

ABSTRAK

Ujaran kebencian atau *hate speech* merupakan tindakan komunikasi yang bertujuan untuk menghina, memprovokasi ataupun menghasut korban. Dengan bantuan media sosial seperti Twitter, menyebarkan ataupun menemukan ujaran kebencian menjadi lebih mudah. Maka untuk mengurangi ujaran kebencian tersebut dibuatlah sebuah sistem pengklasifikasian ujaran kebencian yang mengambil sampel data dari Twitter. Sistem ini menggunakan salah satu pengembangan dari algoritma *machine learning* Gradient Boosting Decision Tree (GBDT) yaitu LightGBM. GBDT merupakan algoritma yang sering digunakan namun masih memberikan hasil yang kurang memuaskan. Alasan utamanya adalah untuk setiap fitur, harus di-*scan* semua contoh data untuk memperkirakan informasi yang didapatkan dari semua kemungkinan titik perpecahan, yang sangat memakan waktu. Maka dari itu pada sistem ini digunakan LightGBM yang menggunakan teknik Gradient-based One-Side Sampling (GOSS) dan juga Exclusive Feature Bundling (EFB). Dengan diimplementasikannya kedua teknik ini pada GBDT, diharapkan hasil yang didapatkan tetap memiliki tingkat akurasi yang tidak berbeda jauh namun dengan waktu komputasi yang jauh lebih cepat.

Kata Kunci: Ujaran Kebencian, Twitter, *LightGBM*