

Optimasi Naïve Bayes dengan Mengimplementasikan Algoritma Genetik pada Analisis Sentimen Review Hotel dengan Teks Bahasa Indonesia

Andri Marwan Tua Silalahi¹, Mahendra Dwifebri Purbolaksono², Moch. Arif Bijaksana³

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

¹andrimsilalahi@student.telkomuniversity.ac.id, ²febhiuss@telkomuniversity.ac.id,
³arifbijaksana@telkomuniversity.ac.id

Abstraksi

Di era web seperti sekarang, sejumlah informasi kini mengalir melalui jaringan. Karena berbagai konten web meliputi opini subjektif serta informasi yang objektif, saat ini umum bagi orang-orang untuk mengumpulkan informasi tentang produk dan jasa yang mereka ingin beli. Namun karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap. Analisa sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan review pengguna menjadi opini positif atau negatif. Pengklasifikasi naive bayes adalah teknik *machine learning* yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, naive bayes memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah. Oleh karena itu, dalam penelitian ini digunakan metode pemilihan fitur, yaitu *genetic algorithm* agar bisa meningkatkan akurasi pengklasifikasi naive bayes. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari review hotel. Pengukuran berdasarkan akurasi naive bayes sebelum dan sesudah penambahan metode pemilihan fitur. Hasil penelitian menunjukkan peningkatan akurasi naive bayes dari 93.43% menjadi 97.47%.

Kata Kunci: Analisis sentimen, algoritma genetik, review hotel, naive bayes, klasifikasi teks

Abstract

In the era of the web as it is now, a number of information is now flowing through the network. Since various web content includes subjective opinions as well as objective information, it is now common for people to gather information about the products and services they wish to purchase. However, because quite a lot of information exists in text form without any numerical scale, it is difficult to classify evaluations of information efficiently without reading the full text. Sentiment analysis aims to address this issue by automatically categorizing user reviews into positive or negative opinions. The naive bayes classifier is a popular machine learning technique for text classification, because it is very simple, efficient and performs well in many domains. However, naive bayes has the disadvantage that it is very sensitive to too many features, which results in low classification accuracy. Therefore, in this study, a feature selection method, namely genetic algorithm, was used in order to increase the accuracy of the naive bayes classifier. This research produces text classification in the form of positive or negative from hotel reviews. Measurements based on naive bayes accuracy before and after the addition of the feature selection method. The results showed an increase in the accuracy of naive bayes from 93.43% to 97.47%.

Keywords: Sentiment analysis, genetic algorithm hotel reviews, naive bayes, text classification.
