

BAB I PENDAHULUAN

1.1. Latar Belakang

Di era digital saat ini, informasi menjadi kebutuhan penting bagi semua orang. Perkembangan teknologi internet telah memicu peningkatan pesat dalam penyebaran informasi, terutama pada berita elektronik berbahasa Indonesia [1]. Berita merupakan informasi yang disampaikan oleh berbagai sumber, dengan banyak definisi yang berbeda. Namun, secara umum, berita dapat disimpulkan sebagai sebuah opini atau informasi yang menarik dan akurat, sehingga dianggap penting bagi banyak orang yang gemar membaca, mendengar, atau menonton [2]. Banyaknya orang yang mengakses portal berita menyebabkan persaingan sengit antar media untuk menarik sebanyak mungkin pembaca atau pengunjung. Akibatnya, penggunaan judul berita *clickbait* ini sudah semakin meningkat [3].

Clickbait merupakan teknik membuat judul atau gambar yang sengaja dirancang untuk menarik perhatian dan membangkitkan rasa penasaran, sehingga membuat penasaran orang ingin mengklik dan membaca berita tersebut. Judul *clickbait* sering kali terdengar berlebihan. Tujuan utamanya yaitu menarik minat pembaca, meskipun kadang isi artikel tidak se-sensasional judulnya. [4]. Penggunaan *clickbait* yang berkelanjutan dapat merusak pengalaman pengguna dan kemungkinan bahwa pembaca akan menjadi lebih skeptis terhadap semua jenis berita, dapat merusak pengalaman pengguna dan meningkatkan skeptisisme pembaca terhadap semua berita, bukan hanya yang bersifat *clickbait*. Hal ini dapat menurunkan minat baca dan mengurangi keterlibatan publik dalam isu penting. Selain itu, dominasi judul *clickbait* membuat masyarakat semakin sulit membedakan berita asli dari yang manipulatif, karena *clickbait* memang dirancang untuk menarik perhatian. Selain itu, Indonesia masuk dalam daftar negara dengan tingkat literasi yang rendah menurut survei *Program for International Student Assessment* (PISA) yang diterbitkan oleh *Organization for Economic Co-operation and Development* (OECD) pada tahun 2019 [5].

Berdasarkan permasalahan judul berita *clickbait* tersebut, diperlukan penelitian untuk mengklasifikasikan judul berita ke dalam kategori *clickbait* atau *non-clickbait*. Pengelompokan berita ini dilakukan menggunakan teknik *text mining*, yang bertujuan menemukan dan mengekstrak informasi penting dari teks bebas atau tidak terstruktur. *Text mining* melibatkan berbagai tahapan, mulai dari pengumpulan informasi, dokumen, atau data web hingga proses klasifikasi.

Klasifikasi teks adalah salah satu tugas dalam *natural language processing* (NLP) yang berfokus pada pengelompokan teks secara otomatis berdasarkan konteksnya, menggunakan teknik *machine learning* atau *deep learning*. Karena semakin banyaknya konten yang diunggah pengguna di berbagai *platform* media sosial, klasifikasi teks menjadi bidang penelitian yang terus berkembang dalam pemrosesan bahasa alami [6].

Pada tahun 2023 Fikri Alwan Ramadhan *et. al.* Penelitian ini mengkaji penerapan metode *multinomial naïve bayes* untuk mengklasifikasikan judul berita *clickbait* dengan menggunakan TF-IDF (*Term Frequency - Inverse Document Frequency*). Data yang digunakan berupa judul berita online yang dikumpulkan dari berbagai situs web. TF-IDF digunakan untuk memberikan bobot pada kata-kata selama proses pelatihan, pengujian, dan klasifikasi. *Dataset* yang dipakai berjumlah 1.000 data, dengan 800 data untuk pelatihan dan 200 data untuk pengujian. Hasil pengujian menggunakan *confusion matrix* menunjukkan akurasi sebesar 65%, serta *recall* dan *precision* masing-masing sekitar 65,69%. [7].

Selain itu ditemukan juga penelitian mengenai klasifikasi judul berita *clickbait* pada tahun 2024 oleh Dicky Satria Mahendra *et. al.*, Penelitian ini mengimplementasikan metode *Multinomial Naive Bayes* untuk klasifikasi judul berita *clickbait*. Hasil pengujian menunjukkan bahwa algoritma *Multinomial Naive Bayes* mampu mencapai tingkat akurasi yang konsisten sekitar 78%. Upaya optimasi melalui *Grid Search* tidak meningkatkan akurasi secara keseluruhan, namun terdapat peningkatan pada nilai *recall* untuk kelas *non-clickbait*, dari 76% menjadi 80%. Parameter terbaik yang ditemukan dalam optimasi ini adalah nilai *alpha* sebesar 0.15 [4].

