

Analisis Sentimen Pada Twitter Terhadap Transportasi Online di Indonesia Menggunakan Ensemble Stacking

Yahya Setiawan¹, Jondri², Widi Astuti³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹yahyasetiawan@students.telkomuniversity.ac.id, ²pembimbing1@telkomuniversity.ac.id,

³pembimbing2@telkomuniversity.ac.id

1. Pendahuluan

1.1 Latar Belakang

Semakin populernya transportasi online tidak sedikit dari perusahaan transportasi online membuat akun media sosial untuk memberitahukan promo atau inovasi terbaru dari produk atau layanan mereka, salah satu media sosial yang populer dimanfaatkan adalah twitter [1]. Twitter merupakan media sosial yang populer di dunia, menurut data yang ada terdapat 500 juta pengguna aktif media sosial tersebut dan lebih dari 500 juta tweet per harinya [1]. Banyak orang yang menuliskan pendapat, pikiran, perasaan serta protes mereka di twitter, sehingga twitter dapat digunakan sebagai sumber data yang dapat di analisis [2]. Sumber data tersebut dapat digunakan untuk menganalisis sentimen seseorang mengenai suatu topik salah satunya yakni kepuasan pelanggan terhadap transportasi online [3]. Analisis sentimen adalah suatu teknik yang dapat menganalisis suatu pendapat atau opini individu ataupun kelompok terhadap produk [1]. Analisis sentimen dapat diekstrak secara otomatis melalui tulisan, dalam penelitian ini tweet pengguna akan digunakan untuk mengetahui keadaan emosionalnya [4]. Dalam penelitian yang dilakukan oleh Annisa Octaviana Sari dan Yuliant Sibaroni, analisis sentimen dapat dikategorikan menjadi positif, negatif dan netral [2]. Penelitian pada analisis sentimen terhadap pengguna jasa transportasi pernah dilakukan sebelumnya dengan menggunakan metode naive bayes. Analisis pada penelitian tersebut menggunakan TF-IDF dan information gain, dengan kesimpulan bahwa tingkat akurasi yang kurang baik [5]. Selanjutnya terdapat penelitian lain terhadap pengguna jasa transportasi yang menggunakan metode *Logistic regression* dengan kesimpulan bahwa metode tersebut cukup baik digunakan untuk melakukan klasifikasi data mining atau teks mining namun bisa ditingkatkan lagi dengan menggunakan pembelajaran mesin lanjut [6]. Terdapat metode lanjutan dari *Logistic regression* yakni metode ensemble stacking. Dalam implementasi terkait penelitian pengguna review film didapatkan metode tersebut dapat meningkatkan akurasi sebesar 20% lebih tinggi dari metode *Logistic regression* namun, hasil tersebut sangat dipengaruhi oleh keheterogenan metode dasar dan ketidak seimbangan dataset [4]. Berdasarkan permasalahan dan penelitian sebelumnya di rancanglah sistem yang digunakan untuk menganalisis sentimen jasa transportasi online di Indonesia pada media sosial twitter. Metode yang digunakan ensemble staking dengan fitur TF-IDF.

1.2 Topik dan Batasannya

Berdasarkan permasalahan dan penelitian sebelumnya di rancanglah sistem yang digunakan untuk menganalisis sentimen jasa transportasi online di Indonesia pada media sosial twitter. Metode yang digunakan ensemble staking dengan tambahan fitur seperti TF-IDF yang akan membantu meningkatkan kinerja sistem analisis sentiment Berdasarkan uraian permasalahan di latar belakang, berikut rumusan masalah dalam penelitian ini:

1. Bagaimana desain sistem analisis sentimen pengguna jasa transportasi online menggunakan metode ensemble stacking?
2. Bagaimana hasil tingkat akurasi analisis sentimen terhadap transportasi online di indonesia dengan metode ensemble stacking?

Penulis menyadari dengan adanya batasan masalah ini ditujukan agar tugas akhir akan lebih terarah dan dipermudahkannya dalam pembahasannya. Berikut ini batasan batasan yang ada pada tugas akhir ini:

1. Crawling data dilakukan berdasarkan tweet pengguna yang melakukan mention ke akun official transportasi Gojek dan Grab.
2. Hanya membandingkan 2 *meta classifier* untuk mengetahui mana yang paling baik dari tingkat akurasi
3. Dalam implementasinya penulis hanya melakukan percobaan dimana peserta didik tidak *heterogen*.

1.3 Tujuan

Berikut ini Tujuan dari yang ingin dicapai dalam penulisan tugas akhir ini:

1. Mendesain sistem analisis sentimen pengguna jasa transportasi online menggunakan metode *ensemble stacking*.
2. Mengetahui tingkat akurasi pada metode *ensemble stacking* dari desain yang diterapkan.
3. Melakukan perbandingan tingkat akurasi dari dua *meta classifier* yang berbeda.

1.4 Organisasi Tulisan

Penelitian ini dibagi menjadi 5 bagian. Bagian pertama mengenai latar belakang mengapa penelitian dilakukan. Pada bagian kedua dijelaskan mengenai studi terkait dengan penelitian. Pada bagian ketiga dijelaskan mengenai sistem yang dibangun. Pada bagian keempat dijelaskan mengenai hasil dan evaluasi dari sistem yang telah dibangun. Pada bagian kelima dijelaskan hasil penelitian yang telah dilakukan.