

ABSTRAK

Twitter merupakan salah satu media sosial yang banyak digunakan di Indonesia. Dengan adanya media sosial *Twitter*, banyak pengguna menjadikan *Twitter* sebagai tempat untuk mencari hiburan, mencari informasi atau berita terbaru, menuangkan opini atau pendapat yang dirasakan oleh pengguna. Dengan tingginya topik mengenai kuliah online maka dibutuhkan upaya untuk melakukan klasifikasi setiap *tweet* yang mengandung kata kunci ‘kuliah online’ sehingga dapat diketahui *tweet* tersebut positif, negatif atau netral. Data yang diambil berupa *Tweet* dengan kata kunci ‘kuliah online’ dengan menggunakan *twitter API*, kemudian data dilakukan pelabelan otomatis dan dilakukan *preprocessing*. Setelah melalui tahap *preprocessing* selanjutnya data dilakukan perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF). Setelah itu data yang telah diproses diklasifikasi dengan algoritma *Decision Tree* dan hasil dievaluasi menggunakan *confusion matrix* dan *k-fold cross validation*. Hasil penelitian menunjukkan rata-rata akurasi terbaik adalah 92.10%, dengan nilai *recall* sebesar 92.10%, presisi sebesar 90.09% dan F1-Score sebesar 90.18% dengan menerapkan SMOTE dan *Pruning*.

Kata kunci : Klasifikasi, Twitter, Tf-idf, Descision Tree.

ABSTRACT

Twitter is one of the most widely used social media platforms in Indonesia. Through Twitter, many users seek entertainment, access the latest news or information, and express their opinions or thoughts. With the increasing discussion about online lectures, it is essential to classify tweets containing the keyword 'online lecture' to determine whether the tweets are positive, negative, or neutral. The data is collected using the Twitter API with tweets containing the keyword 'online lecture,' followed by automatic labeling and preprocessing. After preprocessing, the data undergoes Term Frequency-Inverse Document Frequency (TF-IDF) calculation. The processed data is then classified using the Decision Tree algorithm, and the results are evaluated using the confusion matrix and k-fold cross-validation. The study results show that the best average accuracy is 92.10%, with a recall value of 92.10%, precision of 90.09%, and F1-score of 90.18%, achieved by applying SMOTE and pruning.

Keywords: Classification, Twitter, TF-IDF, Decision Tree