

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Transportasi yang sedang berkembang di Indonesia selama ini semakin pesat. Fenomena perkembangan teknologi terus berkembang setiap tahunnya. Saat ini yang paling ramai dibicarakan dan menjadi bahan diskusi banyak kalangan adalah pengembangan teknologi ke arah bisnis transportasi yang modern dengan menggunakan kecanggihan aplikasi di dunia virtual secara daring (Online).

Masyarakat saat ini sangat dimudahkan dengan adanya sarana transportasi ini, terutama untuk pemesanannya. Dimanapun dan kapanpun juga secara cepat dan real time, masyarakat mudah melakukan mobilisasi kemana saja dengan memiliki aplikasi ini. Bisnis yang memanfaatkan aplikasi virtual untuk memudahkan pemesanan sarana transportasi ini adalah bisnis Gojek dan Grab.

Seiring dengan berjalannya teknologi yang semakin maju setiap tahunnya, kreatifitas anak bangsa terus berinovasi dalam mengembangkan transportasi berbasis online. Aplikasi Gojek diluncurkan ke masyarakat Indonesia pertama kali pada tahun 2010, tepatnya di Kota DKI Jakarta. Hingga hari ini, aplikasi Gojek Indonesia telah diunduh dan dipakai oleh pengguna smartphone Android dan iOS hingga lebih dari 50 juta kali. Pun aplikasi gojek Indonesia sudah beroperasi di bidang transportasi online di lebih dari 50 kota yang tersebar di Indonesia.

GOJEK merupakan salah satu badan usaha yang mengkonsentrasikan salah satu bidang usahanya di sektor industri jasa transportasi yaitu layanan ojek online. Ojek adalah transportasi umum informal di Indonesia yang berupa sepeda motor. Disebut informal karena keberadaannya tidak diakui pemerintah dan tidak ada izin untuk pengoperasiannya. Penumpang biasanya satu orang namun kadang bisa berdua. Biasanya disertai dengan tawar menawar dengan sopirnya dahulu setelah itu sang sopir akan mengantar ke tujuan yang diinginkan penumpangnya.

Tak hanya GO-JEK, pada tahun 2014 silam perusahaan transportasi online dari negara tetangga (Malaysia) mulai memasuki negara Indonesia yang dikenal dengan nama GRAB. Bermula pada kerja sama dengan berbagai perusahaan taksi, pemilik GRAB ingin mengembangkan bisnisnya. Tak lagi sekadar bekerja sama dengan perusahaan taksi, ia juga ingin mengajak perusahaan rental mobil atau orang

pribadi yang ingin menjadikan mobilnya sebagai kendaraan bisnis untuk bergabung di GrabCar.

Sebagai layanan ojek online yang berbasis aplikasi, performa dari aplikasi GO-JEK dan GRAB pun dinilai oleh para penggunanya. Ulasan pengguna dapat dianalisa kemudian dijadikan masukan bagi perusahaan dalam menentukan pengembangan pada platform ojek online tersebut serta menjadi dasar pengguna untuk memilih menggunakan platform yang mana. Kita sebagai manusia dengan mudah memahami performa platform dengan melihat ulasan para penggunanya.

Media sosial kini sudah dijadikan wadah berbagi informasi dan mengekspresikan opini, bukan lagi hanya digunakan untuk mem-posting foto liburan atau bahkan meme lucu. Media sosial memungkinkan semua orang untuk saling berinteraksi satu sama lain. Salah satu media sosial yang populer dan banyak digunakan untuk menyampaikan opini adalah Twitter. Berdasarkan data dari PT Bakrie Telkom, Indonesia memiliki 19,5 juta pengguna Twitter dari total 500 juta pengguna global (Kominfo.go.id, 2013). Hal ini membuat Indonesia menempati peringkat lima pengguna Twitter terbesar di dunia.

Twitter merupakan platform yang memungkinkan penggunanya dapat mengirim dan membaca informasi ataupun opini terhadap isu-isu yang bermacam-macam dalam bentuk 280 karakter pesan twitter yang dinamakan tweet. Twitter dikenal sebagai sarana penyebaran informasi dan mengekspresikan opini. Tweet yang disebar oleh pengguna twitter ini mengandung konten yang berbeda-beda sesuai dengan kebiasaan dan gaya menulis pengguna. Konten yang disebar oleh pengguna dapat dikategorikan secara umum yaitu informasi/berita, opini, promosi, random tweet, obrolan, pernyataan dan tweet mengenai pengguna tersebut.

Analisis sentimen dapat menggunakan teknik data *mining*. Yang dapat mengklasifikasikan konten opini dari sumber data yang sangat banyak. Metode text mining untuk melakukan analisis terhadap tweet yaitu menggunakan algoritma naïve bayes dan Decision Tree (C4.5) karena proses analisis bersifat klasifikasi. Akurasi dalam klasifikasi yang dilakukan oleh Abdan Syakuro pada tahun 2017 menggunakan algoritma naïve bayes mencapai 88.8% dan pada penelitian yang dilakukan oleh Arif dan Rifqi pada tahun 2019 menggunakan algoritma c4.5 pada klasifikasi sentimen mencapai akurasi 80.00%, Kedua Algoritma tersebut memiliki

tingkat akurasi yang cukup tinggi. Hal ini merupakan alasan penelitian ini di buat, karena ingin menggabungkan output dari masing-masing algoritma untuk dianalisis tingkat akurasinya .

