

ABSTRAK

Hotel merupakan tempat tinggal umum bagi wisatawan yang memberikan jasa penginapan, makanan, minuman dan akomodasi, dimana pengguna hotel perlu membayar. Dan saat ini sudah banyak website yang menyediakan layanan untuk pemesanan kamar hotel. Pemesanan hotel Royal Ambarrukmo juga dapat dilakukan melalui website seperti tripadvisor. Dalam website itu pengguna juga dapat memberikan komentar atau tanggapan mengenai hotel. Dari website itu didapatkan data ulasan hotel Royal Ambarrukmo sebanyak 1521 data yang kemudian diolah. Data tersebut melalui beberapa tahap diantaranya labelling untuk memberikan label sentimen, kemudian tahap preprocessing, kemudian perhitungan bobot dengan TF-IDF dan pembagian data training, testing dengan *cross validation*. Selanjutnya dilakukan klasifikasi menggunakan algoritma *Naïve Bayes* dan *Random Forest Classifier*. Pada algoritma *Naïve Bayes* dilakukan pengujian dengan mengubah parameter K-Fold dan alpha. Sedangkan pada algoritma *Random Forest Classifier* dilakukan pengujian dengan mengubah parameter K-Fold dan jumlah pohon.

Terdapat 2 skenario untuk Pengujian pada penelitian ini. Skenario Pengujian pertama dilakukan dengan tujuan untuk mengetahui tingkat akurasi algoritma yang lebih tinggi di antara algoritma *Random Forest* dan *Naïve Bayes* pada sentimen analisis Hotel Royal Ambarrukmo Yogyakarta. Skenario pengujian kedua dilakukan perhitungan nilai akurasi dari hasil sentimen netral dan negatif. Tujuan dari skenario penelitian tersebut yaitu melihat ketepatan hasil prediksi metode Vader Lexicon pada sentimen negatif dan netral.

Hasil percobaan skenario 1 tersebut menunjukkan nilai akurasi yang tertinggi algoritma *Naïve Bayes* yaitu 99,66% dengan parameter K-Fold = 3 dan parameter alpha = 0,1. Sedangkan pada algoritma *Random Forest* menunjukkan nilai akurasi yang tertinggi yaitu 99,86% dengan parameter K-Fold = 3 dan parameter jumlah pohon = 30. Dan pada skenario 2 hasil pengujian dengan data testing sentiment negatif menunjukkan akurasi untuk algoritma *Naïve Bayes* yaitu 100% dan algoritma *Random Forest* yaitu 60%. Sedangkan hasil pengujian dengan data testing sentiment netral menunjukkan akurasi untuk algoritma *Naïve Bayes* yaitu 100% dan algoritma *Random Forest* yaitu 37,5%.

Kata Kunci: Sentimen Analisis, *Web Scraping*, TF-IDF, *Vader Lexicon*, *Cross Validation*, Algoritma *Random Forest*, Algoritma *Naïve Bayes*, *Confusion Matrix*.

ABSTRACT

A hotel is a public residence for travelers that provides lodging, food, beverage and accommodation services, for which hotel users need to pay. And nowadays there are many websites that provide services for booking hotel rooms. Royal Ambarrukmo hotel reservations can also be made through websites such as Tripadvisor. On that website users can also provide comments or feedback about the hotel. From the website, the Royal Ambarrukmo hotel review data was obtained as much as 1521 data which was then processed. The data goes through several stages including labelling to provide sentiment labels, then the preprocessing stage, then calculating weights with TF-IDF and dividing training data, testing with cross validation. Furthermore, classification is carried out using the Naïve Bayes and Random Forest Classifier algorithms. The Naïve Bayes algorithm is tested by changing the K-Fold and alpha parameters. While the Random Forest Classifier algorithm is tested by changing the K-Fold parameter and the number of trees.

There are 2 scenarios for testing in this research. The first test scenario was carried out with the aim of knowing the higher accuracy of the algorithm between the Random Forest and Naïve Bayes algorithms in the sentiment analysis of the Royal Ambarrukmo Yogyakarta Hotel. The second test scenario calculated the accuracy value of the neutral and negative sentiment results. The purpose of the research scenario is to see the accuracy of the prediction results of the Vader Lexicon method on negative and neutral sentiments.

The results of the scenario 1 experiment show that the highest accuracy value of the Naïve Bayes algorithm is 99.66% with the K-Fold = 3 parameter and the alpha = 0.1 parameter. While the Random Forest algorithm shows the highest accuracy value of 99.86% with the K-Fold = 3 parameter and the number of trees = 30 parameter. And in scenario 2, the test results with negative sentiment testing data show the accuracy for the Naïve Bayes algorithm is 100% and the Random Forest algorithm is 60%. While the test results with neutral sentiment testing data show the accuracy for the Naïve Bayes algorithm is 100% and the Random Forest algorithm is 37.5%.

Keywords: Sentiment Analysis, Web Scraping, TF-IDF, Vader Lexicon, Cross Validation, Random Forest Algorithm, Naïve Bayes Algorithm, Confusion Matrix.