

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Tempat wisata menjadi salah satu tujuan utama masyarakat untuk menghabiskan waktu libur dan mencari kesenangan di luar rutinitas sehari-hari. Selain bersantai, kunjungan ke tempat wisata juga berfungsi untuk menyegarkan pikiran dan mempererat hubungan sosial dengan orang-orang terdekat [1]. Salah satu provinsi di Indonesia yang memiliki banyak destinasi wisata menarik perhatian adalah Jawa Timur, dengan Kabupaten Magetan sebagai salah satu daerah wisatanya. Kabupaten Magetan memiliki wisata alam yang mempesona dan menarik perhatian yaitu Telaga Sarangan atau yang dikenal sebagai Telaga Pasir. Telaga Sarangan sendiri terletak di kaki Gunung Lawu memiliki kedalaman 28 meter dengan luas 30 hektar, berjarak 16 kilometer dari Kota Magetan, dan berada pada ketinggian 1287 Mpd [2]. Keindahan alam dan kesejukan udara menjadikan Telaga Sarangan menjadi wisata yang populer baik kalangan wisatawan lokal maupun luar daerah.

Perkembangan teknologi mempermudah akses informasi tentang tempat wisata. Aplikasi *Google Maps* merupakan sebuah aplikasi berbasis internet diakses melalui web browser maupun ponsel untuk membantu pengguna menemukan informasi suatu daerah atau tempat yang belum diketahui sebelumnya, termasuk Telaga Sarangan [3]. Pada *Google Maps* pengguna melihat ulasan yang dapat dijadikan sebagai referensi dalam menentukan destinasi wisata sesuai dengan preferensi mereka. Sehingga dapat disimpulkan bahwa ulasan pada *Google Maps* berperan penting sebagai bahan pertimbangan bagi pihak pengelola Telaga Sarangan serta diperlukan analisa terkait ulasan dengan menggunakan analisis sentimen untuk memahami pandangan pengunjung secara lebih mendalam [4].

Analisis sentimen merupakan salah satu bidang *Natural Language Processing* (NLP) yang banyak diterapkan pada penelitian dengan mengidentifikasi opini atau emosi dari teks ulasan media sosial, seperti *Google Maps*. Melalui analisis sentimen, informasi yang tersebar di media sosial dapat diubah menjadi data yang lebih terstruktur dan bermanfaat sehingga memungkinkan pengambilan keputusan yang lebih baik [5]. Salah satu pendekatan analisis sentimen adalah multi-class

yang mengkategorikan teks kedalam sentimen positif, negatif, atau netral sehingga memungkinkan penanganan data teks dalam skala besar. Tujuan dari sentimen multi-class adalah membangun prediktif model secara akurat dan efisien kedalam beberapa kelas berdasarkan konteks tertentu [6].

*Robustly Optimized BERT Pre-training Approach (RoBERTa)* adalah pengembangan dari model BERT yang dimodifikasi untuk meningkatkan kinerja melalui perbaikan prosedur pra-pelatihan. RoBERTa berfokus pada peningkatan skala *batch* dan *byte*, sehingga memungkinkan pelatihan dengan *dynamic masking* tanpa memerlukan prediksi terhadap kalimat berikutnya [7]. Penggunaannya semakin populer dalam penelitian terbaru karena kemampuannya yang unggul dalam berbagai aplikasi *Natural Language Processing* (NLP), termasuk analisis sentimen. Model ini sangat efektif dalam memahami konteks bahasa dan menjawab pertanyaan singkat tanpa memerlukan data yang dianotasi [8].

Penelitian dengan judul “*Detection of Indonesian Hate Speech in the Comments Column of Indonesian Artists' Instagram Using the RoBERTa Method*” menunjukkan performa model RoBERTa bervariasi tergantung pada metode pre-processing dan *hyperparameters* yang dipakai. Pada skenario pertama, dengan penerapan full pre-processing, akurasi tertinggi yang dicapai adalah 84,81% menggunakan *batch\_size* 8 dan *test\_size* 0,1. Namun, akurasi mengalami penurunan seiring dengan peningkatan *batch\_size* dan *test\_size*. Di sisi lain, pada skenario kedua yang hanya menggunakan pre-processing hingga tahap normalisasi, akurasi lebih tinggi tercapai, yaitu 85,09% dengan *batch\_size* 16, menunjukkan bahwa RoBERTa bekerja lebih efektif pada data yang tidak diproses secara berlebihan [9]. Dalam penelitian ini, RoBERTa digunakan untuk menganalisis sentimen ulasan Telaga Sarangan karena kemampuannya yang kuat dalam menangkap konteks bahasa sehingga menghasilkan hasil yang lebih akurat.

