

1. Pendahuluan

Latar Belakang

Di masa digitalisasi seperti saat ini sejumlah kabar kini disampaikan melalui jaringan. Umum bagi berbagai kalangan untuk mencari berbagai informasi mengenai produk dan jasa yang ingin mereka gunakan melalui berbagai media digital karena melalui media tersebut disajikan opini subjektif serta informasi yang objektif. Namun dikarenakan sebagian besar informasi yang ada dalam bentuk teks bukan dalam skala numerik, evaluasi informasi ini lebih sulit untuk diklasifikasikan tanpa membaca teks secara lengkap [1].

Naïve bayes merupakan teknik *machine learning* yang umum untuk klasifikasi teks serta memiliki hasil yang baik pada banyak domain, hal tersebut dibuktikan oleh S. Habib et al. dalam penelitiannya yang berjudul Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier didapatkan bahwa metode naïve bayes classifier terbukti akurat digunakan dalam proses klasifikasi berita dimana proses klasifikasi menggunakan metode naïve bayes classifier menghasilkan akurasi tertinggi yaitu 94% dengan pembagian data latih 90% dan 10% data uji dari dataset berita yang digunakan [2]. Selain itu, keunggulan naïve bayes adalah dapat mengklasifikasikan teks dengan sangat sederhana dan efisien [3]. Namun, naïve bayes memiliki kekurangan yaitu sangat sensitif dalam pemilihan fitur [4]. Waktu perhitungan akan meningkat dan akurasi klasifikasi juga akan menurun apabila jumlah fitur terlalu banyak [5].

Hal lain yang terdapat pada analisis sentimen adalah pemilihan fitur. Adanya pemilihan fitur klasifikasi teks dapat menjadi lebih efektif dan efisien dengan mengurangi jumlah data yang dianalisis serta mengidentifikasi fitur yang sesuai untuk dipertimbangkan dalam proses pembelajaran. Terdapat dua jenis metode utama pemilihan fitur dalam *machine learning* yaitu wrapper dan filter. Sebagai fungsi evaluasinya, wrapper menggunakan akurasi klasifikasi dari beberapa algoritma. Wrapper bekerja dengan fitur yang di evaluasi secara berulang sehingga menghasilkan akurasi klasifikasi yang tinggi. *Genetic Algorithm* (GA) merupakan salah satu metode wrapper yang bisa digunakan dalam pemilihan fitur.

Topik dan Batasannya

Algoritma naïve bayes sangat sederhana, efisien dan merupakan teknik *machine learning* yang umum untuk klasifikasi teks, serta memiliki performa yang baik pada banyak domain. Namun, naïve bayes sangat sensitif pada fitur yang terlalu banyak, sehingga mengakibatkan akurasi klasifikasi menjadi rendah. Pada penelitian ini akan dibandingkan naïve bayes sebelum menggunakan metode pemilihan fitur dan setelah menggunakan fitur *genetic algorithm* yang diterapkan untuk mengklasifikasikan review hotel untuk meningkatkan akurasi analisis sentimen dan dilihat bagaimana pengaruh metode pemilihan fitur *genetic algorithm* pada akurasi analisis sentimen review hotel dengan teks bahasa Indonesia menggunakan algoritma naïve bayes. Batasan masalah dari latar belakang masalah tersebut adalah komentar pengguna didapat dari situs Kaggle.com, Bahasa yang digunakan dalam komentar pengguna adalah Bahasa Indonesia, Bahasa pemrograman yang digunakan yaitu Bahasa Python.

Tujuan

Tujuan dari penelitian ini adalah untuk mengetahui pengaruh metode pemilihan fitur *genetic algorithm* dalam menganalisa sentimen pada review hotel dengan teks Bahasa Indonesia menggunakan algoritma naïve bayes. Begitu juga beberapa manfaat penelitian, yaitu: Membantu konsumen dalam pengambilan keputusan saat hendak mencari hotel agar bisa menurunkan waktu dalam membaca review dan komentar terhadap suatu hotel dan membantu para pengembang sistem yang berkaitan dengan review hotel, baik dari sosial media lainnya seperti Twitter, Blog, dan lain-lain.