

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Aplikasi TikTok telah menjadi sangat populer di kalangan pengguna di seluruh dunia dalam beberapa tahun terakhir. Sejak kemunculannya pada tahun 2017, popularitas TikTok meningkat pesat. Per Maret 2023, aplikasi ini telah diunduh lebih dari 500 juta kali. Didistribusikan lebih dari 150 negara, terdapat hampir 800 juta pengguna aktif bulanan di seluruh dunia. Selain itu, pada tahun 2018 dan 2019, TikTok melampaui Facebook, YouTube, dan Instagram untuk menduduki posisi teratas sebagai aplikasi seluler yang paling banyak diunduh. Pandemi COVID-19 menjadi salah satu faktor yang meningkatkan pertumbuhan popularitas TikTok yang luar biasa (Isnaini dkk., 2023). Pengguna TikTok dapat memproduksi, melihat, berbagi dan terlibat dengan berbagai macam konten kreatif, termasuk *Lip-sync*, tarian, seputar informasi, komedi, tantangan dan hal-hal lain sesuai minatnya. Faktanya, TikTok telah menawarkan alat perdagangan sosial sejak September 2021, memungkinkan pengguna dan produsen mengiklankan dan menjual barang mereka melalui TikTok Shop (Aisyah Putri & Wayan Supriana, 2024). Sehingga pada penelitian akan digunakan data yang diambil dari review tiktok pada *google playstore* dengan kurun waktu maret 2023 sampai oktober 2024.

Pelabelan merupakan proses memberikan label atau kategori tertentu pada data agar dapat diidentifikasi dan digunakan dalam analisis lebih lanjut. Dalam pengolahan data, terutama dalam *machine learning*, pelabelan dilakukan dengan

menandai data mentah berdasarkan karakteristik atau kelas tertentu yang akan Dianalisis (Tri Julianto dkk., 2024). Misalnya, dalam analisis sentimen, teks atau ulasan diberi label seperti "positif," "negatif," atau "netral" berdasarkan sentimen yang dihasilkan dari teks tersebut. Pelabelan sangat penting karena mempermudah sistem dalam mengenali pola, mengelompokkan data, dan melakukan analisis yang lebih akurat (Harris Pratama & Hayaty, 2023a). Dalam *machine learning*, data yang telah diberi label membantu model untuk mengenali pola-pola yang sesuai dengan label. Pelabelan memungkinkan pemrosesan data menjadi lebih terstruktur, sehingga model *machine learning* dapat melakukan prediksi atau keputusan yang lebih andal (Azwarni & Shah, 2024).

*TextBlob* dan *VADER (Valence Aware Dictionary and sEntiment Reasoner) Lexicon* adalah dua teknik yang digunakan dalam analisis sentimen, terutama untuk pemrosesan bahasa alami. *TextBlob* adalah pustaka Python yang menyediakan alat untuk analisis sentimen dengan pendekatan berbasis aturan dan kamus *lexicon* (Nur Syahirah Wan Min & Zareen Zulkarnain, 2020). *TextBlob* dipilih pada penelitian ini karena memberikan skor polaritas dan subjektivitas pada teks, sehingga cocok untuk analisis sentimen umum. Namun, *TextBlob* memiliki keterbatasan dalam menangani bahasa gaul, slang, atau nuansa emosi dalam teks informal, yang dapat mengurangi akurasi pada teks yang tidak baku atau konteks khusus (Biswas dkk., 2022). Sedangkan, Pemilihan *VADER Lexicon* pada penelitian ini dikarenakan algoritma yang digunakan berbasis aturan yang dirancang untuk menangani teks di media sosial, yang sering menggunakan bahasa informal dan singkat. *VADER Lexicon* unggul dalam mengenali intensitas

sentimen dengan mempertimbangkan kata kunci seperti kapitalisasi, tanda seru, dan emotikon, Sehingga memberikan analisis sentimen yang lebih akurat untuk teks singkat seperti tweet atau ulasan singkat. Namun, *VADER Lexicon* kurang akurat ketika digunakan pada teks panjang atau teks yang kompleks, seperti artikel atau esai, karena kesulitan dalam menangani konteks yang lebih mendalam (Nurchayawati & Mustaffa, 2023).