Algoritma C4.5 adalah salah satu sistem untuk membangun model klasifikasi yang umum digunakan (W.Xindong, 2007) . Algoritma ini sudah banyak digunakan dalam implementasi penentuan keputusan karena memiliki banyak kelebihan. Salah satu kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain (J. Purnomo, 2014).

Naive Bayes hanya membutuhkan jumlah data latih (*Training Data*) dalam jumlah kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari, 2012). Selain itu mudah di pahami dan pengkodean yang sederhana.

Penelitian ini menggunakan metode-metode tersebut karena Naive Bayes adalah metode pengklasifikasian yang populer, metode Naive Bayes adalah metode klasifikasi yang sangat terukur (Pak et al., 2014). Decision Tree (C4.5) merupakan metode pengklasifikasian data yang mudah diaplikasikan dan populer (Rane et al., 2018). Decision Tree (C4.5) sangat populer digunakan untuk klasifikasi data (Nugroho et al., 2017).

Dalam penelitian ini pesan-pesan twitter akan dikumpulkan terlebih dahulu. Proses pengumpulan data tersebut dinamakan crawling. Tweet yang telah dikumpulkan akan memasuki tahap preprocessing. Selanjutnya tweet akan memasuki tahap pembobotan data menggunakan information gain, agar dapat dianalisis dengan menggunakan algoritma C.5 dan Naive Bayes. Dan hasil dari masing-masing algoritma disimpan di database, lalu dalam penelitian ini akan membandingkan kedua hasil seleksi menggunakan metode Voting –by-majority yaitu dengan memberikan kelas baru pada setiap kelas yang terdapat pada sebuah dataset. hasilnya menggambarkan bahwa metode Voting –by-majority lebih baik dari yang individu (Alaa Khudhair Abbas, 2020).

Ulasan pengguna data seluler juga bisa digunakan sebagai sumber data analisis sentimen dengan memanfaatkan API Twitter serta metode Naïve Bayes. Dan hasil akhir yang dihasilkan ialah nilai akurasi sangat tinggi sebesar 99,66% (Febriyani et al., 2018). Selain ulasan pengguna, analisis sentimen juga bisa menggunakan sumber data lain seperti contoh data review film dengan menggunakan metode C4.5 mendapatkan akurasi sebesar 60% (Ngoc et al., 2017).

Berdasarkan uraian di atas, maka penulis tertarik menggunakan metode Naïve bayes dan Decision Tree (C4.5) untuk melakukan analisis sentimen terhadap ojek online menggunakan data twitter. Penelitian ini berfokus terhadap hasil perbandingan akurasi dari masing-masing metode yang penulis gunakan. Diharapkan nantinya penelitian ini dapat membantu perusahaan dalam hal pengembangan layanannya serta dapat menjadi referensi tambahan bagi masyarakat untuk memilih menggunakan layanan ojek online yang mana.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas, adapun permasalahan yang dikaji dalam penelitian ini adalah :

1. Bagaimana mengklasifikasi teks berdasarkan kelasnya (negatif dan positif) ?
2. Mengetahui algoritma mana yang paling baik dalam klasifikasi sentimen pada media Twitter di antara naïve bayes dan C4.5?

## **1.3 Batasan Masalah**

Berdasarkan rumusan masalah di atas maka permasalahan yang akan di bahas akan dibatasi dengan batasan masalah berikut :

1. Yang ingin dianalisis adalah sentimen positif dan sentimen negatif pada ulasan pengguna platform ojek online.
2. Pengambilan data menggunakan Twitter Streaming API. Tweet yang dianalisis mengandung kata nama GO-JEK dan GRAB.
3. Tweet yang dijadikan dataset adalah tweet yang diambil dari tanggal 1 Desember 2020 – 1 Agustus 2021.
4. Tweet yang digunakan adalah tweet yang menggunakan bahasa Indonesia sesuai kamus bahasa Indonesia (KBBI) dan bahasa Indonesia slang (alay).

5. Aplikasi ojek online yang dianalisis yaitu GO-JEK dan GRAB.
6. Metode yang digunakan penelitian ini adalah metode Naïve Bayes dan Decision Tree (C4.5).
7. Melihat dan membandingkan hasil nilai akurasi, presisi dan recall yang dihasilkan dari masing-masing metode.

#### **1.4 Tujuan**

- a. Untuk mengetahui performa Metode Naïve Bayes dan C4.5 (*Decision Tree*) dalam mengklasifikasi *tweet* sentimen mengenai gojek & grab.
- b. Untuk menjadi referensi masyarakat sebelum menggunakan platform ojek online.

#### **1.5 Manfaat**

Manfaat dari penelitian ini adalah :

- a. Mengetahui performa dari masing-masing metode.
- b. Mengetahui metode mana yang cocok untuk permasalahan ini.
- c. Membantu perusahaan platform ojek online agar mengetahui pendapat pengguna mengenai produk mereka.