



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam dekade terakhir. Twitter telah menjadi sumber yang signifikan untuk mendapatkan *user-generated-data*. Jumlah pengguna aktif bulanan Twitter mencapai 330 juta pada kuartal ketiga 2017, dan jumlah pengguna aktif harian yang berjumlah 157 juta pada kuartal kedua 2017. Selain itu, hampir 500 juta tweet per hari dibagikan di Twitter. Dengan demikian, kemajuan teknis yang signifikan telah dibuat untuk memproses dan menganalisis data dari media sosial menggunakan teknik dari berbagai bidang, seperti pembelajaran mesin, pemrosesan bahasa alami, statistik dan web semantik. Data yang didapat dari Twitter dapat digunakan untuk melakukan analisis prediktif di berbagai bidang, mulai dari pribadi, sosial hingga kesehatan masyarakat dan politik (Kursuncu dkk, 2018). Di Twitter, setiap *user* bebas mengungkapkan pendapat mereka, salah satunya yaitu dengan mengulas suatu *brand* berdasarkan apa yang telah *user* alami dengan *brand* tersebut.

Analisis sentimen, dalam kaitannya dengan ulasan pelanggan, dibangun berdasarkan premis bahwa informasi yang diberikan melalui teks dapat bersifat subjektif atau objektif. Ulasan subjektif didasarkan atas opini, perasaan pribadi, kepercayaan, dan penilaian tentang suatu entitas atau peristiwa. Sedangkan, ulasan objektif didasarkan pada fakta, bukti, dan pengamatan yang terukur (Feldman, 2013). Ulasan konsumen dan posting media sosial sering mencerminkan kebahagiaan, frustrasi, kekecewaan, kegembiraan, dan perasaan lainnya (O'Leary, 2011).

Salah satu cara untuk mendapatkan *user-generated-data* adalah menggunakan Text mining. Text mining merupakan sebuah metode turunan dari *Data Mining* yang bertujuan untuk mencari pola atau informasi menarik dari sekumpulan data yang berbentuk *natural language text* (Adib dkk, 2015). Dengan mendapatkan data dari Twitter dan mengolahnya, perusahaan transportasi *online* bisa mendapatkan *brand image* mereka di media sosial Twitter, dan dapat merencanakan langkah bisnis selanjutnya berdasarkan analisis sentimen yang didapat.

Gojek merupakan aplikasi yang berasal dari Indonesia yang membantu pengguna aplikasi untuk melakukan pemesanan terhadap *motorcycle taxi* atau biasa disebut dengan ojek. Sebagai *startup* lokal pertama yang berasal dari Indonesia yang memberikan pelayanan seperti diatas, gojek memiliki *brand recognition*, dan *user* di berbagai kota besar di Indonesia. Gojek telah di unduh oleh 10 juta pengguna. Banyak fitur didalam aplikasi gojek dapat membuat *multiplier effect* untuk industry dan ekonomi (Adhim dkk, 2018).

Requirements Engineering (RE) digunakan untuk menggambarkan proses yang sistematis dari pengembangan berdasarkan persyaratan melalui kerja sama iteratif antara menganalisis masalah, mendokumentasikan hasil pengamatan dalam berbagai format representasi, dan memeriksa keakuratan pemahaman yang diperoleh (Satria, 2003). Requirements Engineering (RE) adalah tahap pertama dari proses rekayasa perangkat lunak, di mana persyaratan pengguna dikumpulkan, dipahami, dan ditentukan (Pandey & Pandey, 2012). Dalam kaitannya dengan penelitian kali ini, persyaratan pengguna dikaitkan dengan tweet pada Twitter, dimana pada penelitian kali ini akan diambil data tweet pada perusahaan penyedia jasa transportasi online. Tweet yang di kirimkan ke akun resmi perusahaan

penyedia jasa transportasi online yang berupa keluhan atau kekecewaan atau buruknya sistem yang dibuat perusahaan penyedia jasa transportasi online akan digunakan untuk melakukan sentimen analisis.

Pada penelitian sebelumnya, analisis sentimen pada Twitter telah diselesaikan dengan algoritma Naïve Bayes (Brata & Muslim, 2018). Algoritma Naïve Bayes banyak sekali digunakan dalam permasalahan analisis sentimen karena memiliki performa dan tingkat akurasi yang lebih baik ketimbang algoritma pengklasifikasian lainnya yaitu K-Nearest Neighbor, K-Means, dan LVQ (Brata & Muslim, 2018).

Pada penelitian kali ini, analisis sentimen pada Twitter akan dilakukan dengan menggunakan algoritma Multinomial Naïve Bayes. Lalu akan dilakukan *text preprocessing* terhadap dataset dengan tahapan *stopword removal*, *case folding*, *filtering*, *stemming* dan *tokenizing*. Merujuk pada penelitian sebelumnya oleh (Sari, 2018), dataset yang digunakan berjumlah 700 data. Dataset yang digunakan selanjutnya akan dilakukan pengklasifikasian dan uji performa dengan menggunakan *confusion matrix* untuk mendapat skor *accuracy*, *precision*, *recall* dan *f-measure*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah yang diangkat adalah :

1. Bagaimana mengimplementasikan *Multinomial Naive Bayes* untuk melakukan analisis sentimen jasa transportasi *online* pada media sosial Twitter?

2. Bagaimana hasil *accuracy, precision, recall, dan F-measure* dari metode pengklasifikasian teks berbasis Multinomial Naïve Bayes untuk melakukan analisis sentimen jasa transportasi *online* pada media sosial Twitter.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Sistem analisis sentimen ini ditujukan untuk *tweets* dari pengguna transportasi online.
2. Dataset yang digunakan dalam penelitian ini diambil menggunakan fitur Twitter API dengan menggunakan kata kunci @gojekindonesia.
3. Dataset yang digunakan berjumlah 700 menurut Sari (2018).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Mengimplementasikan algoritma *Multinomial Naive Bayes* untuk analisis sentimen jasa transportasi *online* pada media sosial Twitter.
2. Mengukur *accuracy, precision, recall, dan F-measure* dari metode pengklasifikasian teks berbasis Multinomial Naïve Bayes untuk melakukan analisis sentimen jasa transportasi online pada media sosial Twitter

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Mengetahui tingkat kepuasan masyarakat terhadap jasa transportasi online.
2. Mengetahui performa algoritma *Multinomial Naive Bayes* untuk permasalahan klasifikasi.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan laporan skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini menjelaskan terkait latar belakang dari penelitian ini, rumusan masalah, Batasan masalah, tujuan masalah, dan manfaat penelitian.

BAB II LANDASAN TEORI

Bab ini menjelaskan terkait landasan teori yang digunakan pada penelitian ini. Teori yang digunakan antara lain, *sentiment analysis*, *text preprocessing*, *Multinomial Naïve Bayes Classifier*, dan *Confusion Matrix*.

BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab ini menjelaskan metode yang digunakan dalam merancang dan membangun sistem Metode yang digunakan adalah *flowchart* untuk training, testing.

BAB IV IMPLEMENTASI DAN UJI COBA

Bab ini menjelaskan mengenai implementasi dari perancangan yang sudah dibuat diikuti dengan *testing* terhadap algoritma *Multinomial Naïve Bayes* untuk analisis sentimen jasa transportasi *online* pada media sosial Twitter.

BAB IV SIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil uji coba dan analisis yang telah dilakukan dalam penelitian, serta saran untuk pengembangan sistem