Analisis sentimen adalah proses memahami, mengekstraksi, dan menganalisis materi tekstual secara otomatis untuk mengumpulkan informasi tentang kecenderungan opini terhadap suatu item, apakah kecenderungan tersebut positif, negatif, atau netral (Parlika dkk., 2020). Meskipun TikTok telah menjadi salah satu aplikasi media sosial paling populer dengan jutaan pengguna aktif, masih terdapat beragam pandangan dan kontroversi seputar platformnya. Oleh karena itu, analisis sentimen terhadap ulasan TikTok di *Googleplay Store* memiliki peran penting dalam memahami pengalaman dan persepsi pengguna (Aisyah Putri & Wayan Supriana, 2024). Dengan menggunakan analisis ini dapat mengidentifikasi isu-isu spesifik serta fitur-fitur TikTok yang mendapat tanggapan positif, negatif dan netral dari pengguna. TikTok dapat menggunakan hasil analisis untuk menentukan arah pengembangan platform, memprioritaskan perbaikan fitur tertentu, atau merumuskan kebijakan penggunaan yang lebih sesuai dengan kebutuhan dan harapan pengguna (Sukirman dkk., 2024). Dengan demikian, analisis sentimen menjadi alat yang efektif untuk meningkatkan kualitas aplikasi dan kepuasan pengguna secara keseluruhan.

XGBoost (*EXtreme Gradient Boosting*) adalah salah satu algoritma

*Machine learning* yang sangat populer dan kuat, terutama untuk tugas-tugas prediksi dan klasifikasi. Algoritma ini menggunakan teknik boosting, di mana model dibangun secara bertahap dengan menambahkan pohon keputusan baru yang memperbaiki kesalahan dari pohon sebelumnya. *XGBoost* bekerja dengan membentuk serangkaian pohon keputusan secara berurutan. Setiap pohon berusaha untuk memperbaiki kesalahan prediksi dari model sebelumnya dengan memberikan bobot yang lebih tinggi pada data yang sulit diprediksi (Natras dkk., 2022). Model pertama mungkin hanya memiliki akurasi rendah, tetapi dengan setiap iterasi baru, model ditingkatkan dengan menambahkan pohon yang fokus pada kesalahan yang terjadi sebelumnya. Secara umum, *XGBoost* menggunakan teknik *gradient boosting*, di mana kesalahan yang diperbaiki pada setiap langkah adalah residual atau selisih antara nilai prediksi dan nilai aktual (Aqbar & Andrianti Supono, 2023).

Penelitian sebelumnya yang dilakukan (Siva Asvia, 2023) mengenai Perbandingan *Naïve Bayes* dan *Random forest* pada Analisis Sentimen Reputasi *Brand Provider Live.on* menghasilkan *Random forest* memiliki akurasi sebesar 80,37%. Ketika teknik SMOTE diterapkan, *Random forest* tetap mempertahankan performa yang lebih baik dengan akurasi 78%, Sementara *Naïve Bayes* Memiliki performa sama pada semua metrik dengan nilai 73%. Berdasarkan perhitungan NRS dengan menggunakan hasil prediksi *Naïve Bayes*, reputasi *provider Live.On* memiliki nilai -33,08%, dan hasil prediksi menggunakan *Random Forest*, reputasi mencapai -41,08%. Karena memiliki nilai reputasi dibawah rata-rata maka, Penelitian ini dianggap memiliki reputasi yang buruk sehingga dibutuhkan untuk

meningkatkan reputasinya. Penelitian berikutnya yang dilakukan (Yuda Maulana, 2023) mengenai Optimalisasi *Support Vector Machine* (SVM) Menggunakan Pelabelan Vader pada Analisis Sentimen Ulasan Google Classroom menghasilkan nilai akurasi 90,15%. Hasil tersebut membuktikan bahwa dengan menerapkan metode Vader dapat mengoptimasi performa algoritma *Support Vector Machine* (SVM).

Berdasarkan seluruh uraian pekerjaan terkait yang telah dibahas tersebut, terdapat kesenjangan penelitian yang masih tersisa, dari penelitian (Siva Asvia, 2023) pelabelan hanya dilakukan berdasarkan *rating* dan komentar yang diketahui ketika membaca teks secara manual. Selanjutnya penelitian (Yuda Maulana, 2023) hanya menggunakan metode pelabelan *VADER* sehingga diperlukan metode pelabelan lain, sebagai pembanding terhadap optimalisasi kinerja algoritma klasifikasi sehingga dapat terlihat metode pelabelan mana yang memiliki optimalisasi yang lebih baik. Untuk melengkapi penelitian sebelumnya, penelitian ini akan dilakukan perbandingan antara pelabelan *VADER lexicon* dan *textblob* untuk menemukan performa kerja terhadap klasifikasi algoritma yang lebih baik.

