PERBANDINGAN METODE *RANDOM FOREST* DAN *K-NEAREST NEIGHBORS* TERHADAP ANALISIS SENTIMEN PADA TWITTER MENGENAI PROMO GOJEK

Ashil Hafidh Dhiya

ABSTRAK

Layanan transportasi *online* merupakan salah satu topik yang sedang hangat dibicarakan. Banyak orang mengandalkan lalu lintas *online*, atau pengemudi dan pelanggan. Transportasi *online* menjadi topik yang ramai diperbincangkan karena kebutuhan transportasi umum di beberapa daerah sulit terpenuhi. Respon masyarakat terhadap pelayanan yang diberikan oleh penyedia jasa transportasi *online* bervariasi, ada yang memberikan hasil positif dan negatif. Analisis sentimen merupakan salah satu cara untuk memahami pendapat seseorang atau sekelompok orang. Data berupa tweet akan dikumpulkan melalui Twitter sebanyak 453 tweet dan dibagi dua menjadi 80% data latih dan 20% data uji. Setelah itu data akan di *preprocessing* menggunakan *case folding*, *cleansing data*, normalisasi, *stemming*, *stopword removal* dan tokenisasi. Setelah data di *preprocess*, maka data akan diberi *term weighting* menggunakan TF-IDF. Setelah data diberi bobot, data akan diolah menggunakan *K-Nearest Neighbors* dan *Random Forest*. Penelitian ini diharapkan bisa mendapatkan informasi akan sentimen opini publik terhadap promo Gojek serta mengetahui performa metode *K-Nearest Neighbors* dan *Random Forest*. Pada pengujian pertama yaitu menggunakan metode *K-Nearest Neighbors* diperoleh nilai akurasi sebesar 75%, nilai presisi sebesar 80%, nilai recall sebesar 64%, dan nilai *f-1 score* sebesar 71%. Pengujian kedua yaitu dengan menerapkan metode Random Forest dan diperoleh nilai akurasi sebesar 77%, nilai presisi sebesar 75%, nilai recall sebesar 77%, dan nilai *f-1 score* sebesar 75%. Berdasarkan perolehan hasil evaluasi tersebut, metode *Random Forest* lebih baik dibandingkan dengan metode *K-Nearest Neighbors* dengan nilai akurasi sebesar 77%.

Kata Kunci: Promo Gojek, Analisis Sentimen, *K-Nearest Neighbors*, *Random Forest*

COMPARISON OF RANDOM FOREST AND K-NEAREST NEIGHBORS METHODS TOWARDS ANALYSIS OF SENTIMENT ON TWITTER ABOUT GOJEK PROMOTION

Ashil Hafidh Dhiya

ABSTRACT

Online transportation services are one of the hotly discussed topics. Many people rely on online traffic, or drivers and customers. Online transportation is a topic that is often discussed because public transportation needs in some areas are difficult to meet. Public responses to the services provided by online transportation service providers vary, some giving positive and negative results. Sentiment analysis is one way to understand the opinion of a person or group of people. Data in the form of tweets will be collected via Twitter as many as 453 tweets and divided into 80% training data and 20% test data. After that, the data will be preprocessed using case folding, data cleansing, normalization, stemming, stopword removal and tokenization. After the data is preprocessed, the data will be given term weighting using TF-IDF. After the data is weighted, the data will be processed using K-Nearest Neighbors and Random Forest. This research is expected to be able to obtain information on the sentiment of public opinion towards the Gojek promo and determine the performance of the K-Nearest Neighbors and Random Forest methods. In the first test, using the K-Nearest Neighbors method, the accuracy value is 75%, the precision value is 80%, the recall value is 64%, and the f-1 score is 71%. The second test is by applying the Random Forest method and obtained an accuracy value of 77%, a precision value of 75%, a recall value of 77%, and an f-1 score of 75%. Based on the evaluation results, the Random Forest method is better than the K-Nearest Neighbors method with an accuracy value of 77%.

Keyword: *Gojek Promotion, Sentiment Analysis, K-Nearest Neighbors, Random Forest.*