

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan komunikasi telah memicu munculnya berbagai tren dalam kehidupan sehari-hari, contohnya seperti aktivitas penggunaan layanan transportasi *online* yang kini merambah ke masyarakat. Aktivitas transportasi *online* tumbuh signifikan seiring dengan kebutuhan masyarakat untuk melakukan penggunaan transportasi, pengiriman barang, pengiriman makanan, dll. secara *online*. Hal ini menyebabkan penyedia jasa transportasi *online* berinovasi dengan membuat harga promo yang dapat menarik minat pelanggan.

Harga promosi merupakan harga suatu produk atau jasa yang diberikan untuk meningkatkan penjualan. Harga promosi yang ditawarkan biasanya lebih murah dari harga reguler yang memiliki tujuan untuk meningkatkan minat pembeli terhadap produk atau jasa yang ditawarkan. Perusahaan biasanya menggunakan harga promosi dengan tujuan yang berbeda-beda seperti : peluncuran produk baru; meningkatkan penjualan; menarik minat pelanggan baru; menjaga stabilitas penjualan; menunjukkan keunggulan produk dibandingkan dengan produk pesaing; dan menjaga loyalitas pelanggan. Salah satu perusahaan yang menyediakan harga promo adalah Gojek.

Pengguna layanan Gojek sangat dimudahkan karena beberapa alasan yaitu salah satunya penawaran yang memiliki keuntungan yang lebih murah dan lebih terjangkau. Aplikasi ini sangat menguntungkan, namun juga memiliki sisi negatif, sisi negatif dari pemberian harga promo dari Gojek adalah tidak meratanya pemberian harga promo pada pengguna. Hal ini membuat respon masyarakat terhadap pelayanan yang diberikan oleh penyedia jasa transportasi *online* pun bervariasi. Salah satu tempat untuk menaruh respon atau opini masyarakat terkait masalah ini adalah melalui media sosial.

Dalam beberapa tahun terakhir, pengguna situs jejaring sosial telah meningkat pesat. Saat ini, jejaring sosial menghasilkan banyak data. Dengan jejaring sosial, masyarakat bisa mudah untuk mengekspresikan pandangan dan

pendapat mereka tentang masalah mengenai keresahan saat menggunakan sebuah aplikasi tertentu. Salah satu media sosial yang banyak digunakan oleh masyarakat untuk mengekspresikan pendapatnya adalah Twitter.

Twitter memungkinkan pengguna untuk menulis dan mendiskusikan berbagai topik masalah yang terjadi. Tidak seperti beberapa kebutuhan media sosial lainnya, dimana kedua belah pihak sepakat untuk saling menghubungi, Twitter memungkinkan pengguna untuk melacak *posting* (tweet) dari pengguna lainnya yang belum memberikan persetujuan. Hal ini merupakan salah satu alasan mengapa Twitter banyak digunakan dan sebagai tempat untuk arus informasi (Facta *et al.*, 2017). Jutaan tweet di Twitter dapat digunakan untuk mengekstrak informasi yang berguna untuk berbagai bidang. Salah satu informasi yang dapat diperoleh dari analisis data Twitter adalah sentimen orang tentang suatu topik. Dengan memahami sentimen publik, masalah dapat dievaluasi untuk perbaikan. Berdasarkan hal tersebut, sentimen analisis adalah salah satu cara untuk memahami pendapat seseorang atau sekelompok orang yang menangani masalah, produk, layanan, atau kelompok tertentu. Analisis sentimen, juga disebut penambangan opini, yang berarti proses pemahaman. Secara otomatis mengekstrak dan memproses data teks untuk mendapatkan informasi emosional yang terkandung dalam kalimat, apakah itu positif atau negatif.

Dalam penelitian ini, dataset Twitter akan digunakan dan dianalisis menggunakan pembobotan TF-IDF. Selanjutnya data tweet akan diklasifikasikan menggunakan metode *Random Forest* dan *K-Nearest Neighbors* yang memberikan deteksi positif dan negatif pada tweet. Lalu penulis akan membandingkan diantara metode tersebut, metode manakah yang lebih baik untuk mengolah data promo Gojek pada Twitter.

