

ABSTRACT

A Comparative Study of Naïve Bayes Classifier and Semi-Supervised Labeling for Prediction Hotel Cleanliness Based on the Sentiment of User Review

By

Aditya Prasetyo Putra
19/444301/PA/19363

The development of internet review sites has had a big impact on the hospitality sector lately. One of the key elements affecting a hotel's reputation is cleanliness. It can be helpful to categorize hotel reviews for a variety of reasons. For instance, it can assist hotels and other hospitality firms in understanding the thoughts and feelings of their clients, pinpointing potential improvement areas, and delivering better client service. Sentiment analysis is a technique for disentangling users' feelings and emotions intellectually. Use sentiment analysis to categorize hotel/housing cleanliness into three polarities using semi-supervised learning (clean, unclean, null) on an online housing platform that provides a range of services. Following the preparation of the data, we will determine how clean a hotel will be based on guest reviews using pseudo labeling and semi-supervised learning algorithms Label Propagation and Label Spreading, which will be trained using the Naive Bayes algorithm. It will then be assessed using metrics including accuracy, precision, recall, and f1-score in conjunction with a confusion matrix to evaluate the modeling.

INTISARI

Studi Perbandingan Naïve Bayes Classifier dan *Semi-Supervised Labeling* untuk Prediksi Kebersihan Hotel Berdasarkan Sentimen Ulasan Pengguna

By

Aditya Prasetyo Putra
19/444301/PA/19363

Perkembangan situs review internet memberikan dampak yang besar pada sektor perhotelan akhir-akhir ini. Salah satu elemen kunci yang mempengaruhi reputasi hotel adalah kebersihan. Akan sangat membantu untuk mengategorikan ulasan hotel karena berbagai alasan. Misalnya, ini dapat membantu hotel dan perusahaan perhotelan lainnya dalam memahami pemikiran dan perasaan klien mereka, menunjukkan area peningkatan potensial, dan memberikan layanan klien yang lebih baik. Analisis sentimen adalah teknik untuk mengurai perasaan dan emosi pengguna secara intelektual. Dengan menggunakan analisis sentimen untuk mengategorikan kebersihan hotel/perumahan menjadi tiga polaritas menggunakan *semi-supervised learning* (bersih, tidak bersih, *null*) pada platform perumahan online yang menyediakan berbagai layanan. Setelah menyiapkan data, kami akan menentukan tingkat kebersihan sebuah hotel berdasarkan ulasan tamu menggunakan pseudo labeling dan algoritma pembelajaran *semi-supervised Label Propagation* dan *Label Spreading*, yang ketiganya akan dilatih menggunakan algoritma Naïve Bayes. Hasil dari algoritma tersebut kemudian akan dinilai menggunakan *confusion matrix* termasuk akurasi, presisi, daya ingat, dan skor-f1 untuk mengevaluasi model.