendapat dan emosi manusia terkait berbagai peristiwa. Ulasan aplikasi Grab di *Google Play Store* pada penelitian ini berada di tingkat kalimat pada analisis sentimen[6]. Setiap sentimen positif, negatif dan netral terhadap aplikasi Grab akan merujuk pada fitur atau pelayanan yang ada di aplikasi Grab.

Analisis sentimen pada ulasan aplikasi Grab di *Google Play Store* salah satunya dapat dilakukan menggunakan metode klasifikasi. Beberapa metode klasifikasi diantaranya, *Naïve Bayes*, *KNN*, *Random Forest*, *SVM* dan lainnya. Menurut penelitian sebelumnya[7] metode *Random Forest* memiliki keunggulan, yaitu mampu menghasilkan kesalahan yang relatif rendah, memiliki kinerja yang baik dan cocok untuk digunakan pada data dalam jumlah besar.

Pada penelitian sebelumnya[7] tentang analisis sentimen, dilakukan perbandingan antara metode *Random Forest* dengan *Naïve Bayes*. Penelitian tersebut menggunakan 2.453 data ulasan aplikasi *BCA Mobile* dan menggunakan ekstraksi fitur *TF-IDF*. Dari perbandingan metode *Naïve Bayes* dengan *Random Forest* tersebut disimpulkan bahwa metode *Random Forest* memberikan hasil yang lebih tinggi dalam hal nilai akurasi, *recall* dan *f1-score* daripada metode *Naïve Bayes* yang hanya unggul pada nilai presisi.

Pada penelitian yang lain[8] dilakukan perbandingan antara metode *KNN* (*K-Nearest Neighbor*) dan *Random Forest* untuk analisis sentimen twitter yang terdiri dari 1000 data komentar. Dari penelitian tersebut metode *Random Forest* menghasilkan nilai akurasi yang lebih tinggi daripada metode *KNN*.

Pada penelitian[9] dan penelitian[10] digunakan teknik *sampling* untuk mengatasi jumlah data yang tidak seimbang antar kelas. Teknik yang digunakan yaitu *SMOTE* (*Synthetic Minority Oversampling Technique*). Teknik *SMOTE* menyeimbangkan jumlah data antar kelas dengan menambah data baru pada kelas minoritas dan melakukan *resampling*. Dari penelitian tersebut, didapatkan hasil bahwa penggunaan teknik *sampling SMOTE* dapat menaikkan akurasi sistem.

Bagian yang penting dalam meningkatkan performa klasifikasi analisis sentimen adalah melalui seleksi fitur. Untuk mengurangi dimensi ruang fitur yang besar, salah satu cara yang efektif adalah dengan menghilangkan fitur atau atribut yang tidak begitu relevan. Jika menggunakan algoritma seleksi fitur yang sesuai, dapat meningkatkan tingkat akurasi klasifikasi secara signifikan. Dalam klasifikasi, *Chi Square* merupakan salah satu jenis seleksi fitur *supervised* yang dapat mengurangi jumlah fitur dengan signifikan tanpa menurunkan tingkat akurasi. Selain itu, *Chi Square* juga dikenal dapat memberikan kinerja yang optimal dalam meningkatkan keakuratan klasifikasi[11]. Dengan menggunakan algoritma seleksi fitur *Chi Square*, data yang besar dapat direduksi sehingga proses yang dibutuhkan dapat berjalan lebih cepat. Pada penelitian[12], menunjukkan bahwa dengan menggunakan *Chi Square* kinerja sistem meningkat dibandingkan sistem tanpa melalui tahapan seleksi fitur. Pada penelitian[11], analisis sentimen menggunakan *Random Forest* dan seleksi fitur *Chi Square* menghasilkan nilai akurasi yang tinggi yaitu sebesar 88.29%. Dari uraian latar belakang di atas, maka penulis mengambil

judul Analisis Sentimen Ulasan Pengguna Aplikasi Grab menggunakan Metode *Random Forest*.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Berdasarkan uraian latar belakang permasalahan, diperoleh bahwa sentimen pada ulasan aplikasi Grab di *Google Play Store* perlu dianalisa secara baik. Maka dari itu, dibutuhkan sistem pemodelan yang dapat mengelompokkan sentimen dengan metode yang sesuai.

