

INTISARI

Media sosial merupakan platform yang paling banyak diakses oleh pengguna internet di dunia. Twitter adalah salah satu media sosial yang populer di Indonesia dengan 24 juta pengguna. Oleh karena Twitter berbasis teks, membuat pengguna memiliki kebiasaan menuliskan berkomentar sebagai reaksi atas suatu topik atau fenomena yang terjadi di dunia. Komentar dalam jumlah besar dapat dikumpulkan untuk dijadikan sebuah dataset untuk dilakukan analisis sentimen. Analisis sentimen dilakukan untuk mengetahui bagaimana respon atau komentar pengguna Twitter atas suatu topik tertentu. Selain itu analisis sentimen dapat digunakan untuk menguji performa sebuah metode klasifikasi. Penelitian ini akan menguji performa metode klasifikasi Random Forest setelah melakukan pengumpulan data yang berasal dari Twitter dengan topik kriminalitas, karena topik tersebut selalu ada dan ramai diperbincangkan. Data tersebut akan diolah agar diketahui bagaimana sentimen pengguna Twitter pada topik tersebut. Data yang dapat dikumpulkan sebanyak 3500 data, dan selanjutnya data akan melewati tahap pre-processing seperti cleansing, case folding, tokenizing, stopword removal, normalisasi, dan stemming. Klasifikasi data dibagi ke dalam tiga kelas yaitu positif, netral, dan negatif. Adapun hasil akurasi model dengan metode Random Forest diperoleh sebesar 84%.

Kata kunci: analisis sentimen, Twitter, klasifikasi data, Random Forest

ABSTRACT

Social media is the most widely accessed platform by internet users in the world. Twitter is one of the most popular social media in Indonesia with 24 million users. Because Twitter is text-based, users have a habit of writing comments as a reaction to a topic or phenomenon that occurs in the world. A large number of comments can be collected to be used as a dataset for sentiment analysis. Sentiment analysis is done to find out how Twitter users respond or comment on a particular topic. The research aims to explore the effectiveness of the Random Forest algorithm in sentiment analysis. This research collects data from Twitter on the topic of crime, because the topic is always present and discussed. The data will be processed to find out how the sentiment of Twitter users on the topic. The data that can be collected is 3500 data, and then the data will go through pre-processing stages such as cleansing, case folding, tokenizing, stopword removal, normalization, and stemming. Data classification is divided into three classes, namely positive, neutral, and negative, using the Random Forest method. The accuracy of the model with the Random Forest method is 84%.

Keywords: sentiment analysis, Twitter, data classification, Random Forest