

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Pada akhir November 2022, laboratorium kecerdasan buatan OpenAI yang didirikan oleh Elon Musk mengumumkan peluncuran robot kecerdasan buatan mereka, ChatGPT. Peluncuran *chatbot* cerdas ini menarik perhatian banyak orang di seluruh dunia dan diklaim berhasil mendapatkan 57 juta pengguna di bulan pertama setelah perilsan [1]. Keunggulan *chatbot* ini yang menjadi perbincangan yaitu dapat melakukan komunikasi tertulis yang interaktif dengan manusia dan menghasilkan respons jawaban yang sangat mirip dengan manusia [2].

Setelah sudah tiga bulan peluncuran ChatGPT, perbincangan mengenai ChatGPT masih dapat ditemukan pada forum diskusi, portal berita, dan media sosial. Warganet Indonesia juga turut memberikan reaksi dan opini mereka pada media sosial mengenai robot cerdas buatan OpenAI tersebut. Salah satu media sosial yang digunakan untuk menunjukkan reaksi dan opini mereka adalah Twitter. Twitter merupakan platform *microblogging* terbesar di dunia [3], *microblogging* platform Twitter memungkinkan penggunanya membuat pesan atau konten yang singkat baik berupa teks, foto dan video. Twitter juga merupakan salah satu media sosial yang paling banyak digunakan di Indonesia dengan jumlah pengguna mencapai 18,45 juta per Januari 2022. Ini menjadikan Indonesia menjadi negara urutan ke-lima dalam daftar pengguna twitter terbanyak di seluruh dunia [4]. Dengan jumlah ini, besar kemungkinan bahwa masyarakat Indonesia dapat mengungkapkan reaksi dan opini mereka terhadap ChatGPT melalui Twitter sehingga respon dan opini mereka dapat dijadikan sebagai entitas untuk mengetahui sentimen masyarakat mengenai hal ini.

Analisis sentimen adalah metode komputasi yang bertujuan untuk mengidentifikasi dan mengklasifikasi pendapat, emosi, evaluasi, sikap, dan penilaian orang terhadap topik, produk, layanan, organisasi, individu, atau kegiatan tertentu[5]. Tujuannya adalah untuk menelusuri pesan yang diposting di media sosial dan mengenali serta mengelompokkan pendapat orang sebagai positif, negatif, atau netral[6]. Dalam melakukan analisis sentimen dengan pendekatan *machine learning*, terdapat beberapa algoritma yang dapat digunakan. Namun untuk melakukan analisis sentimen pada penelitian ini digunakan dua klasifikasi

yaitu Multinomial Naive Bayes dan Logistic Regression untuk menemukan algoritma terbaik untuk melakukan analisis sentimen terkait ChatGPT.

Naive Bayes merupakan salah satu metode klasifikasi yang populer untuk klasifikasi teks dan merupakan metode yang paling sering digunakan sebagai teknik klasifikasi, hal ini dikarenakan keunggulan metode ini yang efisien, cepat dan sederhana [7]. Pada penelitian sebelumnya yang dilakukan oleh Dinar Ajeng, dkk., terdapat *gap* yang jauh yaitu 18.50% dari perbandingan akurasi menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM) sehingga dapat disimpulkan bahwa algoritma Naive Bayes lebih superior terhadap SVM pada penelitian tersebut [7]. Sedangkan pada penelitian lainnya yang dilakukan oleh Muhammad Yusril, dkk., ditemukan bahwa performa yang dimiliki algoritma Logistic Regression sangat tinggi pada akurasi dan presisi sehingga algoritma Logistic Regression lebih unggul dibandingkan Naive Bayes [8].

Berdasarkan dari kesimpulan penelitian sebelumnya, penelitian ini memfokuskan penerapan algoritma *Multinomial Naive Bayes* yang merupakan variasi dari *Naive Bayes* dan membandingkannya dengan algoritma *Logistic Regression* untuk mengetahui algoritma yang lebih unggul jika digunakan pada studi kasus analisis sentimen terkait ChatGPT. Penerapan metode TF-IDF juga dilakukan dalam penelitian analisis sentimen terhadap ChatGPT ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, maka berikut merupakan masalah yang dapat dirumuskan pada penelitian ini.

1. Bagaimana cara mengimplementasikan algoritma Multinomial Naive Bayes dan Logistic Regression pada analisis sentimen terhadap ChatGPT pada platform Twitter?
2. Bagaimana cara mengukur performa algoritma Multinomial Naive Bayes dan Logistic Regression pada analisis sentimen terhadap ChatGPT?

1.3 Batasan Permasalahan

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang diambil untuk penelitian ini berupa twit dengan kata kunci yang berhubungan dengan ChatGPT.

2. Data yang diambil merupakan twit sebanyak 16500 yang berbahasa Indonesia.
3. Algoritma yang digunakan pada penelitian ini adalah Multinomial Naive Bayes dan Logistic Regression.
4. label sentimen pada penelitian ini terbagi dua, yaitu positif dan negatif. *Labelling* data akan dilakukan secara otomatis dengan bantuan *library* daftar kata opini berbahasa Indonesia.
5. Hanya twit yang diunggah pada tanggal 30 November 2022 hingga 28 Februari 2023 yang akan diambil.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Melakukan analisis sentimen terhadap opini masyarakat Indonesia mengenai *chatbot* ChatGPT menggunakan algoritma Multinomial Naive Bayes dan Logistic Regression pada media sosial Twitter.
2. Menemukan dan membandingkan performa dari Multinomial Naive Bayes dan Logistic Regression dalam sentimen analisis pada media sosial Twitter

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Mengetahui hasil sentimen masyarakat Indonesia terhadap aplikasi *chatbot* ChatGPT.
2. Menjadi referensi pada penelitian selanjutnya dengan tema yang berkaitan dengan topik ini.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penelitian ini terdiri dari 5 bab dengan rincian sebagai berikut.

- Bab 1 PENDAHULUAN
Bab 1 terdiri dari latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
- Bab 2 LANDASAN TEORI
Bab 2 menyajikan landasan teori dalam penelitian yang dilakukan. Seperti pengertian rumus Multinomial Naive bayes, Logistic Regression, dan Confussion Matrix.
- Bab 3 METODOLOGI PENELITIAN
Bab 3 membahas perancangan dan implementasi penelitian yang alurnya dijelaskan dengan menggunakan Diagram alur.
- Bab 4 HASIL DAN DISKUSI
Bab 4 menyajikan hasil implementasi, pengujian, serta evaluasi dari pengujian menggunakan metode Multinomial Naive bayes dan Logistic Regression.
- Bab 5 SIMPULAN DAN SARAN
Bab 5 membahas kesimpulan serta saran untuk pengembangan di waktu yang akan datang.

