

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, informasi merupakan sebuah kebutuhan bagi setiap orang. Dengan kemajuan teknologi internet akan menghasilkan peningkatan besar dalam penyebaran informasi, terutama informasi berita elektronik Berbahasa Indonesia. Sebelum memasuki era digital, untuk mendapatkan berita terbaru dan relevan masyarakat biasanya mencari berita melalui media cetak, televisi, koran, ataupun radio. Hal ini membuat terkadang masyarakat kurang *update* terhadap berita karena keterbatasan aksesibilitas dan kecepatan distribusi informasi. Mereka harus menunggu edisi cetak berikutnya atau siaran berita berikutnya untuk mendapatkan pembaruan terkini. Saat ini, dengan masuknya era digital dapat memudahkan masyarakat untuk mendapatkan berita terupdate dengan menggunakan Media Online. Media online bagaikan tambang informasi yang digemari banyak orang, tua maupun muda, menjadikannya sumber utama untuk mencari berbagai pengetahuan.. Menurut survei *Reuters Institute* terbaru bertajuk *Digital News Report 2023*, Media Online memiliki tingkat survei tertinggi dengan angka mencapai 84% dibandingkan dengan Media Sosial sebesar 65%, Televisi sebesar 54%, dan Media Cetak sebesar 15% (Annur, 2023). Survei ini dilakukan pada tahun 2023 terhadap 2012 responden. Dengan berkembangnya informasi melalui media online saat ini, Portal berita menjadi salah satu pilihan bagi masyarakat untuk memperoleh informasi terbaru dan relevan..

Berita merupakan sebuah kumpulan informasi yang membahas mengenai suatu fenomena, pandangan, kegemaran, suasana, kejadian, yang menjadi sebuah tren atau perbincangan di kalangan masyarakat. Berita dibuat oleh seorang jurnalis yang telah mengumpulkan dan meneliti informasi melalui wawancara dengan narasumber, observasi, dan pengamatan. Informasi yang terkumpul nantinya akan diproses dan disajikan dalam bentuk artikel yang memuat fakta-fakta yang relevan dan penting. Perpindahan berbagai berita dari media cetak ke media online membuat informasi lebih mudah menyebar di kalangan masyarakat. Portal berita online atau perusahaan media digital dapat memanfaatkan kondisi ini untuk

bersaing menarik perhatian pembaca agar mengunjungi situs mereka, yang pada akhirnya dapat menyesatkan pembaca. Fenomena tersebut disebut dengan *Clickbait* yang digunakan media online untuk meningkatkan jumlah kunjungan pengunjung

Clickbait merupakan jenis konten web yang dirancang dengan judul menggunakan formula dan tata bahasa tertentu dengan maksud untuk menarik perhatian pembaca agar mengklik tautan (Palau & Sampio, 2016). Hal ini dapat dilihat dari isi konten berita dengan judul berita yang tidak relevan dan terlihat melebih-lebihkan untuk menarik minat pembaca. Kejadian seperti ini sering terjadi agar portal berita online dapat meraup keuntungan dan menaikkan *rating* maupun jumlah pembaca saat mengeklik berita yang tidak sesuai dengan judul beritanya. Penggunaan *clickbait* yang berlebihan di media online dianggap berbahaya, mengingat masih banyaknya masyarakat yang tingkat literasinya tidak sebanding dengan jumlah pengguna internet. Kebanyakan masyarakat masih belum mampu membedakan antara berita hoaks, sindiran, atau bahkan tidak membaca berita secara menyeluruh, hanya melihat judulnya dan menarik kesimpulan dari judul tersebut. Maka dari itu, perlu adanya sistem otomatis yang dapat mengklasifikasi berita *clickbait* dengan memanfaatkan teknik *text mining*.

