

Abstrak

Diabetes adalah penyakit yang menimbulkan ketidaknormalan berupa kadar glukosa yang tinggi pada darah dan menyebabkan generalisasi kerusakan pembuluh darah yang mempengaruhi jantung, mata, ginjal dan saraf serta mengakibatkan berbagai komplikasi. Identifikasi awal diabetes dapat dilakukan dengan menggunakan model untuk prediksi atau diagnosis pengidap diabetes. Pada penelitian ini, Random Forest digunakan untuk memprediksi pengidap diabetes berdasarkan konten ulasan obat. Ekstraksi fitur dilakukan dengan menggunakan N-gram dan *term frequency – inverse document frequency* (TF-IDF). dari Hasil dari percobaan yang dilakukan didapat akurasi model terbaik sebesar 95,2 % dengan ekstrasi fitur unigram.

Kata Kunci: klasifikasi, *random forest*, N-gram, TF-IDF
