

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah membawa perubahan signifikan dalam gaya hidup masyarakat, terutama dalam bidang kesehatan dan kebugaran [1]. Salah satu bentuk inovasi yang telah dikembangkan adalah aplikasi pelacak kebugaran (*fitness tracker*), yang membantu pengguna memantau aktivitas fisik seperti jarak yang telah ditempuh, kualitas tidur, asupan kalori, serta detak jantung [2]. Menurut Statista, jumlah pengguna di pasar kesehatan digital di Indonesia dinyatakan mencapai 37,85 juta pengguna pada tahun 2024. Mengikuti tren yang meningkat menunjukkan pencapaian sebesar 32,49 juta pengguna sejak tahun 2017. Jumlah pengguna juga akan meningkat sebanyak 14,41 juta orang di antara tahun 2024 dan 2029, ini menunjukkan peningkatan yang konsisten terhadap penggunaan aplikasi kebugaran digital [3]. Salah satu aplikasi tersebut adalah Google Fit, yang dikembangkan oleh Google LLC sebagai platform kesehatan yang dapat terhubung dengan berbagai perangkat dan aplikasi kebugaran lainnya [4], [5]. Aplikasi ini telah diunduh oleh lebih dari 100 juta pengguna dan memperoleh lebih dari 600 ribu *rating* dan ulasan, mencerminkan tingginya popularitas dan keinginan pengguna untuk memberikan pendapat mereka terhadap aplikasi. *Rating* tersebut umumnya diberi oleh pengguna untuk menunjukkan kepuasannya terhadap aplikasi. Namun menurut penelitian Lai, sulit untuk mengetahui perasaan pengguna sesungguhnya terhadap aplikasi Google Fit melalui *rating* ini. Oleh sebab itu, ulasan pengguna terhadap Google Fit perlu dianalisis karena ulasan-ulasan tersebut mengandung opini pengguna dalam bentuk teks dan juga membahas pengalaman pengguna terhadap fitur dan kinerja aplikasi yang dapat mempengaruhi keputusan calon pengguna untuk mengunduh aplikasi [6], [7], [8]. Menurut penelitian Ahn dan Park, data ulasan menjadi sumber utama untuk memahami emosi, harapan, masalah dari pengguna aplikasi, sehingga menjadi topik yang penting diteliti untuk mengetahui pandangan pengguna terhadap aplikasi [9].

Penelitian sebelumnya telah banyak membahas analisis sentiment pada ulasan aplikasi seperti SATUSEHAT, e-government, atau aplikasi layanan publik lainnya. Namun, masih terdapat keterbatasan penelitian yang secara khusus menganalisis ulasan pada aplikasi kebugaran atau pelacak kesehatan seperti Google Fit [10], [11], [12], [13]. Google Fit memiliki ekosistem yang terintegrasi dengan berbagai perangkat Android, Fitbit,

smartwatch, dan aplikasi pihak ketiga [14], Integrasi ini menghasilkan pengalaman pengguna yang berbeda, sehingga ulasan terhadap Google Fit cenderung lebih variatif, serta menunjukkan sentimen yang beragam, baik positif maupun negatif [7], [15], [16], [17]. Selain itu, jumlah ulasan yang besar dan tidak terstruktur, penggunaan bahasa gaul, singkatan, hingga sarkasme dapat menjadi tantangan dalam proses analisis sentimen [18]. Oleh sebab itu, dibutuhkan pendekatan berbasis *Natural Language Processing* (NLP) yang modern seperti *Bidirectional Encoder Representations from Transformers* (BERT), yang mampu memahami konteks kata secara dua arah dan menghasilkan representasi teks yang lebih akurat dibandingkan metode konvensional lainnya seperti SVM atau Naive Bayes [19]. Keunggulan BERT dalam analisis sentimen juga dibuktikan pada penelitian Joshy, yang menunjukkan bahwa BERT memperoleh performa lebih tinggi dibanding DistilBERT dan RoBERTa dengan akurasi percobaan 92.76% untuk data set *Sentiment 140* serta akurasi percobaan 90.43% untuk data set *Coronavirus tweets NLP* [20].

Selain analisis sentimen menggunakan BERT, identifikasi topik utama dalam ulasan juga penting untuk mengetahui aspek aplikasi yang paling banyak diperhatikan pengguna. Dalam hal ini, metode BERTopic dapat digunakan karena merupakan pendekatan *topic modeling* modern yang memanfaatkan representasi *embedding* berbasis *Transformer* serta teknik *clustering* untuk mengelompokkan ulasan berdasarkan kesamaan konteks [21], [22], [23]. Kombinasi BERT dan BERTopic memungkinkan analisis yang lebih komprehensif karena mampu mengidentifikasi sentimen sekaligus topik utama dalam ulasan pengguna Google Fit. Untuk itu, dilakukanlah penelitian berjudul “Analisis Sentimen pada Ulasan Pengguna Aplikasi Google Fit Menggunakan Metode BERT dan BERTopic”.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah yang dapat dijabarkan adalah tantangan dalam memahami sentimen pengguna serta mengidentifikasi topik utama yang dibahas dalam ulasan aplikasi Google Fit yang cenderung lebih bervariasi dan memiliki sentimen yang beragam.

1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah mengembangkan sebuah website yang dapat digunakan untuk menganalisis sentimen dan memodelkan topik pada ulasan aplikasi Google Fit dengan menggunakan BERT dan BERTopic.

1.4 Manfaat

Hasil dari penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan gambaran mengenai sentimen pengguna terhadap aplikasi Google Fit berdasarkan ulasan yang diberikan.
2. Membantu mengidentifikasi topik utama yang sering dibahas oleh pengguna dalam ulasan aplikasi Google Fit.
3. Menjadi referensi bagi penelitian selanjutnya yang membahas analisis sentimen dan identifikasi topik pada ulasan aplikasi kebugaran atau aplikasi digital sejenis.

1.5 Ruang Lingkup

Batasan masalah untuk penelitian ini adalah sebagai berikut:

1. Data ulasan bersumber dari kolom ulasan sebanyak 10.928 data yang diambil dari rentang waktu 1 Januari 2023 sampai 30 November 2025 di Google Play Store.
2. Pengambilan data menggunakan teknik *scraping* dengan *library* Google-Play-Scraper.
3. Ulasan yang dianalisis adalah ulasan pengguna berbahasa Inggris karena merupakan bahasa yang umum digunakan di dalam ruang Internet.
4. Ulasan yang dianalisis dikategorikan menjadi 3 kelas, yaitu positif, negatif, dan netral.
5. Pada model BERT, Pengujian menggunakan model *bert-base-uncased*. Parameter pelatihan yang diterapkan meliputi jumlah *epoch* 10, ukuran *batch* 16, 32, dan 64, nilai *learning rate* 1e-5, 3e-5, dan 5e-5, serta penggunaan *optimizer AdamW* [24], [25].
6. Pada model BERTopic, Pengujian menggunakan model *all-MiniLM-L6-v2*. *Embedding* tersebut kemudian diproses dengan konfigurasi UMAP (*n_neighbors*=20, *n_components*=10, dan *min_dist*=0.1), dan HDBSCAN (*min_cluster_size*=75) [26].