Penelitian ini menggunakan algoritma *XGBoost* karena berdasarkan survey paper (Khatib Sulaiman et al., n.d.-a) diperoleh hasil algoritma *XGBoost* lebih akurat dibandingkan algoritma *Support Vector Machine* (SVM). Karena *XGBoost* menggunakan teknik ensemble yang lebih canggih dan optimasi yang lebih baik. Penelitian berikutnya dilakukan oleh (Christanto dkk., 2023) diperoleh *XGBoost* sebagai algoritma terbaik untuk analisis *sentimen review* hotel di trip advisor dengan akurasi mencapai 99% dibandingkan dengan algoritma Decision

Tree dengan akurasi 97% dan algoritma *Support Vector Machine* (SVM) dengan akurasi 98% oleh karna itu, penelitian ini dipilih algoritma *XGBoost* untuk membandingkan teknik labelling *VADER lexicon* dan *textblob* yang diharapkan mampu menghasilkan nilai yang lebih akurat daripada penelitian sebelumnya.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka permasalahan yang akan dibahas adalah sebagai berikut:

1. Bagaimana cara pengaplikasian dari labelling *VADER lexicon* dan *textblob* menggunakan algoritma *XGBoost* pada analisis sentimen ulasan aplikasi tiktok untuk menganalisis sentimen positif, negatif dan netral?
2. Bagaimana performa yang dihasilkan dari perbandingan pelabelan menggunakan *VADER lexicon* dan *textblob* pada ulasan aplikasi tiktok menggunakan algoritma *XGBoost* ?

## 1.3 Tujuan Penelitian

1. Mengaplikasikan cara kerja dari labelling *VADER lexicon* dan *textblob* menggunakan algoritma *XGBoost* pada analisis sentimen ulasan aplikasi tiktok untuk menganalisis sentimen positif, negatif dan netral.
2. Membandingkan pelabelan menggunakan *VADER lexicon* dan *textblob* pada ulasan aplikasi tiktok menggunakan algoritma *XGBoost*.

## 1.4 Manfaat Penelitian

Penelitian ini memiliki manfaat yang signifikan, baik secara teoritis maupun praktis antara lain:

1. Manfaat penelitian ini secara teoritis adalah mampu memahami bagaimana

kinerja labelling *VADER lexicon* dan *textblob* pada analisis sentimen ulasan aplikasi tiktok dengan menggunakan algoritma *Xgboost*.

2. Secara praktis, penelitian ini dapat Mengetahui berbagai pendapat masyarakat baik komentar positif, netral ataupun negatif dari aplikasi tiktok, sehingga dari banyaknya pandangan tersebut dapat digunakan untuk meningkatkan kualitas aplikasi tiktok itu sendiri.

### 1.5 Batasan Masalah

Penelitian ini nantinya dilakukan pada data yang akan di peroleh dari ulasan aplikasi tiktok pada *google playstore*.

1. Penelitian ini akan dilakukan berdasarkan labelling antara *VADER lexicon* dan *Textblob* mengenai analisis sentiment.
2. Penelitian ini nantinya menggunakan algoritma *XGBoost (eXtreme Gradient Boosting)*.
3. Data yang digunakan terbatas terhadap review aplikasi tiktok pada *google play store* dari Maret 2023 sampai Oktober 2024.

### 1.6 Sistematika Penulisan

Sebagai langkah untuk mempermudah pemahaman serta penyusunan laporan tugas akhir ini, disusun sistematika penulisan yang menjelaskan isi dari setiap bagian dalam penelitian. Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut:

## BAB I PENDAHULUAN

Bab ini menguraikan gambaran umum penelitian yang mencakup latar belakang yang menjadi dasar pelaksanaan penelitian, perumusan masalah, tujuan penelitian,

manfaat penelitian, batasan masalah, serta sistematika penulisan sebagai panduan dalam penyusunan laporan.

## **BAB II TINJAUAN PUSTAKA**

Bab ini menguraikan perencanaan penelitian, tahapan-tahapan yang dilalui, serta metode yang diterapkan dalam penelitian. Pembahasan mencakup secara menyeluruh peta jalan penelitian serta langkah-langkah sistematis yang dilakukan dari awal hingga akhir proses penelitian.

## **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan rancangan penelitian, tahapan yang dilakukan, serta metode yang digunakan dalam penelitian. Pembahasan dalam bab ini mencakup prosedur pelaksanaan penelitian secara sistematis agar dapat dipahami oleh peneliti lain.

## **BAB IV HASIL DAN PEMBAHASAN**

Bab ini memuat pembahasan dan analisis terhadap hasil yang diperoleh dari proses pengujian, sesuai dengan tujuan dan rumusan masalah penelitian.

## **BAB V KESIMPULAN DAN SARAN**

Bab ini menguraikan kesimpulan yang diperoleh dari hasil penelitian, serta dilengkapi dengan saran yang dapat dijadikan bahan pertimbangan untuk pengembangan dan penelitian lebih lanjut.