

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sebagai media sosial, YouTube merupakan salah satu *platform* yang paling banyak digunakan di seluruh dunia. YouTube menyajikan berbagai video dengan durasi tertentu dan berbagai macam *genre*. YouTube juga menyediakan fitur kolom komentar yang ditujukan pada pengguna YouTube untuk memberikan komentar umpan balik (*feedback*) dalam bentuk *like* atau *dislike* terhadap video yang ditayangkan di *platform* tersebut. Sehingga pengguna dapat berinteraksi dan menyampaikan opini mereka tentang video tersebut. Komentar yang diberikan dapat bersifat positif, seperti berupa saran konstruktif atau motivasi yang mendukung kreator video. Namun, terdapat pula komentar yang bersifat negatif, terutama yang mengandung faktor SARA (Suku, Agama, Ras, dan Antargolongan) berpotensi menimbulkan konflik serta menjatuhkan reputasi konten kreator. video Youtube[1].

Platform YouTube memiliki popularitas yang sangat tinggi di Indonesia. Berdasarkan jumlah pengguna, Indonesia menduduki posisi keempat secara global sebagai negara dengan pengguna YouTube terbanyak, melampaui negara-negara seperti Meksiko dan Jepang. Pada bulan april 2024, jumlah pengguna YouTube di Indonesia mencapai 139 juta pengguna[2]. YouTube juga digunakan sebagai sarana untuk menyebarkan konten pendidikan yang bermanfaat oleh berbagai channel, termasuk channel YouTube Ruangguru. Ruangguru adalah perusahaan teknologi pendidikan terkemuka di Indonesia yang menawarkan berbagai layanan pembelajaran berbasis teknologi. Platform ini dirancang untuk menyediakan berbagai layanan pendidikan, mencakup kelas virtual, fasilitas ujian daring, video pembelajaran berbasis langganan, marketplace untuk layanan les privat, serta berbagai konten edukasi lain yang dapat diakses melalui aplikasi dan situs website resmi Ruangguru[3].

Ruangguru memiliki channel YouTube dengan jumlah pengikut sebanyak 1,91 juta *subscriber* per tanggal 12 Agustus 2024. Channel ini aktif mengunggah berbagai jenis konten edukasi pendidikan, salah satu video terbaru yang sukses menarik perhatian dan minat besar dari masyarakat Indonesia adalah serial yang berjudul “Clash of Champions”, sebuah serial video yang berhasil menjadi salah satu *trending* di YouTube dengan jumlah jutaan *views*. Serial tersebut pada episode 1 telah ditonton sebanyak 10 juta kali dan memperoleh lebih dari 17 ribu komentar. Dengan tingginya popularitas dan jumlah penonton yang mencapai jutaan pada serial tersebut, berbagai tanggapan, baik positif maupun negatif, bermunculan dari pengguna YouTube, khususnya masyarakat Indonesia. Oleh karena itu, diperlukan analisis sentimen

terhadap komentar pada channel YouTube Ruangguru untuk memahami tanggapan masyarakat Indonesia serta mendapatkan *feedback* mengenai kualitas dan relevansi konten video yang disajikan oleh Ruangguru.

Analisis sentimen bertujuan untuk mengembangkan suatu *system* yang bisa mengidentifikasi serta menginterpretasikan pendapat / pandangan seseorang yang terkandung dalam suatu kalimat atau kata didalam dokumen atau sebuah tulisan . proses ini bisa menganalisa kalimat serta mengelompokkannya secara otomatis kedalam kategori positif, netral, dan negatif [4]. Namun Komentar pada video YouTube tidak selalu relevan dengan topik utama video. Banyak komentar yang membahas topik lain seperti pengalaman pribadi, promosi, atau opini yang tidak berhubungan langsung dengan konten video. Untuk mengakomodasi komentar-komentar semacam ini, diperlukan kelas tambahan, yaitu "*Others*," atau kalimat diluar pembahasan topik yang akan membantu meningkatkan akurasi analisis sentimen dengan menghindari *missclassification*. Tujuan analisis sentimen adalah untuk memahami sudut pandang secara menyeluruh terhadap produk atau layanan yang sedang dikaji / *Point of View* (POV), sehingga dapat memberikan wawasan yang lebih mendalam mengenai pengalaman yang dirasakan oleh pengguna.[5]. Pada penelitian ini metode yang digunakan untuk melakukan analisis sentimen komentar YouTube pada video Clash of Champions yaitu menggunakan Algoritma BERT. BERT adalah model pembelajaran mesin yang dirancang dan dikembangkan untuk membangun representasi dua arah dari teks tanpa label. Proses ini dilakukan dengan mempertimbangkan konteks di sisi kanan dan kiri secara bersamaan pada setiap lapisan pemrosesan [6].

