

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jumlah pengguna media sosial di Indonesia saat ini mencapai 191,4 juta pengguna pada awal tahun 2022, dan jumlah pengguna media sosial di Indonesia diperkirakan akan tumbuh sebanyak 21 juta (+12,6 persen) pada tahun 2021 – 2022. Media sosial paling populer adalah Twitter dengan total 18,45 juta pengguna di Indonesia pada awal tahun 2022 [1]. Twitter mencatat pertumbuhan kuat di Indonesia dengan *volume* Tweet sebesar 145,5 persen dan dapat menjadi peluang bagi para brand untuk menerapkan strategi pemasaran jelang momen belanja. Banyak orang datang ke Twitter untuk menemukan dan mencari rekomendasi tentang *brand* tertentu, melakukan *review* produk melalui utas, serta mendiskusikan produk yang sedang populer atau ramai dibicarakan [2].

Twitter memungkinkan brand dan perusahaan untuk terlibat dengan konsumen dengan cara yang otentik, memungkinkan brand untuk berbagi informasi tentang produk terbaru nya, menjawab pertanyaan, atau berinteraksi dengan konsumen dan bahkan dengan brand lain. Di Twitter, brand atau perusahaan dapat mencari nama mereka dan mengetahui tentang apa yang masyarakat katakan tentang produknya dan mempertimbangkan ide serta kritik [3].

Google Playstore ialah platform bagi masyarakat pengguna *smartphone* terkhususnya perangkat Android untuk men-*download* dan mencari aplikasi. Masyarakat dapat memberikan penilaian serta ulasan dan juga komentar terkait aplikasi yang mereka gunakan. Hal ini memberikan kesempatan bagi para peneliti, penguji, serta pengembang aplikasi untuk melakukan Analisa terkait aplikasi pengalaman masyarakat terhadap penggunaan aplikasi yang ada, yang mana hal tersebut tentunya memberikan kesan sentimen antara positif negatif ataupun netral, dan juga dapat menjadi umpan balik untuk pengembangan aplikasi kedepannya bagi para pengembang. Melalui kedua *platform* tersebut, Masyarakat dapat

beropini perihal produk, layanan, ataupun aplikasi yang digunakannya. Masyarakat dapat menuangkan pendapat, ulasan, dan menyebarkan pengalaman mereka kepada masyarakat.

Oleh karenanya, Penggunaan Twitter dan Google Playstore menjadi media yang bernilai sebagai sumber untuk mengumpulkan berbagai respon tanggapan masyarakat yang dapat digunakan untuk melakukan analisis sentimen guna pengembangan produk yang lebih baik. Oleh sebab itu dibutuhkan *website* yang dapat digunakan untuk menganalisis sentimen berbagai topik yang dapat digunakan secara *realtime*. Sebagian besar analisis sentimen sulit dilakukan dengan cara manual dikarenakan analisis data yang begitu banyak dari Twitter, maka dari itu *deep learning* dibutuhkan agar analisis dapat dilakukan lebih efektif dan efisien [4].

Aplikasi akan dibuat menggunakan IndoBERT yang memiliki beberapa keunggulan untuk analisis sentimen yaitu, IndoBERT dilatih khusus untuk bahasa Indonesia sehingga memiliki pemahaman bahasa yang lebih baik dibandingkan model bahasa umum lainnya, Memiliki pemahaman konteks yang lebih baik karena dapat memahami kalimat secara keseluruhan membuatnya lebih akurat dalam menilai kalimat yang panjang, serta untuk pelatihan lebih lanjut dapat Menggunakan *fine – tuning* untuk tujuan analisis sentimen tertentu dengan data yang relevan.

1.2 Tujuan dan Manfaat

1.2.1. Tujuan

Penelitian ini bertujuan untuk membuat sistem yang menampilkan menu analisis sentimen aplikasi analisis sentimen otomatis berbasis IndoBERT berdasarkan tanggapan pengguna Twitter dan Google Playstore. Penelitian ini dibutuhkan untuk memudahkan pengambilan keputusan dan analisa produk bagi para pelaku bisnis, layanan jasa, dan barang serta pengembang aplikasi dalam membaca data sentimen pada Twitter dan Google Playstore secara *realtime*.

Adapun tujuan dari pembuatan aplikasi ini adalah:

- a. membantu untuk mengetahui pendapat dan kepuasan masyarakat terhadap suatu produk, layanan, dan jasa tertentu.

- b. membantu pengidentifikasian tren yang sedang berkembang melalui analisis sentimen,
- c. membantu dalam mengevaluasi dan memantau reputasi suatu merek.

