

BAB I PENDAHULUAN

1.1 Latar Belakang

Clickbait adalah jenis judul atau konten teks yang dirancang untuk menciptakan *information gap*—celah antara apa yang diketahui pembaca dan apa yang ingin mereka ketahui—guna memancing rasa penasaran dan mendorong pengguna untuk mengklik sebuah tautan[1], [2]. Judul-judul *clickbait* umumnya menggunakan bahasa yang hiperbolik, menggugah emosi, atau bersifat provokatif, seperti pertanyaan atau ungkapan menggantung, namun sering kali tidak merefleksikan isi sebenarnya dari berita yang dimuat[2], [3]. Dalam konteks pragmatik, *clickbait* mengeksploitasi prinsip relevansi: pembaca diarahkan pada interpretasi yang tampaknya memiliki nilai berita tinggi, padahal isi sebenarnya bersifat trivial atau mengecewakan.

Di tengah kompetisi atensi yang semakin intens di ruang digital praktik *clickbait* kian meluas tidak hanya digunakan oleh media digital baru tetapi juga oleh media arus utama dan *outlet* berita dengan reputasi mapan. Dalam konteks Indonesia, fenomena ini juga berkembang pesat. Sebagaimana dicatat oleh Fakhruzzaman *et al.* (2021), banyak situs berita daring nasional menggunakan judul-judul *clickbait* untuk meningkatkan jumlah klik, yang secara langsung berkorelasi dengan pendapatan dari iklan digital berbasis tayangan atau interaksi. Bahkan, beberapa situs berita yang sebelumnya dikenal kredibel pun turut menggunakan strategi ini, yang akhirnya menurunkan persepsi publik terhadap integritas media[4].

Fenomena peningkatan penggunaan *clickbait* juga terjadi pada level global. Penelitian berbasis data Common Crawl menemukan bahwa tingkat *clickbait* meningkat secara signifikan setelah peristiwa besar, seperti deklarasi COVID-19 sebagai pandemi dan Pemilu AS 2020[1]. Hal ini menunjukkan bahwa *clickbait* sering dimanfaatkan oleh media untuk memanfaatkan momen-momen yang memicu perhatian publik secara masif. Meskipun penelitian tersebut berfokus pada

situs berita berbahasa Inggris, pola serupa dapat diamati di Indonesia, di mana lonjakan produksi berita daring saat pandemi COVID-19 diikuti oleh meningkatnya praktik *clickbait* untuk mengejar keterlibatan pengguna dan pendapatan iklan.

Penggunaan *clickbait* juga tidak bisa dilepaskan dari pengaruh algoritma platform media sosial seperti Facebook dan Twitter. Seperti diuraikan oleh Lischka dan Garz (2021), sistem kurasi algoritmik mendorong konten yang memiliki tingkat interaksi tinggi, sehingga memberi insentif kepada media untuk memproduksi konten yang "*clickworthy*" alih-alih informatif. Dalam lingkungan ini, banyak media cenderung mengikuti standar industri tidak tertulis dalam penggunaan *clickbait* demi mempertahankan visibilitas dan daya saing di platform digital[5].

Di Indonesia fenomena *clickbait* juga berkembang pesat. Banyak situs berita daring menggunakan judul-judul *clickbait* untuk mengejar klik demi pendapatan dari iklan daring, meskipun hal ini berisiko menurunkan kualitas jurnalistik dan kepercayaan pembaca. Beberapa penelitian terdahulu telah mencoba mendeteksi *clickbait* dalam bahasa Indonesia menggunakan berbagai pendekatan berbasis *machine learning*, seperti *neural network* dan model berbasis *transformer*[3]. [4]. Namun, pendekatan alternatif dengan model *tree-based ensemble*, seperti *Gradient Boosting Tree*, masih jarang dieksplorasi padahal berpotensi memberikan hasil klasifikasi yang kompetitif dan interpretatif[6].

Gradient Boosting Tree (GBT) merupakan algoritma *ensemble* yang menggabungkan sejumlah model pohon keputusan secara bertahap untuk membentuk *prediktor* yang kuat dan akurat, terutama dalam menangani data dengan pola yang kompleks dan non-linear. Dalam konteks klasifikasi berbasis teks, GBT telah menunjukkan performa yang menjanjikan, salah satunya dalam studi yang memanfaatkan analisis sentimen dari data Twitter untuk memprediksi pergerakan harga saham perusahaan teknologi berbasis percakapan tentang *ChatGPT*. Studi tersebut membuktikan bahwa GBT mampu mengolah fitur-fitur berbasis teks seperti skor sentimen positif dan negatif serta metrik keterlibatan pengguna, dan menghasilkan performa klasifikasi yang kompetitif, khususnya dalam mengidentifikasi tren pasar keuangan[7]. Berdasarkan efektivitas tersebut,

algoritma ini menjadi kandidat yang potensial untuk diterapkan dalam tugas klasifikasi judul berita *clickbait* berbahasa Indonesia, yang juga merupakan permasalahan berbasis teks dengan pola linguistik yang kompleks.

