

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam skripsi Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan *Bidirectional Gated Recurrent Unit*.

1.1 Latar Belakang

Seiring dengan perkembangan zaman, teknologi merupakan bagian besar yang mengalami perubahan dengan pesat. Perkembangan teknologi meningkatkan jumlah portal berita daring yang menyediakan berbagai informasi. Pertumbuhan berita daring memiliki dampak dalam pertumbuhan konten berita daring dengan judul *clickbait*. Di era digital, maraknya konten daring telah meningkatkan kebutuhan akan metode yang kuat untuk mengklasifikasikan dan menyaring informasi. Tugas klasifikasi dalam mengidentifikasi judul *clickbait* menjadi suatu kepentingan karena dampaknya terhadap keterlibatan pengguna dan penyebaran informasi yang tidak tepat. *Clickbait*, yang dicirikan oleh sifatnya yang berlebihan atau menyesatkan, dapat secara signifikan mempengaruhi perilaku pembaca dan kredibilitas konten.

Klasifikasi teks adalah proses mengkategorikan data teks ke dalam kategori label yang telah ditentukan sebelumnya sehingga teks dapat digunakan untuk berbagai hal, seperti penyaringan spam dan deteksi topik (Dzisevic & Šešok, 2019). Proses klasifikasi teks melibatkan transformasi teks mentah menjadi data numerik yang dapat diterima dan dikenali polanya oleh algoritma *machine learning* atau *deep learning*. Model *deep learning* dapat melakukan tugas klasifikasi teks secara signifikan dengan menangkap hubungan yang kompleks dan nuansa kontekstual dalam teks yang membuat klasifikasi lebih akurat dan mudah beradaptasi. Pemrosesan data teks mencakup praproses teks (seperti *case folding*, tokenisasi dan *stemming* kata), ekstraksi fitur (menggunakan teknik seperti *word embedding*), pemilihan model *deep learning* yang sesuai, melatih model pada contoh data latih berlabel, dan mengevaluasi kinerjanya menggunakan metrik seperti nilai akurasi (Maslej-Krešňáková dkk., 2020).

Clickbait pada judul dirancang untuk menarik perhatian dan mengundang pengguna untuk mengklik, seringkali tidak selalu mengandung informasi yang akurat atau berharga. *Clickbait* pada umumnya tidak menjelaskan isi dari berita yang sebenarnya sehingga menciptakan ketidakselarasan antara judul berita dan isi konten berita (Wei & Wan, 2017). Pada zaman digital ini dimana segala hal dapat dengan cepat disebarkan, penggunaan *clickbait* menyebabkan banyak konten disesuaikan untuk menciptakan sensasi dengan menjadi viral daripada menyuguhkan berita yang akurat dan berkualitas. Dengan mengenali judul berita yang merupakan *clickbait* dan tidak, maka kualitas, integritas, dan keakuratan berita, kepercayaan pembaca terhadap media daring dapat terjaga (Rahman, 2023).

Penelitian klasifikasi judul berita *clickbait* yang dilakukan oleh Chowanda dkk. (2022) menggunakan beberapa teknik *machine learning* yang populer, yaitu *Naïve Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Artificial Neural Network*, *Random Forest*, *Gradient Boosted Trees*, *Decision Trees* dan *Support Vector Machine*. Dengan menggunakan data 32.000 judul berita *clickbait* berbahasa Inggris, *Fast Large Margin* memberikan *testing accuracy* sebesar 90%. Bhoj dkk. (2021) melakukan penelitian menggunakan dataset berbahasa Inggris yang diperoleh dari *website* hiburan dan berita dengan mengusulkan model *Naïve Bayes*, *Random Forest*, dan *Long Short Term Memory*. Berdasarkan hasil penelitian tersebut model *Long Short Term Memory* menghasilkan akurasi sebesar 95,03%. Studi terhadap 1000 judul berita daring berbahasa Indonesia yang dilakukan oleh Isyriyah dkk. (2024) menggunakan metode *K-Nearest Neighbors* untuk tugas klasifikasi menghasilkan akurasi sebesar 80,5%.

