
Abstrak

Tingginya jumlah email spam telah menyebabkan peningkatan triase email, menyebabkan kerugian sebesar USD 355 juta per tahunnya. Salah satu cara untuk mengurangi kerugian ini adalah dengan mengklasifikasikan email spam ke dalam kategori penipuan atau promosi yang dilakukan oleh pihak yang tidak diinginkan. Pengembangan awal klasifikasi email spam didasarkan pada metode sederhana seperti filter kata. Saat ini, metode yang lebih kompleks telah muncul seperti pemodelan kalimat menggunakan pembelajaran mesin. Beberapa metode yang paling terkenal untuk menangani masalah klasifikasi teks adalah jaringan dengan Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU). Studi ini berfokus pada klasifikasi email spam, sehingga metode LSTM dan GRU digunakan. Hasil penelitian ini menunjukkan bahwa, dalam skenario tanpa putus, LSTM dan GRU memperoleh nilai akurasi yang sama yaitu 99.02%, lebih unggul dari XGBoost, model dasar. Sementara itu, dalam skenario dropout, LSTM mengungguli GRU dan XGboost dengan masing-masing memperoleh akurasi 98.6%, 98.58%, dan 98.53% masing-masing. Skor GRU Recall lebih baik daripada LSTM dan XGBoost dalam skenario dengan dropout, masing-masing memperoleh nilai 98.98%, 98.92%, dan 98.14%. Dalam skenario tanpa dropout, LSTM lebih unggul dari GRU dan XGBoost, dengan masing-masing memperoleh nilai 98.39%, 98.39%, dan 98.15%.

Kata Kunci: GRU, LSTM, Spam Classification
