

1. Pendahuluan

Setiap harinya selalu ada berita tentang terorisme, bencana alam, kasus hak asasi manusia, dan kecelakaan transportasi. Kehadiran berita buruk ini lebih mendominasi populasi artikel berita. Studi menunjukkan bahwa berita bisa mempengaruhi keadaan mental dari pembacanya[1]. Membaca berita buruk akan menimbulkan efek negatif terhadap kesehatan mental jika terus dilakukan secara terus menerus[2].

Pada era digital dimana berita datang dari banyaknya sumber, diperlukan metode yang bisa mengkategorikan berita menjadi baik dan buruk. Suatu sistem yang dapat menyaring berita berdasarkan konten bisa membantu pembaca dalam mengendalikan jumlah berita buruk yang dibaca. Selain itu dengan adanya sistem ini maka pembaca bisa menghindari kelebihan konsumsi terhadap berita buruk.

Untuk implementasi sistem pengkategorian berita yang merupakan masalah pada *Natural Language Processing* (NLP) penulis sebelumnya menggunakan 300 artikel berita yang dikategorikan menjadi good, bad, dan neutral. Penulis sebelumnya menggunakan beberapa metode *machine learning* yaitu; *naïve bayes*, *Logistic regression*, *Support Vector Machine* dan *deep learning Long-Short Term Memory* (LSTM). Penelitian tersebut menunjukkan hasil terbaik didapatkan oleh metode *machine learning naïve bayes* dengan akurasi 82% dan hasil terburuk didapatkan metode *deep learning LSTM* dengan akurasi 59%[3].

Menggunakan algoritma *deep learning* dalam *Natural Language Processing* (NLP) sudah umum dilakukan. Dari sejumlah algoritma *deep learning*, *Recurrent Neural Network* (RNN) merupakan arsitektur yang paling lazim dipakai dikarenakan kemampuan dalam menangani naskah teks *variable-length* secara sekuensial. RNN Namun metode RNN biasa tidak cukup kuat dalam menangani permasalahan NLP dengan data berjumlah relatif besar. Dalam langkah sekuen yang panjang RNN akan menghadapi permasalahan ledakan dan difusi gradien[4].

Algoritma Long-Short Term Memories (LSTM) merupakan algoritma turunan RNN yang mampu mengatasi mengurangi permasalahan gradien. LSTM mampu menangkap dependensi panjang pada data sekuen dengan mekanisme gerbang. Dalam sistem LSTM terdapat 3 gerbang yaitu forget gate, input gate, dan output gate. Gerbang ini berfungsi untuk memutuskan bagaimana data dalam *memory cell* digunakan dan diperbarui[5].

Pada penelitian tugas akhir ini penulis akan melihat performa dari metode *deep learning Long-Short Term Memory* pada trend jumlah data tertentu dengan jumlah data yang lebih besar. Jumlah berita akan ditambah menjadi sebanyak 5000 artikel berita, yang sebelumnya hanya menggunakan 300 artikel berita. Penambahan jumlah dataset yang digunakan pada penelitian ini bertujuan untuk meningkatkan stabilitas performa dari *deep learning Long-Short Term Memory* yang berkerja lebih baik pada jumlah dataset yang besar. Dataset ini akan diuji dalam setiap range 500 data pada 5000 data dan diperhatikan bagaimana tren akurasi dari algoritma LSTM dalam setiap range.

Bagian selanjutnya akan membahas studi terkait dengan penelitian yang membahas studi literatur yang mendukung penelitian. Selanjutnya bagian Sistem yang dibangun akan menjelaskan tentang gambaran dan penjelasan dari implementasi yang dilakukan dalam penelitian. Setelah itu hasil implementasi akan di evaluasi dan dijelaskan pada bagian Evaluasi. Kemudian bagian Kesimpulan akan menjelaskan tentang rangkuman dari evaluasi serta saran untuk penelitian selanjutnya.