Terdapat juga penelitian mengenai klasifikasi berita hoaks pada tahun 2023 oleh Ari Muzakir *et. al.* Penelitian ini menggunakan model *Bidirectional Long Short-Term Memory* (BILSTM) dengan representasi kata menggunakan *embeddings* GloVe. Hasil evaluasi menunjukkan model mampu beradaptasi dengan baik pada data pelatihan, ditandai dengan kerugian (*loss*) terendah sebesar 0,30% dan akurasi yang sangat tinggi, yaitu 99,14% [8].

Dari beberapa penelitian yang disebutkan sebelumnya dapat disimpulkan bahwa penggunaan metode BILSTM memiliki hasil yang cukup bagus dalam klasifikasi. Akan tetapi, model tersebut masih memiliki potensi untuk dikembangkan dan ditingkatkan akurasinya. Berdasarkan penjelasan tersebut,

dalam penelitian ini saya tertarik menggunakan metode BiLSTM dengan representasi vektor dari *FastText*. BiLSTM adalah model yang memproses data secara berurutan dari dua arah, yaitu dengan menggunakan lapisan maju (*forward*) untuk menangkap informasi dari masa depan dan lapisan mundur (*backward*) untuk menangkap informasi dari masa lalu. Pendekatan ini memudahkan model memahami hubungan antara kedua arah tersebut dan membantu dalam memprediksi kata berikutnya dalam teks berita. Selain itu, BiLSTM efektif mengatasi masalah *vanishing gradient* karena menggunakan mekanisme gerbang (*gate*) yang memilih informasi penting untuk disimpan dalam memori sel. Dengan cara ini, model dapat mengurangi penumpukan lapisan yang biasanya memperlambat proses pelatihan. BiLSTM juga termasuk algoritma *deep learning* yang membutuhkan input dalam bentuk vektor [9]. Oleh karena itu, dalam penelitian ini dibutuhkan metode vektorisasi yang tepat. Sebagai solusi, penelitian ini akan menggunakan *FastText* untuk mengubah kata-kata menjadi representasi vektor. *FastText* memiliki keunggulan mampu memberikan representasi untuk kata-kata yang tidak ada dalam *dataset*, sehingga dapat mengatasi masalah *out of vocabulary* dengan lebih baik. Kata-kata yang tidak ditemukan selama proses tersebut akan dipecah menjadi *n-gram*, yaitu kumpulan urutan suku kata, untuk memperoleh embedding vektornya. [10].

Hasil yang diharapkan dari penelitian ini meliputi beberapa aspek terukur. Metode BiLSTM dan vektorisasi *FastText* diharapkan dapat mencapai akurasi yang lebih tinggi dari penelitian sebelumnya dalam mengklasifikasikan berita *clickbait*, yang menunjukkan peningkatan signifikan dibandingkan dengan metode konvensional sebelumnya. Penerapan model ini juga diharapkan dapat secara substansial mengurangi jumlah berita *clickbait* yang beredar di platform berita, memberikan dampak positif terhadap kualitas informasi yang diterima publik. Pencapaian hasil ini akan membuktikan efektivitas dan efisiensi dari metode BiLSTM dalam klasifikasi berita *clickbait*.

1.2. Rumusan Masalah

1.2.1 Permasalahan

Tingginya jumlah berita *clickbait* yang beredar di platform berita berbahasa Indonesia. Jenis berita ini sering kali sulit dibedakan dari berita yang asli karena judulnya yang sensasional. Dampak ini merugikan standar etika jurnalistik, sehingga perlu lebih kritis dalam menyikapi praktik *clickbait*. Analisis teks

otomatis untuk mengidentifikasi berita *clickbait* memerlukan pemahaman kontekstual yang mendalam dan kemampuan untuk menangkap pola linguistik yang kompleks. Saat ini, metode klasifikasi teks yang umum digunakan seperti *Naive Bayes* atau SVM serta representasi teks seperti TF-IDF dan *Bag-of-Words* sering kali mengalami keterbatasan dalam hal akurasi terutama dalam bahasa Indonesia yang memiliki struktur linguistik dan kata yang unik. Oleh karena itu, diperlukan pendekatan yang dapat meningkatkan akurasi klasifikasi serta efisiensi pemrosesan.

