

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Kemajuan teknologi digital mengalami pertumbuhan yang pesat, terdapat peningkatan para pengguna aplikasi *mobile*. Hal ini mendorong pentingnya pemahaman terhadap tingkat kepuasan dan keluhan pengguna dalam memakai sebuah aplikasi. *Google* menghadirkan *Play Store* sebagai platform dengan berbagai fitur dan jenis konten, seperti game, film, buku, dan berbagai aplikasi, dengan variasi kategori sesuai dengan keinginan pengguna. Satu fitur yang paling menonjol di *Play Store* adalah sistem *review* dan rating, dimana pengguna dapat memberikan *review* atau tanggapan mereka terhadap aplikasi yang mereka gunakan [1].

Aplikasi ChatGPT dikembangkan oleh OpenAI pada November 2022 yang merupakan aplikasi kecerdasan buatan (AI) [2]. ChatGPT dapat merespons interaksi manusia dan mesin dengan model jaringan saraf tiruan mendalam berparameter 175 miliar, fokus pada pemrosesan bahasa alami [3]. Keunggulan utama dari ChatGPT adalah memiliki fleksibilitas dan kemudahan penggunaan [4]. Setiap aplikasi memiliki aspek positif, netral dan negatif yang dapat menimbulkan berbagai tanggapan dari pengguna, seperti rasa puas atau kekecewaan terhadap suatu aplikasi [5]. Dilengkapi AI (Artificial Intelligence) dan NLP (Natural Language Processing), aplikasi ini berubah menjadi asisten pintar yang mampu menjawab berbagai macam jenis pertanyaan dari pengguna [6]. Mendeteksi suatu opini terhadap suatu subyek (misalnya individu, organisasi atau produk)

merupakan analisis sentimen [7]. Dalam menentukan ulasan termasuk dalam kategori positif, netral dan negatif suatu produk merupakan sebuah analisis sentimen.

Pada awal 2024, *ChatGPT* meraih rating 4.8 di *Google Play Store* yang sebagian besar merupakan ulasan positif, namun juga mendapat kritik negatif terkait jawaban tidak akurat dan dugaan kebohongan. Menurut penelitian [8] ketika diminta menjawab beberapa pertanyaan medis, *ChatGPT* memberikan jawaban yang kurang berkualitas sehingga terbatas jika digunakan untuk keperluan publikasi ilmiah. Namun, *ChatGPT* memberikan referensi yang tampak nyata. Pengguna *ChatGPT* harus memeriksa referensi yang diberikan sebelum ingin digunakan ke dalam naskah. Dengan banyaknya *review* yang diterima di *google play store* perlu dilakukannya analisis sentimen untuk mengetahui banyaknya masing-masing sentimen. Karena pengungkapan emosi seseorang banyak yang diekspresikan melalui bentuk tulisan termasuk ungkapan kepuasan ataupun kekecewaan yang disampaikan oleh pengguna di kolom komentar sebuah aplikasi. Maka dari itu, proses analisis sentimen yang dilakukan pada penelitian ini menggunakan teknik *text mining*. *Text mining* merupakan proses penambangan data (data mining) yang mengambil informasi dari data berupa teks seperti email, dokumen teks, file HTML serta data dari ulasan komentar [9].

Kualitas dari suatu kalimat sangat membantu dalam analisis sentimen, penggunaan metode *Term Frequency Invers Document Frequency* (TF-IDF) adalah sebuah metode yang banyak digunakan untuk menentukan seberapa jauh keterkaitan kata (*term*) pada suatu dokumen dengan memberi bobot pada setiap kata

[10]. TF-IDF merupakan perhitungan atau pembobotan untuk sebuah kata melalui teknik tokenisasi, stopwords, dan stemming, dan frekuensi munculnya kata dalam dokumen yang diberikan menunjukkan pentingnya kata itu di dalam sebuah dokumen[10].

Studi sebelumnya yang dilakukan oleh [11] dengan analisa relevansi tweet terhadap hashtag dengan metode *logistic regression*, mengumpulkan data menggunakan *twitter scraping* dan menghasilkan data sebanyak 2883 data. Penggunaan TF-IDF yang dimana setiap kata akan dihitung untuk mengukur tingkat signifikansi suatu kata dalam sebuah dokumen yang dapat membantu menaikkan akurasi. Akurasi yang didapatkan sebesar 74,53%.

Penelitian lain oleh [12] yang mengumpulkan data yang di ambil dari portal berita online dengan topik pemilu 2019. Data tersebut kemudian dilabeling menjadi tiga kelas yaitu positif, negatif dan netral. Pada penelitian ini menggunakan *logistic regression*. Dengan menggunakan Confusion Matrix, penelitian tersebut menemukan persentase akurasi 86%, persentase presisi 88%, dan persentase recall 86%.