Teknik *text mining* adalah model keterbaruan dari metode data mining dan dapat digunakan untuk mengatasi permasalahan dalam mengklasifikasi berita *clickbait*. *Text mining* merupakan sebuah proses yang dilakukan untuk mengidentifikasi informasi atau tren terbaru yang belum terungkap dengan menganalisis dan memproses sejumlah besar data. (Alhaq et al., 2021). *Text Mining* memiliki algoritma yang dirancang untuk dapat mengidentifikasi data yang bersifat semi-terstruktur, seperti sinopsis, abstrak, atau isi dari dokumen-dokumen (Gupta & Lehal, 2009). Dalam menganalisis teks yang bersifat tidak terstruktur, *text mining* akan menghubungkan antara berbagai bagian teks sesuai dengan aturan-aturan tertentu. Biasanya *text mining* melibatkan proses seperti klasifikasi teks atau informasi, ekstraksi entitas atau konsep, pembentukan taksa umum, dan pengelompokan teks. Selain itu, *text mining* dapat didefinisikan sebagai proses mengekstraksi data dari teks atau dokumen untuk menemukan berbagai kalimat yang bisa mendukung isi dalam dokumen sehingga memungkinkan untuk dilakukan analisis hubungan dalam *text mining* (Betesda, 2016). Sebelum melakukan analisis

text menggunakan metode dalam *text mining*, langkah awal yang perlu dilakukan adalah *pre-processing teks*.

Text Processing adalah suatu proses pengolahan teks yang melibatkan berbagai teknik dan algoritma dengan tujuan mengekstrak informasi dari teks. Terdapat beberapa tahapan untuk melakukan pengolahan teks atau *text processing*, seperti pembersihan teks, ekstraksi fitur, dan klasifikasi teks. Setelah itu akan dilakukan metode klasifikasi berita *clickbait*. Klasifikasi adalah proses di mana suatu dokumen ditetapkan ke dalam kategori atau kelas tertentu sesuai dengan karakteristik yang dimilikinya (Nurhadi, A., 2016). Banyak algoritma yang dapat digunakan dalam klasifikasi text, seperti *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbour* (KNN), *Artificial Neural Network* (ANN), dan *Support Vector Machines* (SVM). Pada penelitian ini, akan menggunakan metode *Multinomial Naïve Bayes* karena metode ini sangat efisien dalam pembelajaran, terutama pada dataset yang relatif besar. Selain itu, kemudahan implementasi hasil dari metode *Naïve Bayes* menjadi pilihan yang baik untuk mengklasifikasi judul berita.

Naive Bayes adalah metode klasifikasi yang menerapkan Teorema Bayes dengan prinsip probabilitas yang sederhana, dengan asumsi tingkat tidak ketergantungan (independen) yang tinggi antar-fitur (Nurhuda, Sihwi, & Doewes, 2016). Beberapa kelebihan dari menggunakan metode *Naive Bayes* adalah memiliki perhitungan yang cepat, sederhana, dan memiliki tingkat akurasi yang tinggi (Arifin & Ariesta, 2019). Dalam Metode *Multinomial Naive Bayes*, jumlah data latih yang diproses cenderung kecil dan menjadi dasar untuk menghitung estimasi parameter pada proses klasifikasi data. Hal ini disebabkan oleh asumsi bahwa variabel yang digunakan dianggap sebagai independen, bukan dengan menggunakan keseluruhan matriks kovarians.

Penelitian sebelumnya telah banyak menggunakan Algoritma Multinomial Naive Bayes untuk mengklasifikasikan berita *clickbait*. Contohnya, penelitian oleh Fikri Alwan Ramadhan, Sampe Hotlan Sitorus, dan Tedy Rismawan (Ramadhan a tal., 2023) yang menerapkan Metode Multinomial Naive Bayes dengan Term Frequency - Inverse Document Frequency untuk Klasifikasi Judul Berita *Clickbait*. Mereka menggunakan total 1000 data, dengan 800 data untuk pelatihan dan 200 data untuk pengujian. Evaluasi dilakukan menggunakan Confusion Matrix, yang