Studi oleh Subramani et al. (2023) menunjukkan bahwa *Gradient Boosted Decision Tree* (GBDT) dapat mencapai performa klasifikasi yang sangat tinggi dalam analisis data sosial berbasis teks, khususnya dalam tugas prediksi *influencer* di media sosial. Dengan menggabungkan GBDT dan *Chaotic Gradient-Based Optimizer* (CGBO), pendekatan ini mampu mengklasifikasikan data yang tidak jelas dengan akurasi yang lebih baik dibandingkan algoritma lain seperti *Naïve Bayes*, *SVC*, dan *Artificial Neural Networks*. Eksperimen mereka menunjukkan bahwa model GBDT-CGBO mencapai akurasi hingga 98,76% dengan waktu eksekusi yang lebih efisien, menjadikannya solusi efektif untuk permasalahan klasifikasi berbasis teks dalam lingkungan yang dinamis dan padat informasi[8]. Temuan ini memperkuat relevansi GBT dalam mengidentifikasi pola linguistik yang kompleks seperti yang terdapat dalam judul berita *clickbait* berbahasa Indonesia.

Dengan mempertimbangkan urgensi untuk memfilter dan mengklasifikasi judul berita *clickbait* secara otomatis serta kebutuhan akan metode yang efisien dan dapat dijelaskan, penelitian ini bertujuan untuk mengembangkan model klasifikasi judul berita *clickbait* berbahasa Indonesia menggunakan algoritma *Gradient Boosting Tree*. Penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan literasi digital masyarakat dan membantu platform media serta pengguna untuk lebih selektif dalam mengonsumsi informasi daring.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah ditulis diatas maka penulis dapat merumuskan masalah sebagai berikut.

1. Bagaimana merancang model *machine learning* untuk klasifikasi otomatis judul berita *clickbait* menggunakan kombinasi fitur tekstual dan simbolik?
2. Bagaimana pengaruh *hyperparameter tuning* terhadap peningkatan

performa dan efisiensi model klasifikasi *clickbait*?

3. Bagaimana mengimplementasikan model deteksi *clickbait* dalam bentuk aplikasi *real-time* yang dapat diakses publik?

1.3 Batasan Masalah

Dalam penelitian ini, terdapat sejumlah batasan masalah yang penulis rincikan sebagai berikut:

1. Penelitian ini hanya berfokus pada klasifikasi judul berita *clickbait* yang berbahasa Indonesia.
2. Judul-judul berita pada dataset yang digunakan di penelitian ini hanya bersumber dari beberapa situs berita *online* terpilih.
3. Penelitian ini murni hanya mempertimbangkan judul dari berita, tidak termasuk isi berita maupun gambar dari beritanya.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai oleh penulis dalam penelitian ini adalah untuk mengembangkan sebuah model *machine learning* yang mampu melakukan klasifikasi terhadap judul berita *clickbait* dalam bahasa Indonesia secara akurat. Selain itu, penelitian ini juga bertujuan untuk melakukan kajian dan evaluasi terhadap performa model yang dihasilkan, guna menilai efektivitas serta potensi penerapannya dalam konteks deteksi *clickbait* di media digital.

1.5 Manfaat Penelitian

Melalui penelitian ini, penulis berharap dapat memperoleh pemahaman yang komprehensif mengenai kinerja algoritma *Gradient Boosting Tree* dalam mengklasifikasikan judul berita *clickbait* dalam bahasa Indonesia, sehingga hasil penelitian ini dapat menjadi landasan untuk pengembangan studi serupa di masa mendatang.

1.6 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini disusun ke dalam lima bab utama

yang saling berkaitan satu sama lain dan disajikan secara runtut untuk memudahkan pemahaman pembaca. Adapun uraian secara garis besar dari masing-masing bab adalah sebagai berikut:

1. Bab 1 – Pendahuluan

Bab ini membahas latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Bab ini memberikan dasar dan konteks awal mengenai alasan dilakukannya penelitian serta apa yang ingin dicapai.

2. Bab 2 – Tinjauan Pustaka

Berisi uraian teori-teori dan penelitian terdahulu yang relevan dengan topik penelitian. Tinjauan ini mencakup landasan teori, konsep-konsep penting, dan referensi yang menjadi acuan dalam merancang serta mengembangkan model yang diteliti.

3. Bab 3 – Metode Penelitian

Pada bab ini dijelaskan secara rinci tentang objek penelitian, desain atau alur penelitian, serta alat dan bahan yang digunakan. Selain itu, dijabarkan juga metode pengumpulan dan pengolahan data, sehingga proses penelitian dapat direplikasi atau dikembangkan lebih lanjut.

4. Bab 4 – Hasil dan Pembahasan

Bab ini menyajikan hasil dari eksperimen atau implementasi yang telah dilakukan. Pembahasan dilakukan untuk mengevaluasi dan menginterpretasikan hasil penelitian berdasarkan tujuan dan rumusan masalah yang telah ditetapkan.

5. Bab 5 – Penutup

Bab terakhir memuat kesimpulan dari penelitian yang telah dilakukan, serta saran-saran yang dapat dijadikan masukan untuk penelitian selanjutnya atau pengembangan lebih lanjut dari hasil penelitian yang ada.