## **1.2 Perumusan Masalah**

### **1.2.1 Permasalahan**

Pada aplikasi *Google Maps*, terdapat berbagai ulasan mengenai wisata Telaga Sarangan dengan variasi yang cukup beragam dan jumlah yang banyak. Hal ini menyulitkan pengunjung untuk membaca semua ulasan secara detail karena akan memakan banyak waktu. Oleh karena itu, diperlukan sebuah analisis sentimen

yang dapat mempermudah proses analisis dan memberikan gambaran umum mengenai tingkat kepuasan pengunjung. Permasalahan dalam penelitian ini adalah bagaimana mengkategorikan ulasan wisata Telaga Sarangan pada Google Maps berdasarkan sentimen multi-class secara akurat. Dengan sentimen multi-class diharapkan analisis dapat memberikan gambaran tingkat kepuasan pengunjung lebih terperinci dan bermanfaat bagi pengelola dalam meningkatkan layanan wisata.

### 1.2.2 Metode Usulan

Penelitian ini menggunakan metode *Robustly Optimized BERT Pretraining Approach (RoBERTa)* untuk melakukan analisis sentimen multi-class pada ulasan Telaga Sarangan yang diambil dari *Google Maps*. Metode RoBERTa dipilih karena kemampuannya yang lebih unggul dalam menangkap konteks bahasa dan menghasilkan klasifikasi sentimen yang lebih akurat. Selain itu, penelitian ini juga menerapkan teknik *Synthetic Minority Over-sampling Technique (SMOTE)* untuk mengatasi ketidakseimbangan data pada dataset ulasan yang digunakan.

### 1.2.3 Pertanyaan Penelitian

Berdasarkan uraian latar belakang pertanyaan penelitian yang diajukan adalah bagaimana pengaruh kombinasi *hyperparameter* dan teknik penanganan *imbalance* data terhadap kinerja model RoBERTa dalam klasifikasi sentimen *multi-class* pada ulasan wisata Telaga Sarangan?

## 1.3 Tujuan dan Manfaat

### 1.3.1 Tujuan

1. Mengukur tingkat akurasi metode *Robustly Optimized BERT Pretraining Approach (RoBERTa)* dalam melakukan analisis sentimen *multi-class* terhadap ulasan pengunjung Telaga Sarangan di aplikasi *Google Maps* dengan pengaruh variasi *hyperparameter (batch size dan epoch)* terhadap performa model RoBERTa.
2. Menilai efektivitas teknik *Synthetic Minority Over-sampling Technique (SMOTE)* dalam meningkatkan keseimbangan data dan akurasi klasifikasi sentimen.

### 1.3.2 Manfaat

1. Memberikan wawasan terkait pengaruh hyperparameter dan pemahaman tentang teknik SMOTE serta dampaknya terhadap keseimbangan data dan hasil klasifikasi sentimen dalam konteks ulasan online dalam penerapan model RoBERTa pada klasifikasi sentimen, yang dapat membantu peneliti dan praktisi dalam mengoptimalkan model NLP di bidang pariwisata.
2. Meningkatkan pemahaman tentang teknik SMOTE dan dampaknya terhadap keseimbangan data dan hasil klasifikasi sentimen dalam konteks ulasan online, yang dapat diterapkan pada data ketidakseimbangan dalam penelitian lain.
3. Memberikan kontribusi praktis untuk pengelola wisata dengan menyediakan analisis sentimen yang lebih akurat dan terstruktur dari ulasan pengunjung, yang berguna untuk meningkatkan layanan dan pengalaman wisata di Telaga Sarangan.

### 1.4 Batasan Masalah

1. Data yang digunakan dalam penelitian ini merupakan ulasan Telaga Sarangan pada aplikasi *Google Maps* yang diperoleh melalui *scraping* sebanyak 3875 data.
2. Penelitian menggunakan sentimen *multi-class* dimana terdapat tiga kelas yaitu kelas positif, negatif, dan netral.
3. Metode yang digunakan adalah *Robustly Optimized BERT Pretraining Approach* (RoBERTa) dengan teknik *imbalance* data.

### 1.5 Sistematika Penulisan Proposal Skripsi

Terdapat beberapa bab yang tersusun dalam sistematika penulisan sesuai dengan format pada program studi Sistem Informasi Universitas Trunojoyo Madura, antara lain:

## BAB I PENDAHULUAN

Bab ini berisi tentang penjelasan latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan proposal penelitian

## **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tentang analisis sentimen, *Google Maps*, *Robustly Optimized BERT Pretraining Approach* (RoBERTa), serta teori lainnya yang menunjang penelitian.

## **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan tahapan-tahapan penelitian yang dilakukan, mulai dari pengumpulan data, *pre-processing* data, hingga metode analisis sentimen yang digunakan, yaitu *Robustly Optimized BERT Pretraining Approach* (RoBERTa).

## **BAB IV HASIL DAN PEMBAHASAN**

Bab ini menjelaskan hasil penelitian proses klasifikasi sentiment multi-class terhadap ulasan wisata Telaga Sarangan menggunakan *Robustly Optimized BERT Pretraining Approach* (RoBERTa).

## **BAB V PENUTUP**

Bab ini berisi kesimpulan hasil penelitian serta saran untuk penelitian selanjutnya.