

1. Pendahuluan

1.1 Latar Belakang

Transportasi *online* merupakan inovasi baru yang memanfaatkan kendaraan-kendaraan pribadi sebagai fasilitas layanan transportasi umum berbasis *online* yang dapat diorder oleh masyarakat melalui *smartphone*. Kemudahan dalam pemesanan serta sistem yang lebih transparan membuat transportasi *online* semakin banyak diminati oleh masyarakat. Masyarakat memberikan pendapat dan opini mengenai transportasi *online* melalui berbagai media sosial, salah satunya Twitter. Twitter merupakan salah satu media sosial yang akrab digunakan oleh masyarakat Indonesia. Berdasarkan penelitian Semiocast, jumlah pemilik akun Twitter di Indonesia menempati urutan kelima terbesar di dunia, dan berada pada posisi ketiga negara yang paling aktif mengirim pesan Twitter (*tweet*) perhari [1]. Dengan banyaknya pengguna disertai kemudahan pengaksesan Twitter dalam penyampaian opini, maka akan banyak data opini terkumpul yang kemudian bisa dimanfaatkan sebagai bahan penilaian dan evaluasi atas penggunaan transportasi *online*.

Untuk menghasilkan informasi dari data opini yang ada, maka perlu dilakukan pengolahan data Twitter dengan analisis sentimen, yang merupakan bidang *study* untuk menganalisis pendapat, sentimen, sikap dan emosi seseorang terhadap suatu entitas seperti isu, topik dan lain-lain [2]. Dalam penelitian ini dilakukan analisis sentimen dengan memisahkan opini ke dalam kelas bersentimen positif atau negatif serta ke dalam kategori penilaian kualitas layanan transportasi publik. Sebelumnya telah dilakukan penelitian mengenai analisis sentimen terhadap Twitter dengan berbagai metode seperti *Naive Bayes Classifier*, *Support Vector Machine* dan *K-Nearest Neighbor*. Didapatkan kesimpulan bahwa *Naive Bayes Classifier* menghasilkan performansi paling baik dibandingkan dengan metode yang lain [3] [4]. Atas dasar hal ini, metode yang akan digunakan dalam penelitian tugas akhir ini adalah *Naive Bayes Classifier* yang dikenal sebagai metode sederhana, cepat, tetapi memiliki performansi yang tinggi dalam pengklasifikasian teks [5].

Permasalahan yang sering muncul pada analisis sentimen adalah tingginya dimensi data (banyaknya fitur atau kata dalam satu dokumen) yang tentunya dapat mengganggu proses klasifikasi itu sendiri. Untuk mengurangi tingginya dimensi data perlu dilakukan pemilihan fitur atau *feature selection* yang dapat memberikan pengaruh terhadap hasil klasifikasi [6]. *Feature selection* merupakan proses untuk menyeleksi kata yang dianggap kurang informatif. Terdapat banyak metode *feature selection* seperti, *Chi-square*, *Information Gain*, *Chi-WSS* dan lain-lain. Keunggulan dari *Chi-square* adalah dapat meningkatkan performansi suatu kategori teks paling optimal dibandingkan dengan metode yang lain [6]. Oleh karena itu, pada penelitian tugas akhir ini akan dilakukan analisis sentimen Twitter mengenai penggunaan transportasi *online* menggunakan metode *Naive Bayes Classifier* dengan *feature selection Chi-square*.

1.2 Perumusan Masalah

1. Bagaimana mengimplementasikan *Naive Bayes Classifier* dan *feature selection Chi-square* pada klasifikasi sentimen dan klasifikasi kategori penggunaan transportasi *online*?
2. Bagaimana pengaruh *feature selection Chi-square* terhadap performansi sistem klasifikasi sentimen dan kategori penggunaan transportasi *online*?

1.3 Batasan Masalah

1. Dataset yang digunakan adalah pesan Twitter (*tweet*) berbahasa Indonesia dengan kata kunci @GrabID dan @gojekindonesia.
2. Terdapat dua kelas klasifikasi sentimen yaitu positif dan negatif.
3. Terdapat tujuh kelas klasifikasi kategori yaitu *Accessibility*, *Availability*, *Information*, *Time*, *Customer Service*, *Comfort* dan *Safety*.
4. Satu *tweet* hanya dapat diklasifikasikan dalam satu kelas sentimen dan satu kelas kategori.
5. Perhitungan *feature selection Chi-square* dilakukan berdasarkan kelas sentimen.
6. Tidak membandingkan metode *Naive Bayes Classifier* dengan metode klasifikasi lainnya.

1.4 Tujuan

Penelitian ini bertujuan untuk membuat analisis sentimen penggunaan transportasi *online* dengan mengklasifikasikan sentimen dan kategori mengenai penggunaan transportasi *online* pada Twitter menggunakan metode *Naive Bayes Classifier* dengan *feature selection Chi-square*. Selain itu penelitian ini bertujuan untuk mengukur performansi *Naive Bayes Classifier* dalam mengklasifikasikan sentimen dan kategori mengenai penggunaan transportasi online serta melihat pengaruh *feature selection chisquare* terhadap performansi sistem.

1.5 Sistematika Penulisan Tugas Akhir

Urutan penulisan laporan ini adalah sebagai berikut : Bagian 2 menunjukkan penelitian-penelitian terkait dengan tugas akhir ini. Sistem yang akan diajukan untuk analisis sentimen pada Twitter mengenai penggunaan transportasi *online* dengan metode *Naive Bayes Classifier* dan *feature selection Chi-square* akan dijelaskan pada bagian 3. Pada bagian 4 akan didiskusikan mengenai hasil pengujian dan evaluasi sistem. Dan kesimpulan akan dijelaskan pada bagian 5.