Pada penelitian ini, penulis akan menggunakan metode *Random Forest* dan *K-Nearest Neighbors*. *Random Forest* merupakan gabungan dari sebuah *decision tree* dan juga sebuah algoritma yang digunakan untuk mengolah data dengan skala yang besar. Semakin banyak pohon yang terbentuk, maka akan mempengaruhi hasil akurasi menjadi lebih baik. Pembentukan pohon akan dilakukan dengan menarapkan metode *random feature selection* yang berfungsi untuk meminimalisir

kesalahan. Sedangkan *K-Nearest Neighbors* merupakan algoritma yang berfungsi untuk melakukan klasifikasi pada suatu data berdasarkan *train dataset*, lalu diambil dari k tetangga terdekat. Metode ini mengklasifikasikan data berdasarkan *similarity* atau kemiripan atau kedekatan dengan data lainnya. *K-Nearest Neighbors* juga merupakan metode yang sederhana dan efektif jika data pelatihan besar.

Pada penelitian ini, metode *Random Forest* dan *K-Nearest Neighbors* akan dibandingkan menggunakan evaluasi model dan dilihat metode mana yang terbaik dalam melakukan pengolahan klasifikasi data promo Gojek. Kedua metode tersebut dibandingkan karena pada banyak penelitian yang sudah dilakukan, metode *Random Forest* dan *K-Nearest Neighbors* sama-sama menghasilkan nilai evaluasi yang sangat tinggi, sehingga pada penelitian ini akan membuktikan metode manakah yang terbaik untuk mengolah data promo Gojek.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya maka dapat dibuat rumusan masalah utama dari penelitian ini adalah sebagai berikut :

- a. Bagaimana penerapan dan perbandingan metode *K-Nearest Neighbors* dan *Random Forest* untuk melakukan analisis sentimen pada media sosial Twitter terhadap promo Gojek?
- b. Bagaimana performa metode *Random Forest* dan *K-Nearest Neighbors* terhadap opini masyarakat mengenai promo Gojek?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut :

- a. Menerapkan dan membandingkan metode *Random Forest* dan metode *K-Nearest Neighbors* dalam melakukan analisis sentimen pada media sosial Twitter terhadap promo Gojek.
- b. Mengetahui performa metode *Random Forest* dan *K-Nearest Neighbors* terhadap opini masyarakat mengenai promo Gojek

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

- a. Mengetahui cara kerja metode *Random Forest* dan *K-Nearest Neighbors* dalam melakukan analisis sentimen terhadap promo Gojek.

- b. Mengetahui hasil performa dari metode *Random Forest* dan *K-Nearest Neighbors*.
- c. Mengetahui perbandingan metode *Random Forest* dan metode *K-Nearest Neighbors*.

1.5 Ruang Lingkup

Ruang lingkup pada penelitian ini adalah sebagai berikut:

- a. Data yang digunakan adalah data tweet berbahasa Indonesia yang membahas tentang promo Gojek.
- b. Metode klasifikasi yang digunakan pada penelitian ini adalah *Random Forest* dan *K-Nearest Neighbors*.
- c. Sentimen akan diklasifikasi menjadi 2, yaitu positif dan negatif.
- d. Jumlah tweet yang digunakan sebanyak 453 tweet.
- e. Bahasa pemrograman yang digunakan adalah Python.

1.6 Luaran Yang Diharapkan

Luaran yang diharapkan dari penelitian ini adalah sebuah komparasi yang akan membuktikan, manakah yang lebih baik di antara metode *K-Nearest Neighbors* dan *Random Forest* terhadap sentimen mengenai promo Gojek.

1.7 Sistematika Penulisan

Penelitian ini disusun dalam beberapa bab yang masing-masing menjabarkan beberapa pokok pembahasan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menjelaskan latar belakang, rumusan masalah, ruang lingkup penelitian, tujuan penelitian, manfaat, luaran yang diharapkan dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada bab 2 memaparkan berbagai macam landasan teori yang digunakan sebagai acuan dalam menganalisis dan memecahkan masalah yang mendukung tema penulisan ini.

BAB 3 METODOLOGI PENELITIAN

Pada bab 3 membahas tahapan pada penelitian yang akan dilakukan dan menjelaskan langkah-langkah penelitian agar tercapai tujuan penelitian ini.

BAB 4 HASIL DAN PEMBAHASAN

Pada bab 4 menggambarkan langkah-langkah dari proses yang dilakukan dalam penelitian terhadap masalah yang diangkat untuk mencapai hasil dan tujuan yang ingin diteliti.

BAB 5 PENUTUP

Bab penutup menjelaskan kesimpulan yang didasarkan pada hasil dari proses penelitian yang telah dilakukan, serta saran bagi penelitian selanjutnya.

DAFTAR PUSTAKA

RIWAYAT HIDUP

LAMPIRAN-LAMPIRAN