

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, media sosial telah menjadi sarana utama bagi masyarakat untuk berinteraksi, bertukar pendapat, dan menyampaikan informasi secara terbuka. Salah satu *platform* yang paling populer dan banyak digunakan di Indonesia adalah *YouTube*. *Platform* ini memungkinkan pengguna untuk memberikan komentar, berbagi opini, serta berinteraksi langsung dengan pembuat konten maupun sesama pengguna. Dengan jutaan komentar yang diunggah setiap harinya, kolom komentar *YouTube* menjadi ruang publik yang padat dan dinamis (Qonita *et al.*, 2023).

Namun, terbukanya ruang interaksi ini juga memunculkan permasalahan serius, salah satunya adalah penyalahgunaan kolom komentar oleh akun bot. Bot adalah program otomatis yang dapat menghasilkan komentar secara massal, sering kali menyisipkan tautan spam atau konten manipulatif yang berpotensi mengganggu kenyamanan pengguna dan menurunkan kredibilitas *platform* (Ferrara *et al.*, 2016; Pasieka *et al.*, 2021). Bot juga berkontribusi dalam penyebaran disinformasi dan manipulasi opini publik, yang semakin sulit dikenali karena gaya bahasanya meniru manusia (Ng & Carley, 2025)

Keberadaan akun bot di *YouTube* secara langsung berdampak pada kualitas interaksi digital. Komentar bot dapat menutupi komentar organik dari akun asli, membuat konten berkualitas kalah saing, dan bahkan menyesatkan penonton. Hal ini tentu menjadi tantangan tersendiri bagi pengelola *platform* maupun para peneliti untuk menjaga ekosistem digital yang sehat.

Salah satu kendala utama dalam membangun sistem pendekripsi komentar bot adalah keterbatasan data berlabel. Komentar dari akun bot dan akun asli sulit dibedakan secara manual dalam jumlah besar, sehingga proses pelabelan data menjadi sangat terbatas. Tantangan ini membutuhkan pendekatan yang lebih fleksibel dan efisien, terutama dalam konteks data yang minim.

Untuk menjawab tantangan tersebut, penelitian ini mengusulkan penggunaan *Few-shot learning*, yaitu pendekatan dalam *Machine Learning* (ML) yang memungkinkan model belajar dari sejumlah kecil data berlabel. *Few-shot learning* telah terbukti efektif dalam skenario serupa, seperti deteksi rumor di media sosial (Lu et al., 2021) atau klasifikasi *tweet* bertema *COVID-19* dengan data terbatas (Lwowski & Najafirad, 2020)

Dalam penerapannya, penelitian ini menggunakan model *Transformers-based*, yang dikenal memiliki kemampuan tinggi dalam memahami konteks bahasa alami. Model ini banyak digunakan dalam tugas-tugas *Natural Language Processing* (NLP) seperti klasifikasi teks dan analisis sentimen karena kemampuannya mengolah hubungan antarkata dalam skala besar (Fadhil N, 2024; Gaber et al., 2024). Dengan melakukan *fine-tuning* pada data komentar terbatas, model dapat mengidentifikasi pola dari komentar akun asli dan akun bot secara lebih akurat (Liao et al., 2023; Logeswaran et al., 2020).

Berbeda dengan penelitian terdahulu yang banyak berfokus pada platform seperti Twitter atau Facebook (Rashif et al., 2021; Tunc et al., 2022), penelitian ini menyoroti kesenjangan riset pada *platform YouTube* dalam mendeteksi komentar bot. Selain itu, pendekatan yang digunakan juga lebih adaptif karena mampu memberikan hasil meskipun hanya dengan sedikit contoh data.

Dengan mengombinasikan pendekatan *Few-shot learning* dan model *Transformers-based*, penelitian ini menawarkan solusi klasifikasi komentar untuk membedakan komentar dari akun asli dan akun bot, yang dapat langsung diterapkan pada skenario nyata di *YouTube*. Harapannya, sistem ini mampu meningkatkan kualitas interaksi digital, mengurangi spam, serta mendukung pengembangan lebih lanjut di bidang kecerdasan buatan, khususnya dalam pemrosesan bahasa alami.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah :

1. Bagaimana mengimplementasikan teknik *Few-shot learning* dalam klasifikasi teks komentar akun asli dan akun bot di *platform YouTube*?

2. Bagaimana performa model *Transformers-based* dalam membedakan komentar dari akun asli dan akun bot di *platform YouTube*?

1.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penelitian ini adalah :

1. Mengembangkan model klasifikasi yang mampu membedakan komentar dari akun asli dan akun bot di *YouTube* menggunakan pendekatan *Few-shot learning*.
2. Mengukur dan menganalisis performa model *Transformers-based* dalam tugas klasifikasi teks komentar untuk kategori akun asli dan akun bot dengan menggunakan metrik evaluasi seperti *Confusion Matrix*.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Penelitian ini diharapkan memberikan kontribusi bagi pengembangan teknologi pemrosesan bahasa alami, khususnya dalam penerapan *Few-shot learning* untuk klasifikasi teks dengan data yang terbatas.
2. *Youtube* dan *platform* serupa dapat mengaplikasikan model ini untuk menyaring komentar yang tidak autentik (komentar dari akun bot), sehingga meningkatkan kepercayaan pengguna dan menjaga kualitas interaksi di *platform*.
3. Adanya klasifikasi teks yang lebih akurat, pengguna *YouTube* dapat menikmati pengalaman yang lebih bersih dari spam dan komentar otomatis, sehingga dapat mengakses konten yang lebih relevan dan kredibel.
4. Penelitian ini dapat menjadi dasar bagi pengembangan dan penerapan model *Transformers* pada skenario klasifikasi komentar di *platform* dengan data terbatas pada masa mendatang

1.5 Ruang Lingkup dan Batasan Penelitian

Adapun Ruang Lingkup dan Batasan Penelitian ini agar tujuan tercapai maka:

1. Penelitian ini hanya difokuskan pada klasifikasi teks komentar dari satu video akun *YouTube* dan tidak mencakup *platform* media sosial lainnya yang mungkin memiliki karakteristik dan jenis komentar yang sarupa.

2. Data yang digunakan terbatas pada data publik yang tersedia setelah *scraping* atau *crawling* dan sesuai dengan kebijakan privasi *YouTube*, sehingga tidak termasuk data komentar yang dihapus atau dibatasi oleh *platform*.
3. Model akan mengklasifikasikan teks komentar berdasarkan dua kategori yaitu; akun asli dan akun bot
4. Hanya menggunakan komentar dalam bahasa indonesia, tanpa menerjemahkan atau menganalisis komentar berbahasa asing.
5. Pada penelitian ini model yang digunakan terbatas pada model *Transformers-based* tanpa membandingkan performanya dengan model lain di luar.
6. Pendekatan *Few-shot learning* yang diterapkan menggunakan teknik *fine-tuning* pada model *Transformers-based*, dengan asumsi bahwa data terlabel yang tersedia untuk tiap jenis akun sangat terbatas.
7. Hasil penelitian ini berupa teks komentar dari akun asli dan akun bot
8. Kinerja model akan dievaluasi dengan memakai metrik seperti : Akurasi, *Presisi*, *Recall*, dan *F1-score*.