

Abstrak

Tweet merupakan sebuah pesan berisikan informasi yang dibagikan oleh pengguna Twitter. *Tweet* dapat dibagikan kepada pengguna lain dengan cara *retweet*, sehingga fitur ini berperan sangat penting dalam penyebaran informasi. Penelitian ini membahas prediksi *retweet* menggunakan fitur berbasis pengguna dan fitur berbasis konten, serta ANN sebagai *classifier*. Masalah penelitian yang dikaji di sini adalah cara menanggulangi duplikasi data serta ketidakseimbangan kelas dengan menggunakan *undersampling* dan *oversampling*. Hasil evaluasi menunjukkan bahwa proses klasifikasi mencapai nilai F1 skor 86% pada model dengan melakukan *undersampling* dan menghapus duplikasi pada data.

Kata kunci: prediksi, *retweet*, *undersampling*, *oversampling*, ANN.