

BAB I

PENDAHULUAN

1.1. Latar Belakang

Akses manusia terhadap berita terbaru saat ini semakin mudah dan masif, disebabkan oleh perkembangan teknologi pada masa ini (Setiawan, 2020). Teknologi telah mengubah bentuk penyaluran berita yang dahulunya disalurkan lewat media surat kabar, majalah, radio, dan televisi, hingga saat ini menggunakan sistem berbasis internet (Findra, 2021). Umumnya, berita yang disalurkan lewat sistem berbasis internet atau portal berita dikelompokkan berdasarkan kategori, seperti politik, olahraga, ekonomi, hiburan, teknologi, kesehatan dan lain-lain (Usmani, 2020).

Saat ini pengelompokan berita kedalam kategori-kategori tertentu dilakukan secara manual, sehingga pengelompokan berita harus diawali dengan membaca keseluruhan isi berita agar bisa di kelompokkan dengan tepat. Hal ini sangat merepotkan, ditambah lagi berita yang ingin dikelompokkan berjumlah sangat banyak (Findra, 2021). Selain itu, terkadang ada pihak yang turut sengaja mengelompokkan berita dengan kategori yang populer dikarenakan ingin beritanya dibaca oleh banyak orang (Setiawan, 2020).

Peningkatan dokumen elektronik yang semakin pesat dari waktu ke waktu memberikan perhatian khusus mengenai pentingnya klasifikasi dokumen. Hal ini dikarenakan meningkatnya dokumen tidak terstruktur dalam jumlah yang besar yaitu, salah satunya artikel berita (Ranjan, 2017).

Berdasarkan penelitian sebelumnya oleh (Randhika, 2021) dengan judul Implementasi Algoritma *Complement* dan *Multinomial Naive Bayes Classifier* Pada Klasifikasi Kategori Berita Media *Online* dengan menggunakan metrik evaluasi *F1-Score*,

dimana penelitian tersebut mencoba menggabungkan metode *Multinomial Naive Bayes Classifier* (MNBC) dan *Complement Naive Bayes Classifier* (CNBC). Penelitian ini mencoba mengkategorikan artikel berita kedalam 3 kategori yaitu Hiburan dan Gaya Hidup, Olahraga, dan Indonesiaku. Hasil dari penelitian ini menunjukkan nilai metrik evaluasi *F1-Score* sebesar 90,13%. Meskipun memiliki nilai *F1-Score* yang sangat baik, namun tingkat kesalahan prediksinya masih terbilang tinggi yaitu Olahraga sebesar 5% dan Indonesiaku 20%.

Selanjutnya penelitian serupa terkait kategori berita dilakukan oleh (Sari, 2020) dengan judul Klasifikasi Teks Multilabel pada Artikel Berita Menggunakan *Long Short Term Memory* dengan *Word2Vec*. Pada penelitian tersebut melakukan klasifikasi secara multilabel teksklasifikasi dengan menggunakan *feature ekstraction word embedding Word2Vec* 300 dimensi dengan *tuning hyperparameter* dan membuat perbandingan 8 model LSTM. Penelitian ini menggunakan *AGNews datasets* berbahasa Inggris dengan 400.000 sampel. Kesimpulan dari penelitian ini adalah metode *Long Short Term Memory* (LSTM) dengan fitur *word embedding Word2Vec* berhasil mengklasifikasi artikel berita dengan akurasi yang dihasilkan 95.38%.

Salah satu metode *deep learning* yang diusulkan pada penelitian ini adalah *Recurrent Neural Network* (RNN) dengan menerapkan arsitektur *Long Short Term Memory* (LSTM). LSTM dikembangkan untuk menangani masalah *exploding* dan *vanishing gradient* yang dapat mengatasi saat melatih RNN tradisional. LSTM dapat digunakan untuk menangani kompleksitas yang luar biasa menjadikannya sebagai algoritma pembelajaran mesin yang sesuai untuk klasifikasi dokumen (Sari, 2020).

Berdasarkan permasalahan diatas, penulis melakukan penelitian yaitu "Klasifikasi Teks Pada Artikel Berita Menggunakan Metode *Long Short Term Memory* (LSTM)" untuk

mengelompokan artikel berita bahasa Indonesia sesuai dengan kategori yang ditentukan.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka yang menjadi rumusan masalah pada penelitian adalah bagaimana menerapkan algoritma *Long Short Term Memory* (LSTM) untuk mengklasifikasikan teks pada artikel berita Bahasa Indonesia secara otomatis berdasarkan kategori yang telah ditentukan.

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini, yaitu:

1. Domain penelitian dibangun untuk mengklasifikasi artikel berita berbahasa Indonesia.
2. *Dataset* yang digunakan sebagai data latih dan data uji berasal dari portal berita cnn indonesia, antaranews, *dailysocial*, *goal* indonesia, juara, kumparan, merdeka, poskotanews, rimanews, dan suara.
3. Aplikasi hanya dapat mengklasifikasi artikel dengan kategori hiburan, *showbiz*, tajuk utama, teknologi, dan olahraga.
4. *Output* dari penelitian ini adalah *aplikasi* berbasis *website*.

1.4. Tujuan Penelitian

Adapun tujuan penelitian ini adalah membuat system yang dapat mengklasifikasikan teks pada artikel berita bahasa Indonesia secara otomatis berdasarkan kategori yang telah ditentukan dengan menerapkan algoritma *Long Short Term Memory* (LSTM).

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sebuah *system* yang dapat mengklasifikasikan artikel berita berdasarkan kategori yang telah ditetapkan sebelumnya.

2. Dapat mengimplementasikan algoritma *Long Short Term Memory* (LSTM) dalam bidang *Natural Language Processing* (NLP).
3. Diharapkan perancangan aplikasi ini dapat menambah referensi terhadap penelitian baru dengan masalah yang sama.

1.6. Sistematika Penulisan

Untuk memudahkan pembahasan dalam skripsi ini, sistematika penulis dibagi menjadi 5 (lima) bab yang terdiri dari:

BAB I PENDAHULUAN

Terdiri dari latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Memaparkan teori-teori yang didapat dari sumber-sumber yang relevan untuk digunakan sebagai panduan dalam penelitian serta penyusunan proposal.

BAB III METODE PENELITIAN

Bab ini membahas tentang metode penelitian yang telah dilakukan oleh penulis dengan permasalahan yang diangkat.

BAB IV HASIL DAN PEMBAHASAN

Menjelaskan hasil penelitian yang telah dilakukan yang terdiri dari analisis data, pengujian metode, evaluasi metode, dan implementasi *interface* dan pengujian *black box*.

BAB V PENUTUP

Memuat kesimpulan dari hasil penelitian dan saran untuk penelitian selanjutnya mengenai topik terkait.