menghasilkan tingkat akurasi sekitar 65%, recall sekitar 65,69%, dan presisi sekitar 65,69%. Selanjutnya terdapat penelitian yang sama yang dilakukan oleh (Hovipah, Hearani, Jasril, & Syafria, 2023) Hovipah yaitu Klasifikasi *Clickbait* Menggunakan *Transformers*. Hasil terbaik dalam penelitian ini adalah tingkat akurasi sebesar 63%, dengan *precision* sekitar 0.63 dan *recall* sekitar 1, menggunakan pembagian data sebesar 10% untuk data uji dan 90% untuk data latih (Mori Hovipah et al., 2023). Hal ini menunjukkan bahwa penggunaan Algoritma *Multinomial Naive Bayes* dalam mengklasifikasikan berita *clickbait* telah memberikan hasil yang cukup baik. Harapan dilakukannya penelitian ini semoga bisa memberikan manfaat dalam mengidentifikasi judul berita *clickbait*, sehingga dapat menjadi tambahan informasi yang berguna dalam upaya pencegahan *clickbait* di masa mendatang. Selain itu, diharapkan metode ini dapat berperan sebagai sumber pengetahuan yang berharga terkait penggunaan *Multinomial Naive Bayes*.

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah dari latar belakang yang telah dijelaskan dalam penelitian ini :

1. Bagaimana implementasi algoritma *Multinomial Naïve Bayes* dalam mengklasifikasikan judul berita *clickbait* dan non *clickbait* ?
2. Bagaimana pengaruh tingkat hasil akurasi yang dihasilkan oleh *Grid Search* terhadap algoritma *Multinomial Naive Bayes* pada pengklasifikasian judul berita *clickbait*?

1.3 Tujuan

Berikut merupakan tujuan yang ingin dicapai berdasarkan latar belakang yang telah dijelaskan dalam penelitian ini:

1. Mengetahui implementasi algoritma *Multinomial Naïve Bayes* dalam mengklasifikasikan judul berita *clickbait* dan non *clickbait*.
2. Mengetahui pengaruh tingkat hasil akurasi yang dihasilkan oleh *Grid Search* terhadap algoritma *Multinomial Naïve Bayes* pada pengklasifikasian judul berita *clickbait*.

1.4 Manfaat

Berikut merupakan manfaat yang dapat diperoleh dari penelitian ini :

1. Melindungi pembaca dari informasi yang potensial tidak dapat dipercaya atau bersifat *clickbait*.
2. Memberikan wawasan mendalam tentang karakteristik dan pola yang sering muncul pada judul berita *clickbait*.
3. Memberikan kontribusi terhadap pengembangan model klasifikasi yang handal untuk digunakan dalam menghadapi tantangan judul berita *clickbait* di platform berita online.

1.5 Batasan Masalah

Pada penelitian ini terdapat beberapa batasan masalah sebagai berikut :

1. Penelitian ini terbatas pada penggunaan dataset judul berita *clickbait* yang diambil dari sumber dataset *CLICK-IDE*.
2. Data yang digunakan untuk penelitian ini diambil dalam periode waktu tertentu, yaitu dari bulan Maret hingga November 2019. Penelitian ini hanya mencakup tren *clickbait* yang terjadi dalam periode tersebut, sehingga perubahan atau pembaharuan tren *clickbait* setelah periode tersebut berada di luar lingkup penelitian ini.
3. Penelitian ini berfokus pada implementasi *Metode Multinomial Naive Bayes* dalam klasifikasi judul berita *clickbait*. Perbandingan dengan metode klasifikasi lainnya, seperti *Decision Tree* atau *Support Vector Machine*, berada di luar cakupan penelitian ini.
4. Optimasi yang dilakukan hanya menggunakan teknik *Grid Search* dengan parameter *alpha* pada *Multinomial Naive Bayes*. Penggunaan teknik optimasi lain seperti *Random Search* atau *Bayesian Optimization* serta optimasi parameter lain berada di luar lingkup penelitian ini.