

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Pemrosesan Bahasa Alami (*Natural Language Processing, NLP*) merupakan cabang utama kecerdasan buatan yang berkembang pesat, terutama dalam analisis sentimen. Analisis sentimen digunakan untuk memahami opini dalam teks, seperti komentar video di YouTube, yang mencerminkan persepsi audiens terhadap konten. Dengan meningkatnya volume komentar, diperlukan metode analisis yang akurat dan efisien [1]. YouTube sebagai sumber data menawarkan tantangan unik. Komentar video mencakup beragam opini yang sering kali tidak terstruktur, dengan penggunaan bahasa informal, singkatan, dan emotikon yang membuat analisis sentimen menjadi kompleks[2].

Arsitektur *Long Short-Term Memory (LSTM)* dalam *Recurrent Neural Networks (RNN)* telah terbukti efektif dalam memproses data sekuensial seperti teks. *LSTM* dirancang untuk menangkap hubungan temporal jangka panjang melalui mekanisme khusus yang mengatasi masalah *vanishing gradient* pada *RNN* standar [3]. Namun, kinerja *LSTM* dipengaruhi oleh konfigurasi arsitektur, termasuk jumlah lapisan dalam model.

Penelitian [4] menunjukkan bahwa penambahan jumlah lapisan pada *LSTM* tidak selalu meningkatkan akurasi dan dapat menyebabkan *overfitting*. Model dengan satu lapisan cenderung lebih cepat dan sederhana, tetapi mungkin kurang mampu menangkap fitur kompleks. Sebaliknya, model dengan dua atau tiga lapisan dapat menangkap pola lebih dalam, namun membutuhkan lebih banyak sumber daya komputasi dan rentan terhadap *overfitting*. Oleh karena itu, penting untuk mengevaluasi bagaimana jumlah lapisan dalam arsitektur *LSTM* mempengaruhi akurasi prediksi pada tugas analisis sentimen. Hasil penelitian diharapkan memberikan wawasan tentang konfigurasi arsitektur *LSTM* yang optimal untuk data teks yang kompleks dan tidak terstruktur, serta berkontribusi pada pengembangan teknologi *NLP* yang lebih baik untuk aplikasi berbasis komunitas digital.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah seperti di 1.1, maka rumusan masalah dari penelitian ini, yaitu :

1. Bagaimana pengaruh jumlah lapisan dalam arsitektur *Long Short-Term Memory* (LSTM) (satu, dua, dan tiga lapisan) terhadap akurasi prediksi dalam analisis sentimen pada data komentar video YouTube?

## 1.3 Batasan Masalah

Batasan masalah penelitian ini, yaitu :

1. Dataset yang digunakan terbatas pada komentar pengguna dari video "*Clash of Champions*" di kanal YouTube Ruangguru. Komentar-komentar ini dipilih sebagai sumber data untuk menganalisis opini publik terkait program tersebut.
2. Penelitian ini berfokus pada data teks berbahasa Indonesia yang informal dan tidak baku, sesuai dengan karakteristik komentar YouTube yang diambil. Penggunaan bahasa selain bahasa Indonesia tidak menjadi ruang lingkup penelitian ini.
3. *Fine-tuning* yang dilakukan terbatas pada model *Large Language Model (LLM)* Gemma2 dengan parameter sebesar 9 miliar. Penelitian ini tidak akan membandingkan hasil *fine-tuning* dengan model-model lainnya atau dengan LLM yang lebih kecil atau lebih besar.

