

BAB I

PENDAHULUAN

1.1 Latar Belakang

Informasi selalu hadir dalam kehidupan sehari-hari baik dari orang di sekitar kita maupun dari media massa atau media online. Dipengaruhi oleh kemajuan teknologi, volume informasi mengalami peningkatan yang besar[1]. Namun ketersediaan dokumen digital terutama informasi terkait berita terkini di internet yang berlimpah ini akan menyulitkan masyarakat untuk mengaksesnya jika dokumen tersebut tidak diatur secara layak. Pengaturan berita yang umum adalah dengan mengelompokkan berita-berita berdasarkan tema dari berita tersebut oleh tenaga khusus secara manual. Semakin berkembangnya kategori dan waktu rilis berita yang dituntut cepat akan menyulitkan *editor* untuk mengelompokkan berita, terutama pada artikel yang tidak terlalu berbeda secara jelas. Pengelompokan berita ini akan lebih efisien apabila pengklasifikasian berita diproses secara otomatis dengan komputer menggunakan metode tertentu hingga mengeluarkan hasil kategori terkait berita yang sudah dimasukkan.

Beberapa peneliti sudah melakukan penelitian terkait klasifikasi kategori berita dengan berbagai metode dan bahasa yang digunakan. Pada penelitian [2] menggunakan *Naive Bayes Classifier* dengan seleksi fitur menggunakan *Information Gain* dan *Bayesian Boosting*. Pada penelitian ini menghasilkan akurasi sebesar 73.2%. Namun penelitian tersebut masih menggunakan teknologi *machine learning* dan belum menerapkan pembelajaran yang lebih dalam atau lebih dikenal dengan istilah *deep learning*.

Deep learning adalah salah satu metode implementasi dari *machine learning* yang memiliki tujuan meniru cara kerja otak manusia. *Deep learning* menggunakan sejumlah *neuron* yang terhubung dalam menentukan karakteristik tertentu pada rangkaian data. Pada penelitian [3] melakukan klasifikasi berita dari *dataset AGNews* berbahasa Inggris kedalam empat kategori menggunakan metode *Long Short-Term Memory (LSTM)* dengan ekstraksi fitur menggunakan *Word2Vec*. Hasil penelitian menunjukkan akurasi tertinggi sebesar 95,38%. Salah satu pemodelan *deep learning* yang lain adalah *Transformer*. *Transformer* dengan cepat menjadi model pilihan untuk masalah NLP karena *Transformer* memfasilitasi lebih banyak paralelisasi selama pelatihan sehingga waktu pelatihan berkurang sehingga mengatasi salah satu kekurangan yang dimiliki RNN biasa. *XLNet* adalah salah satu model arsitektur *Transformer* yang baru-baru ini dirilis dan dikembangkan.

XLNet adalah metode pemodelan bahasa *autoregressive* menggunakan konsep model permutasi. Metode *XLNet* sudah diterapkan dalam beberapa penelitian terkait task NLP seperti penelitian yang melakukan klasifikasi kategori berita menggunakan *dataset AGNews* berbahasa Inggris kedalam empat kategori dengan hasil akurasi 95.55%[4].

Berdasarkan pada penelitian yang disebutkan diatas, diketahui bahwa pada penelitian [2] belum menerapkan pembelajaran *deep learning* dan pada penelitian [3] dan [4] sudah menerapkan pembelajaran *deep learning* dengan masing-masing menggunakan *LSTM* dan *XLNet* namun *dataset* yang digunakan menggunakan bahasa Inggris. Oleh karena itu, pada penelitian ini menggunakan *XLNet* dalam

melakukan klasifikasi berita berbahasa Indonesia kedalam lima kategori berita. Penelitian ini diharapkan dapat diketahui nilai akurasi dari metode XLNet dalam mengklasifikasi berita berbahasa Indonesia serta menghasilkan akurasi yang lebih baik.

1.2 Rumusan Masalah

Dari latar belakang yang dipaparkan diatas, dapat dirumuskan permasalahannya adalah berapa nilai akurasi yang diperoleh dari XLNet dalam melakukan klasifikasi kategori berita berbahasa Indonesia.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut

1. Kategori dokumen berita yang digunakan adalah bola, kesehatan, keuangan, otomotif dan properti.
2. Bahasa yang digunakan adalah Bahasa Indonesia.

1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan dari penelitian ini adalah mengetahui nilai akurasi dari XLNet dalam melakukan klasifikasi berita kedalam lima kategori yang sudah ditentukan.

1.5 Manfaat Penelitian

Dengan dilakukannya penelitian ini diharapkan mampu memberikan manfaat berupa diketahui nilai akurasi yang dihasilkan XLNet dalam klasifikasi kategori berita berbahasa Indonesia serta dapat dijadikan referensi untuk penelitian lain yang berkaitan.

1.6 Sistematika Penulisan

Untuk mengetahui uraian singkat yang memuat gambaran singkat secara keseluruhan isi masing-masing bab, maka dibuat sistematika sebagai berikut:

BAB I

PENDAHULUAN

Bab ini berisi tentang uraian latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II

LANDASAN TEORI

Bab ini berisi tentang teori-teori yang digunakan serta konsep dasar yang menjadi landasan dari penelitian ini yang berasal dari berbagai sumber seperti jurnal dan *proceeding*.

BAB III

METODE PENELITIAN

Bab ini berisi alur tahapan penelitian, alat dan bahan yang akan digunakan dalam penelitian, serta langkah-langkah penelitian.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas tentang proses implementasi yang dilakukan secara menyeluruh serta hasil yang didapatkan dari penelitian.

BAB V

PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk memperbaiki kekurangan yang ada pada penelitian.