

ABSTRAK

Pandemi virus corona yang terjadi pada tahun 2020 menyebabkan kegiatan perkuliahan dilakukan secara daring untuk mencegah penyebaran virus corona. Perkuliahan yang dilakukan secara daring mendapat banyak opini dari masyarakat maupun mahasiswa. Banyaknya opini terkait perkuliahan daring dapat dilakukan analisis sentimen untuk mengetahui opini apa yang banyak disampaikan masyarakat dan mahasiswa. Proses analisis sentimen dilakukan menggunakan metode *Multinomial Naïve Bayes*. Data yang digunakan untuk proses analisis sentimen sebanyak 4.014 dengan kata kunci “kuliah daring dan belajar online”. Data akan dilakukan pembersihan data terlebih dahulu melalui proses *preprocessing* dengan 2 skenario. Pada proses *preprocessing* skenario pertama dilakukan menggunakan *stopword removal* dan *stemming* bahasa Indonesia. Pada proses *preprocessing* skenario kedua dilakukan menggunakan *stopword removal* dan *stemming* bahasa Inggris. Data yang telah dilakukan *preprocessing* akan diberi label menggunakan *textblob* dan dikategorikan ke dalam kelas positif, negatif, dan netral. Dalam penelitian ini juga akan dilakukan perbandingan hasil akurasi menggunakan seleksi fitur *Information Gain* dengan harapan dapat meningkatkan hasil akurasi. Berdasarkan hasil pengujian yang telah dilakukan seleksi fitur *Information Gain* terbukti dapat meningkatkan akurasi dalam proses analisis sentimen menggunakan *Multinomial Naïve Bayes*. Hasil akurasi tertinggi menggunakan skenario *preprocessing* bahasa Inggris sebesar 79.54%. Sedangkan hasil akurasi tertinggi menggunakan *preprocessing* bahasa Indonesia sebesar 68.62%.

Kata kunci : analisis sentimen, *information gain*, *multinomial naïve bayes*, *stemming*, *stopword removal*

ABSTRACT

The corona virus pandemic that occurred in 2020 caused lectures to be conducted online to prevent the spread of the corona virus. Lectures conducted online received many opinions from the public and students. The number of opinions related to online lectures can be done by sentiment analysis to find out what opinions are conveyed by the public and students. The sentiment analysis process is carried out using the Multinomial Naïve Bayes method. The data used for the sentiment analysis process is 4,014 with the keywords "online lectures and online learning". The data will be cleaned first through the preprocessing process with 2 scenarios. In the preprocessing process, the first scenario is carried out using stopword removal and stemming in Indonesian. In the second scenario preprocessing process is done using stopword removal and stemming in English. Data that has been preprocessed will be labeled using a textblob and categorized into positive, negative, and neutral classes. In this study, a comparison of the accuracy results using the Information Gain feature selection will also be carried out in the hope of increasing accuracy results. Based on the test results, the Information Gain feature selection has been proven to increase accuracy in the sentiment analysis process using Multinomial Naïve Bayes. The highest accuracy results using the English preprocessing scenario of 79.54%. While the results of the highest accuracy using Indonesian preprocessing of 68.62%.

Keywords: sentiment analysis, information gain, multinomial navie bayes, stemming, stopword removal