1.2. 2. Manfaat

Adapun manfaat dari pembuatan aplikasi ini adalah:

- 1.4.2 masyarakat mendapatkan akses ke ulasan suatu produk, layanan, dan jasa untuk mengetahui pendapat pelanggan,
- 1.4.3 terbantu untuk mengetahui kualitas suatu produk, layanan, dan jasa berdasarkan ulasan dan pendapat pelanggan,
- 1.4.4 terbantu dalam membuat keputusan yang lebih baik berdasarkan ulasan dan pendapat pelanggan.

1.3 Tinjauan Pustaka

Dalam rangka untuk memperkaya pengetahuan, dalam tugas akhir ini, kami mengangkat beberapa penelitian terdahulu untuk dijadikan referensi. R. Mas, dkk. [5] melakukan analisis sentimen terhadap aplikasi Ruang Guru di Google Playstore. *Review* komentar diambil menggunakan teknik *scrapping*. Data yang dihasilkan berjumlah 5.437 *records*. Hasil dari analisa sentimennya dapat dimanfaatkan untuk melihat respon pengguna Ruang Guru lalu pengembang bisa memaksimalkan fitur yang dirasa kurang oleh pengguna.

Dalam upaya untuk memahami analisis sentimen Bahasa Indonesia, pada peneliti terdahulu oleh H. Jayadianti, dkk. [6] dilakukan adaptasi penelitian untuk analisis sentimen pada ulasan Bahasa Indonesia menggunakan *pretrained* IndoBERT dan R-CNN. Penelitian oleh M. I. Akmal, dkk. [7] membahas tentang permasalahan bagaimana sebaran kata yang digunakan masyarakat dalam mengomentari kasus kebocoran data kartu SIM pada media sosial Twitter dengan menggunakan pendekatan empat metode, yaitu *Random Forest*, *Logistic Regression*, *Support-Vector Machine*, dan IndoBERT.

Pada penelitian sebelumnya oleh P. F. Supriyadi, dan Y. Sibaroni [8] dilakukan analisis produk tertentu menggunakan IndoBERT. Penelitian dilakukan pada produk dagang Xiaomi untuk menentukan opini masyarakat terhadap produk

tersebut menggunakan metode Naïve Bayes dengan data sebanyak 2.078 tweet dan menghasilkan akurasi yang cukup baik sebesar 71,88% dengan sebaran persentase sentiment positif 39%, netral 51%, dan 10% negatif. Kemudian mereka melakukan penelitian yang sama namun menggunakan IndoBERT.

Kemudian, terdapat penelitian lainnya oleh M. Ikram, dkk. [9] yang fokus membandingkan hasil akurasi LSTM dan IndoBERT dalam mendeteksi *hoax* menggunakan dataset yang diambil dari Twitter. Model IndoBERT-nya dilatih dengan korpus bahasa Indonesia. Model *fine-tuned* IndoBERT memiliki performa dengan akurasi 92,07%, sementara LSTM memiliki performa dengan akurasi 87,54%. Jadi model IndoBERT menunjukkan hasil performa yang jauh lebih tinggi dari LSTM.

1.4 Data Penelitian

Penelitian ini memanfaatkan dua jenis sumber data utama, yakni data *keyword* yang diperoleh melalui proses *crawling* dari platform Twitter, dan data hasil *scraping* dari Google Playstore. Untuk melatih dan menguji model, penelitian ini menggunakan IndoBERT, sebuah model dari IndoNLU yang merupakan suatu kumpulan sumber daya untuk penelitian dalam bidang Pemahaman Bahasa Alami (*Natural Language Understanding*) khususnya untuk Bahasa Indonesia [10].

IndoBERT, sebagai model dalam penelitian ini, telah terbukti baik dalam berbagai konteks pemahaman bahasa, menjadikannya pilihan yang tepat untuk analisis sentimen dalam konteks Bahasa Indonesia. Tambahan, *dataset* yang digunakan untuk melatih dan menguji model IndoBERT tersebut berasal dari penelitian sebelumnya yang berfokus pada analisis sentiment dengan tiga kelas, yaitu Positif, Negatif, dan Netral. Dataset ini terbagi menjadi data latih sebanyak 11.000, data validasi sebanyak 1.260, dan data uji sebanyak 500 [11].

Data yang didapat melalui proses *crawling* Twitter memberikan gambaran langsung tentang respon pengguna terhadap suatu topik atau kata kunci tertentu. Sementara itu, data yang diperoleh melalui *scraping* Google Playstore memberikan informasi lebih lanjut terkait ulasan dan rating pengguna terhadap aplikasi atau produk tertentu.