Penelitian lain oleh [13] yang menggunakan *crawling* data untuk mengumpulkan data, menghasilkan 350 data *tweet*. Selanjutnya, melakukan pembagian data sebesar 80:20 untuk data pelatihan dan uji coba. Pembobotan tiap-tiap kata dalam *tweet* menggunakan TF-IDF. Setelah penerapan *logistic regression* didapatkan bahwa persentase nilai *accuracy* sebesar 78,57%, persentase nilai

precision sebesar 76,92%, persentase nilai *recall* sebesar 83,3%, dan persentase nilai *f1-score* sebesar 80%.

Studi selanjutnya oleh [14] data di ambil secara manual dari platform twitter sebanyak 161 data. Kumpulan data ini mencakup tweet yang diposting mulai dari tanggal 1 Januari 2022, dengan pencarian difokuskan pada tweet yang mengandung kata '*Tech Winter*' dan '*layoff*'. Persentase nilai hasil penelitian menggunakan algoritma *logistic regression* adalah 83,33%. F1-score untuk kelas positif adalah 0.73, menunjukkan ruang untuk perbaikan dalam keseimbangan antara presisi dan recall.

Dari penelitian-penelitian sebelumnya memiliki metode pengambilan data yang beragam seperti, Twitter, *crawling* data, dan pengumpulan dengan manual dari platform yang diinginkan. Penggunaan TF-IDF pada penelitian diatas membantu dalam pembobotan atau tingkat signifikansi suatu kata dalam dokumen. Pada penelitian ini dilakukan untuk melihat review yang diberikan oleh pengguna yang telah mengunduh Chatgpt melalui *platform Google Play store*.

Dengan pemaparan latar belakang diatas, maka penulis dapat melakukan penelitian analisis sentimen dengan menggunakan algoritma *logistic regression* dengan judul “**Penerapan Algoritma Logistic Regression untuk Analisis Sentimen Pada Review ChatGPT di Play Store**“ untuk melihat *review* termasuk dalam kategori positif, netral dan negatif.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang permasalahan yang telah diuraikan diatas maka dapat diambil rumusan yang akan menjadi pembahasan penelitian ini, yaitu :
“Bagaimana penerapan algoritma *Logistic Regression* untuk analisis sentimen pada *review* aplikasi chatgpt di *google playstore* ?”.

1.3 BATASAN MASALAH

Agar penelitian ini terfokus dan tidak keluar dari pembahan adapun batasan masalah sebagai berikut :

1. Data yang digunakan dalam penelitian ini merupakan dataset publik yang di ambil dari repository kaggle yang di beri judul ChatGPT reviews dengan jumlah data sebanyak 213,189 *review*.
2. Penelitian ini hanya menggunakan fitur yang ada dalam dataset.
3. Output yang dihasilkan dari data chatgpt_reviews berupa *accuracy, precision, recall dan F1-score*, mengetahui kata apa saja yang sering muncul serta mengetahui *review* positif, netral dan negatif.
4. *Tools* yang digunakan adalah *Google Colab*.

1.4 TUJUAN PENELITIAN

Adapun tujuan yang ingin dicapai pada penelitian dan penyusunan tugas akhir ini adalah :

1. Untuk penerapan algoritma *Logistic Regression* untuk analisis sentimen pada *review* aplikasi chatgpt di *google play store* .

2. Untuk mengukur tingkat akurasi dan mengklasifikasi nilai sentimen pada review aplikasi chatgpt di *google play store* mengetahui *review* positif, netral dan negatif serta mengetahui kata apa saja yang sering muncul.

1.5 MANFAAT PENELITIAN

Adapun manfaat yang didapat dari penelitian dan penyusunan tugas akhir ini adalah sebagai berikut :

1. Dapat mengetahui bagaimana tanggapan tentang aplikasi chatgpt yang telah diunduh dan digunakan.
2. Bagi penulis, dapat mengembangkan pengetahuan dan wawasan yang baru bahwa data review juga dapat di olah dan dilihat nilai sentimennya.
3. Bagi pembaca, dengan adanya penelitian ini dapat dijadikan sebagai referensi sumber pembelajaran untuk penelitian selanjutnya.

1.6 SISTEMATIKA PENULISAN

Adapun sistematika penulisan yang digunakan dalam menyusun penelitian ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini penulis membahas tentang latar belakang masalah, perumusan masalah, pembatasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab landasan teori ini membahas tentang teori-teori dan pendapat para ahli yang berhubungan langsung dengan permasalahan yang

dianalisis, antara lain : pengertian AI, text mining, analisis sentimen, *Logistic Regression*, ChatGPT.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang kerangka kerja penelitian, metode pengumpulan data, metode pengklasifikasian, dan alat bantu yang digunakan selama mengerjakan penelitian.

BAB IV : ANALISIS & HASIL

Pada bab ini akan dilakukan tahap analisis menggunakan algoritma *Logistic Regression* dengan memanfaatkan data *review* aplikasi ChatGPT yang telah dikumpulkan.

BAB V : PENUTUP

Pada bab ini membahas tentang kesimpulan dan saran selama melaksanakan proses peneliti.