

Abstrak

Twitter merupakan media sosial dimana penggunanya dapat memposting suatu informasi berupa *tweet*. *Tweet* yang diposting oleh pengguna dapat dibagikan ulang oleh pengguna lain melalui *retweet*, *tweet* yang memiliki banyak *retweet* tersebar lebih cepat dan luas dibandingkan yang tidak mendapat *retweet*. Tetapi tidak semua *tweet* mendapatkan *retweet* karena terdapat faktor-faktor yang mempengaruhinya. Dalam penelitian ini fitur yang digunakan adalah fitur berbasis pengguna, berbasis waktu dan berbasis konten, serta menggunakan algoritma pembelajaran mesin Decision Tree (DT) jenis *classification and regression tree* (CART) untuk memprediksi kelas *retweet*. Masalah yang dihadapi pada penelitian ini adalah data yang tidak seimbang, lalu mengatasinya dengan *oversampling* dan *undersampling*. Setelah mengatasi data yang tidak seimbang performansi dari model meningkat terutama saat melakukan *undersampling* untuk *max_depth = 4* menghasilkan nilai akurasi dan f1 85%.

Kata kunci: *retweet, tweet, DT, CART, oversampling, undersampling*