Hingga saat ini penelitian mengenai *clickbait* berbahasa Indonesia masih sangat sedikit. Meskipun model *deep learning* lainnya, *Long Short Term Memory* dan *Bidirectional Long Short Term Memory*, banyak digunakan dalam berbagai aplikasi pemrosesan bahasa alami, *Gated Recurrent Unit* dan *Bidirectional Gated Recurrent Unit* juga memiliki keunggulan dalam hal efisiensi dan kinerja yang lebih baik (Meriani & Rahmatulloh, 2024). Namun, dibanding dengan *Long Short Term Memory* dan *Bidirectional Long Short Term Memory*, *Gated Recurrent Unit* dan *Bidirectional Gated Recurrent Unit* membutuhkan komputasi yang mahal ketika berhadapan dengan sejumlah besar data sekuensial. Model *deep learning*, seperti *Bidirectional Gated Recurrent Unit*, memiliki potensi dalam menangani data sekuensial dan menangkap hubungan kontekstual

(Li dkk., 2024). Dengan memanfaatkan arsitektur *Bidirectional Gated Recurrent Unit* dapat meningkatkan kontribusi terhadap klasifikasi judul *clickbait* dalam bahasa Indonesia.

Oleh karena itu, pada penelitian skripsi ini diusulkan penerapan model *Bidirectional Gated Recurrent Unit* untuk digunakan dalam melakukan tugas klasifikasi teks judul berita *clickbait* berbahasa Indonesia. Model ini dipilih untuk diimplementasikan dalam penelitian ini berdasarkan keunggulan yang dimilikinya dari antara model lainnya. Penelitian tentang klasifikasi dengan *Bidirectional Gated Recurrent Unit* ini diharapkan dapat menghasilkan performa yang lebih baik daripada penerapan metode klasifikasi sebelumnya serta menambah literatur yang ada dan digunakan untuk tujuan pendidikan.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang, maka dapat dirumuskan masalah yang harus diselesaikan pada skripsi ini adalah bagaimana kombinasi pengaturan *hyperparameter* model *Bidirectional Gated Recurrent Unit* memengaruhi performa optimal dalam klasifikasi judul berita *clickbait*?

1.3 Tujuan dan Manfaat

Tujuan dari penelitian skripsi ini adalah untuk mendapatkan *hyperparameter* yang optimal dari model *Bidirectional Gated Recurrent Unit* untuk klasifikasi pada judul berita *clickbait* berbahasa Indonesia. Manfaat dari penelitian skripsi ini adalah menemukan kombinasi pengaturan *hyperparameter* model *Bidirectional Gated Recurrent Unit* yang menghasilkan performa terbaik dalam klasifikasi judul berita *clickbait* berbahasa Indonesia.

1.4 Ruang Lingkup

Ruang lingkup diperlukan agar penelitian dari skripsi yang dilakukan sesuai dan mencapai sasaran yang diharapkan.

Ruang lingkup dalam penelitian skripsi ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian skripsi ini adalah dataset CLICK-ID, yaitu kumpulan data teks yang terdiri dari judul berita berbahasa Indonesia (William & Sari, 2020).
2. Data terbagi ke dalam dua klasifikasi, yaitu *clickbait* dan *non-clickbait*.

3. Model yang diterapkan dalam penelitian skripsi ini adalah *Bidirectional Gated Recurrent Unit*.

1.5 Sistematika Penulisan

Sistematika penulisan memberikan gambaran laporan dari skripsi disesuaikan dengan sistematika pembahasan untuk memberikan suatu gambaran yang urut dan jelas sebagai berikut:

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan *Bidirectional Gated Recurrent Unit*.

BAB II LANDASAN TEORI

Bab ini menyajikan studi literatur dan membahas tentang kajian teori, meliputi pra-pemrosesan data, pembagian data, model GloVe, pembangunan model menggunakan *Bidirectional Gated Recurrent Unit*, pelatihan model, serta pengujian model dan evaluasi, yang digunakan untuk menunjang laporan dari skripsi Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan *Bidirectional Gated Recurrent Unit*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang tahapan dalam proses penelitian dari skripsi Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan *Bidirectional Gated Recurrent Unit*. Tahapan metodologi meliputi pra-pemrosesan data, pembagian data, model GloVe, pembangunan model menggunakan arsitektur *Bidirectional Gated Recurrent Unit*, pelatihan model, serta pengujian model dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan informasi tentang lingkungan dan perangkat yang digunakan untuk penelitian, skenario eksperimen, serta hasil dan analisis dari penelitian skripsi Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan *Bidirectional Gated Recurrent Unit*.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan kesimpulan dari bab-bab yang sudah dibahas sebelumnya dari penelitian skripsi Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan *Bidirectional Gated Recurrent Unit* dan saran dari penulis sebagai bahan masukan yang dapat dilakukan pada penelitian terkait selanjutnya.