1.2.2 Metode Usulan

Solusi yang diberikan dalam penelitian ini adalah penggunaan metode *TF-IDF* sebagai ekstraksi fitur, *SMOTE* sebagai penyeimbang data, *Chi Square* sebagai seleksi fitur dan metode *Random Forest* sebagai metode analisis sentimen dari ulasan pengguna aplikasi Grab di *Google Play Store*.

1.2.3 Pertanyaan Penelitian

Penelitian ini memiliki pertanyaan yang diperoleh dari permasalahan dan metode usulan yang diuraikan sebelumnya yaitu:

1. Berapa nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* untuk proses klasifikasi ulasan pengguna aplikasi Grab?
2. Berapa nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* dengan menggunakan *SMOTE* untuk klasifikasi ulasan pengguna aplikasi Grab?
3. Berapa nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* dengan menggunakan seleksi fitur *Chi Square* untuk klasifikasi ulasan pengguna aplikasi Grab?
4. Berapa nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* dengan menggunakan seleksi fitur *Chi Square* dan *SMOTE* untuk klasifikasi ulasan pengguna aplikasi Grab?
5. Berapa nilai perbandingan jumlah data latih dan jumlah data uji yang menghasilkan nilai akurasi yang paling tinggi?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Tujuan penelitian ini yaitu:

1. Mengetahui nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* untuk proses klasifikasi ulasan pengguna aplikasi Grab.
2. Mengetahui nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* dengan menggunakan *SMOTE* untuk proses klasifikasi ulasan pengguna aplikasi Grab.
3. Mengetahui nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* dengan menggunakan seleksi fitur *Chi Square* untuk proses klasifikasi ulasan pengguna aplikasi Grab.
4. Mengetahui nilai akurasi, presisi, *recall* dan *f1-score* yang dihasilkan oleh metode *Random Forest* dengan menggunakan seleksi fitur *Chi Square* dan *SMOTE* untuk proses klasifikasi ulasan pengguna aplikasi Grab.
5. Mengetahui pengaruh perbandingan jumlah data latih dan jumlah data uji dalam meningkatkan nilai akurasi.

1.3.2 Manfaat Penelitian

Penelitian ini menghasilkan beberapa manfaat, di antaranya sebagai berikut:

1. Menghasilkan model yang dapat melakukan klasifikasi sentimen suatu ulasan sehingga hasilnya dapat digunakan untuk dianalisis sesuai kebutuhan.
2. Sebagai bahan rujukan untuk penelitian selanjutnya.

1.4 Batasan-Batasan

Dalam penelitian ini terdapat batasan-batasan yang diterapkan, diantaranya sebagai berikut:

1. Data penelitian ini berasal dari ulasan atau komentar pengguna aplikasi Grab di *Google Play Store* yang diambil pada tanggal 31 Januari 2024.
2. Data menggunakan Bahasa Indonesia.
3. Data yang digunakan hanya data yang relevan.
4. Data yang digunakan hanya ulasan yang mempunyai *rating* 1 dan 5.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini mencakup beberapa bab, yang disusun sebagai berikut:

- BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, perumusan masalah yang meliputi permasalahan, metode usulan, dan pertanyaan penelitian, kemudian selanjutnya terdapat tujuan dan manfaat dari penelitian, batasan-batasan penelitian dan sistematika penulisan.

- BAB 2 KAJIAN PUSTAKA

Bab ini berisi mengenai landasan teori atau referensi mengenai materi metode yang digunakan pada penelitian ini, diantaranya meliputi rujukan *paper* yang digunakan terkait analisis sentimen, *text preprocessing*, *TF-IDF*, *SMOTE*, *Chi Square*, *Random Forest*, *Confusion Matrix* dan penelitian terkait.

- BAB 3 METODE PENELITIAN

Bab ini berisi penjelasan tentang arsitektur penelitian dan skenario pengujian.

- BAB 4 HASIL DAN PEMBAHASAN

Bab ini berisi hasil penelitian, meliputi tahapan perancangan sistem disertai *code*, hasil *running*, kemudian terdapat tampilan yang dihasilkan, hasil pengujian sistem disertai analisisnya.

- BAB 5 PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya.