

Abstrak

Membaca berita bisa berdampak buruk terhadap kesehatan mental pembacanya. Oleh sebab itu diperlukan sistem yang dapat mengklasifikasikan berita menjadi good, bad, dan neutral agar pembaca dapat menghindari konsumsi berlebih terhadap berita yang berlabel bad. Dataset ini diklasifikasikan menggunakan algoritma *Long-Short Term Memory* yang merupakan turunan algoritma *Recurrent Neural Network*. Algoritma ini dipilih sebagai metode klasifikasi karena berita ini berfungsi dengan baik pada data yang memiliki informasi yang cukup besar seperti artikel berita. Peningkatan jumlah dataset sebanyak 5000 dilakukan setelah melihat penelitian sebelumnya hanya menggunakan dataset sebanyak 300 yang menyebabkan performa dari LSTM tidak dapat berkerja dengan baik. Penelitian ini akan melihat tren akurasi dari algoritma *Long-Short Term Memory* dalam pengklasifikasian setiap 500 data. Hal ini dilakukan untuk melihat bagaimana jumlah data mempengaruhi akurasi dari algoritma *Long-Short Term Memory*. Pada penelitian ini terdapat beberapa tahap dalam perancangan sistem yaitu *preprocessing*, *feature engineering*, dan *model training* menggunakan *Long-Short Term Memory*. Dari hasil penelitian didapatkan grafik tren akurasi meningkat seiring meningkatnya jumlah data dengan akurasi terbaik sebesar 76.58%.

Kata kunci : *dataset, preprocessing, feature engineering, recurrent neural network, long-short term Memory*