Algoritma BERT memiliki keunggulan dalam menangani teks panjang secara lebih efisien dibandingkan metode RNN, karena memanfaatkan *positional encoding* untuk memahami urutan kata dalam *input*. Selain itu, BERT dibangun di atas *arsitektur Transformer* yang memungkinkan pemrosesan data secara paralel. Dengan jumlah parameter yang lebih efisien dibandingkan model berbasis CNN, BERT mampu menghasilkan performa yang lebih optimal dalam waktu pemrosesan yang relatif lebih cepat[7] .

Analisis sentimen berperan penting dalam memahami tanggapan masyarakat terhadap konten edukasi yang disajikan secara daring. Seiring dengan meningkatnya jumlah pengguna yang menyampaikan pendapat melalui kolom komentar, perlu dilakukan analisis untuk memperoleh pemahaman yang lebih mendalam mengenai persepsi mereka. Hal ini menjadi semakin relevan pada konten edukasi yang disajikan oleh Ruangguru, yang memiliki dampak signifikan bagi peserta didik dan tenaga pengajar. Melalui analisis sentimen, Ruangguru dapat

menerima masukan yang lebih akurat terkait relevansi dan kualitas kontennya, yang nantinya dapat dijadikan dasar untuk melakukan perbaikan di masa mendatang.

Namun, terdapat tantangan yang cukup besar dalam menganalisis komentar-komentar tersebut, terutama karena banyak di antaranya menggunakan bahasa informal, slang, emotikon, serta bentuk lain yang tidak terstruktur. Oleh karena itu, penggunaan model BERT dianggap tepat karena kemampuannya dalam memahami konteks bahasa secara mendalam serta fleksibel dalam menghadapi variasi bentuk bahasa informal yang sering digunakan di media sosial seperti YouTube.

## 1.2. Rumusan Masalah

Dengan mempertimbangkan permasalahan yang dijelaskan di latar belakang, dapat disimpulkan bahwa rumusan masalah adalah sebagai berikut :

- a. Bagaimana performa model BERT dapat diimplementasikan untuk analisis sentimen pada komentar Youtube pada video Clash of Champions di channel Ruang Guru ?
- b. Seberapa efektif model BERT dalam membedakan sentimen positif, negatif atau dalam komentar Youtube pada video Clash of Champions di channel Ruang Guru berdasarkan metrik evaluasi *accuracy, precision, recall, dan F1-score* ?

## 1.3. Tujuan Penelitian

Fokus utama dalam penelitian yaitu mengimplementasikan model BERT (*Bidirectional Encoder Representations from Transformers*) serta mengetahui performanya dalam melakukan analisis sentimen pada komentar YouTube terkait video “Clash of Champions” di channel Ruanguru. Melalui penerapan model BERT, penelitian ini bertujuan untuk mengklasifikasikan sentimen positif, netral, dan negatif dan *others* dengan lebih akurat dibandingkan metode analisis sentimen konvensional..

## 1.4. Batasan Masalah

Di dalam studi penelitian ini, telah ditentukan sejumlah batasan masalah untuk memastikan bahwa penelitian yang dilakukan tetap terarah pada inti pembahasan yang diangkat. berikut batasan masalah dari studi ini :

- a. Metodologi yang akan di implementasikan adalah Algoritma BERT (*Bidirectional Encoder Representations from Transformers*)
- b. Dataset menggunakan format csv.

- c. Data komentar diambil hanya dari Episode 1 pada video “Clash of Champions” di channel YouTube Ruangguru dengan total 17.100 komentar, yang dianggap representatif karena jumlah komentar tertinggi dibandingkan dengan episode lainnya.
- d. Komentar yang tidak relevan dengan topik akan dimasukkan ke dalam kelas “*Others*” untuk menghindari pengaruh noise pada model.
- e. Proses analisis dilakukan menggunakan Google Colaboratory dengan Bahasa Pemrograman Python.