## 1.4 Tujuan Penelitian

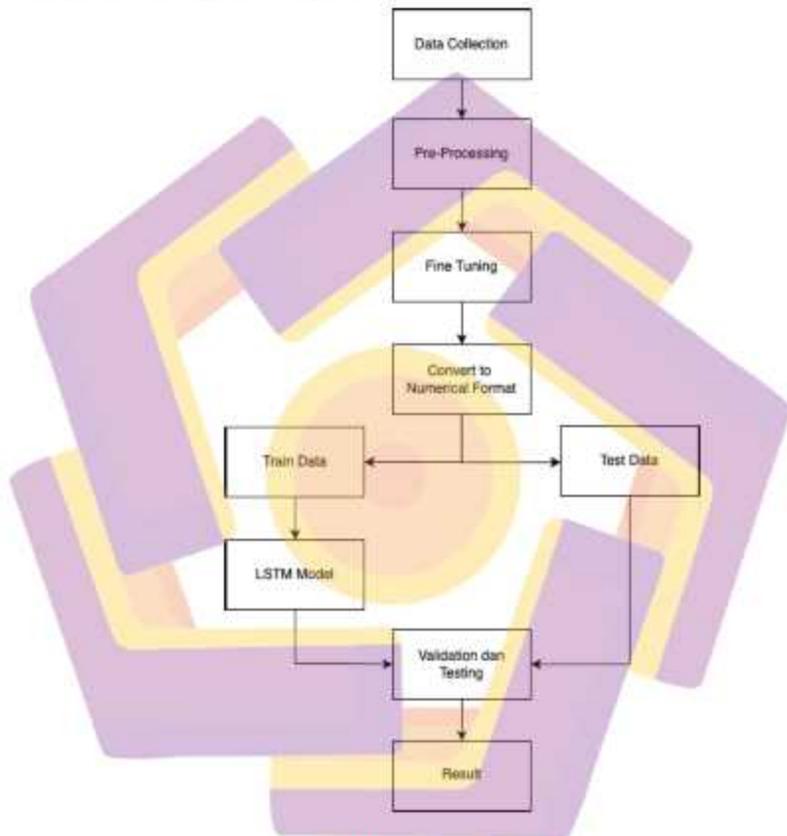
Tujuan penelitian ini, yaitu :

1. Menganalisis pengaruh jumlah lapisan dalam arsitektur *Long Short-Term Memory* (LSTM), yaitu satu, dua, dan tiga lapisan, terhadap akurasi prediksi sentimen pada data komentar pengguna dari video "*Clash of Champions*" di YouTube.

## 1.5 Metodologi Penelitian

Pada penulisan skripsi ini, alur penelitian bertujuan untuk menjelaskan tahapan atau rancangan metodologis yang digunakan dalam menganalisis sentimen terhadap video "*Clash of Champions*" di kanal YouTube Ruangguru dengan

menggunakan algoritma *Long Short-Term Memory (LSTM)*. Tahapan penelitian ini disusun secara sistematis untuk memastikan model yang dikembangkan dapat menangkap pola-pola sentimen secara efektif dan akurat. Gambar 1.1 berikut adalah alur penelitian yang akan dilakukan:



**Gambar 1.1 Alur Penelitian**

Pada penelitian ini, alur penelitian dimulai dengan pengumpulan data dari komentar video "Clash of Champions" di YouTube menggunakan *API YouTube*. Data tersebut kemudian melalui tahap *preprocessing* untuk pembersihan, seperti *tokenization*, *stop word removal*, dan *stemming* lalu melakukan *fine-tuning*. Setelah data dipersiapkan, dilakukan pembagian menjadi data latih dan data uji, diikuti oleh

konversi ke format numerik menggunakan *embedding TF-IDF*. Model *LSTM* dilatih dengan data latih dan diuji dengan data uji untuk mengukur performanya dalam klasifikasi sentimen. Akhirnya, hasil penelitian dievaluasi berdasarkan metrik akurasi untuk menarik kesimpulan mengenai kinerja model.

#### **1.6 Sistematika Penulisan**

Berikut sistematika penulisan agar pembaca lebih mudah memahami isi laporan penelitian :

##### **BAB I PENDAHULUAN**

Pada Bab 1 membahas tentang isi dan rencana dalam pengerjaan skripsi terkait latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan

##### **BAB II TINJAUAN PUSTAKA**

Pada Bab 2 membahas tentang kajian tinjauan pustaka yang mencakup studi literatur, table aslian dan dasar teori. Hal ini bersumber dari jurnal atau referensi yang berkaitan dengan penelitian ini.

##### **BAB III METODE PENELITIAN**

Pada Bab 3 membahas tentang metode penelitian yang digunakan mencakup objek penelitian, alur penelitian, serta alat dan bahan yang akan digunakan.

##### **BAB IV HASIL DAN PEMBAHASAN**

Pada Bab 4 membahas tentang implementasi hasil dan juga pembahasan pada perancangan penelitian.

##### **BAB V PENUTUP**

Pada Bab 5 membahas tentang kesimpulan dan saran yang didapatkan dari hasil penelitian.