1.2.2 Metode Usulan

Penelitian ini menggunakan metode BILSTM dan representasi *FastText* untuk mendeteksi *clickbait* dan *non-clickbait* pada judul berita. BILSTM dipilih karena dapat membaca teks dari dua arah, memahami konteks lebih baik, dan mengatasi masalah *vanishing gradient* dengan menyimpan informasi penting. *FastText* dipilih karena kemampuannya menangani kata-kata yang tidak ada dalam dataset (*out-of-vocabulary*), sehingga memungkinkan pembentukan *embedding* untuk kata baru. Kinerja model dievaluasi menggunakan *confussion matrix*, *precicion* , *recall*, dan *F1-score*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan yang telah di jelaskan sebelumnya, dapat dirumuskan pertanyaan dalam penelitian sebagai berikut:

1. Bagaimana performa yang dihasilkan menggunakan model BILSTM dengan representasi teks *fasttext* dalam klasifikasi judul berita *clickbait* berdasarkan metrik evaluasi seperti akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix*?

1.3. Tujuan dan Manfaat

1.3.1 Tujuan

1. Untuk mengetahui performa yang akan dihasilkan menggunakan model BILSTM dengan representasi teks *fasttext* dalam klasifikasi judul berita *clickbait*.

1.3.2 Manfaat

1. Meningkatkan kualitas berita dengan mendeteksi dan mengurangi penggunaan *clickbait* pada judul berita.
2. Hasil penelitian ini dapat menjadi referensi untuk meningkatkan performa dalam pengembangan teknik *machine learning* dengan memanfaatkan

kemampuan BILSTM dan representasi dari *fasttext*.

1.4. Batasan Masalah

Batasan masalah yang diterapkan dalam penelitian tugas akhir ini antara lain:

1. Data yang diambil merupakan Judul berita yang didapatkan dari 10 website berita, yaitu detik.com, gridoto, juara.net, kapanlagi, kompastv, liputan6, tirtoid, tribunnews, wowkeren, dan lambeturah. Dengan rentan waktu dari bulan Agustus 2024 sampai dengan bulan Oktober 2024.
2. Jumlah label yang diklasifikasikan ada 2, yaitu *clickbait* dan *non-clickbait*.
3. Jumlah judul berita yang digunakan yaitu 1320 dengan rincian kelas *clickbait* 613 judul berita *non-clickbait* 705 judul berita.
4. Fitur yang dipakai dalam penelitian ini adalah judul berita, dan label.
5. Fokus penelitian adalah untuk mengetahui apakah metode yang diusulkan efektif untuk permasalahan dan studi kasus yang dipilih.

1.5. Sistematika Penulisan Proposal

Pada penelitian ini, digunakan sistematika penulisan proposal sebagai berikut:

BAB I PENDAHULUAN

Pada bab pertama merupakan pendahuluan yang memiliki gambaran secara umum mengenai judul yang diangkat dalam penelitian ini, mulai dari pemaparan Maraknya judul berita *clickbait* sampai dengan *deep learning* metode BILSTM. Penulis menyusun pemaparan dan ringkasan dari setiap sub bab yang dibagi dalam lima bab yaitu latar belakang, metode usulan, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab dua berisi tentang kajian penelitian terdahulu yang melandasi peneliti dalam menulis proposal ini, dan memuat kajian pustaka tentang landasan teori *Clickbait*, *Word Embedding*, *FastText*, RNN, LSTM, metode BILSTM, dan Evaluasi Model.

BAB III METODE USULAN

Pada bab tiga menjelaskan tentang metode penelitian yang memuat jenis dan sumber data yang diambil, bab ini juga menguraikan teknik pengolahan data, langkah – langkah dalam penelitian dan evaluasi hasil analisa klasifikasi serta uji coba dalam proses analisa data yang dilakukan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab empat dibahas hasil dari proses klasifikasi berita clickbait dan non-clickbait menggunakan model BILSTM dengan embedding FastText yang dilatih dari data. Proses ini mencakup tahapan pelatihan model, evaluasi hasil, serta analisis dari berbagai kombinasi parameter pelatihan yang diuji.

BAB V PENUTUP

Pada bab lima ini menjelaskan kesimpulan hasil yang diperoleh dari serangkaian pengujian dan analisis, dapat disimpulkan bahwa pemilihan parameter pelatihan memiliki pengaruh yang signifikan terhadap performa model klasifikasi, dan berisi saran untuk penelitian selanjutnya mengembangkan aplikasi yang telah dibuat.