

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan pesat konten video berdurasi pendek di media sosial telah mengubah lanskap komunikasi publik dan cara masyarakat mengonsumsi informasi. Platform seperti Instagram Reels, TikTok, dan YouTube Shorts mampu menarik perhatian pengguna melalui penyajian informasi yang ringkas, cepat, dan mudah diakses. Laporan DataReportal (2025) mencatat bahwa Indonesia memiliki sekitar 143 juta pengguna media sosial aktif, dengan lebih dari 90% menonton konten video setiap minggu. Secara global, *Cisco Visual Networking Index* melaporkan bahwa sejak 2022, video menyumbang lebih dari 82% dari total lalu lintas internet, menegaskan bahwa format video telah menjadi bentuk konten dominan di dunia digital. Kondisi ini menimbulkan urgensi untuk membangun sistem analisis yang mampu memahami isi dan karakteristik video dalam skala besar, terutama karena volume unggahan yang terus meningkat dan heterogenitas bahasa yang digunakan, termasuk *slang*, campuran bahasa daerah, hingga variasi ekspresi yang tidak baku. Analisis manual terhadap konten video dengan karakteristik tersebut tidak hanya memakan waktu, tetapi juga tidak memungkinkan dilakukan secara konsisten dan sistematis dalam konteks volume data yang sangat besar.

Meningkatnya dominasi konten video mendorong kebutuhan akan metode analisis yang mampu memahami isi, pola penyampaian, dan respons audiens secara mendalam. Analisis video tidak hanya berfungsi untuk mengukur performa konten, tetapi juga untuk mengidentifikasi elemen penting seperti struktur naratif, gaya

bahasa, serta karakteristik penyampaian yang memengaruhi keterlibatan pengguna. Penelitian Munaro *et al.* (2024) menunjukkan bahwa setiap gaya linguistik konten kreator memiliki keterkaitan dengan tingkat keterlibatan konsumen digital yang berbeda-beda. Schwenzow *et al.* (2021) mencatat bahwa analisis manual konten video sangat sulit diskalakan karena memakan waktu dan bergantung pada penilaian subjektif. Sebagai solusi atas keterbatasan tersebut, pendekatan otomatis memungkinkan penanganan dataset berskala besar secara lebih konsisten dan objektif. Temuan-temuan ini memperlihatkan bahwa analisis konten video merupakan fondasi penting dalam memahami dinamika komunikasi digital yang semakin berpusat pada format audio-visual.

Kemajuan teknologi *Natural Language Processing* (NLP), yaitu bidang kecerdasan buatan yang berfokus pada pemrosesan, pemahaman, dan generasi bahasa manusia oleh komputer, khususnya melalui pengembangan *Large Language Models* (LLM), telah menghadirkan pendekatan baru yang mampu memahami bahasa alami secara lebih mendalam. Model-model seperti Qwen2.5, GPT-4, dan LLaMA menunjukkan kemampuan yang kuat dalam tugas-tugas peringkasan, klasifikasi, dan penalaran semantik. Model dengan skala parameter sangat besar yang dilatih melalui *fine-tuning* penuh membutuhkan sumber daya komputasi yang sangat besar, sehingga menimbulkan tantangan signifikan bagi penelitian akademik dan pengembangan industri dengan sumber daya terbatas. Pendekatan *Parameter-Efficient Fine-Tuning* (PEFT), seperti *Low-Rank Adaptation* (LoRA), dikembangkan sebagai respons terhadap kendala tersebut untuk mengadaptasi model berskala besar secara lebih efisien melalui pembaruan parameter berperingkat rendah. LoRA bekerja dengan membekukan parameter utama model dan mengoptimalkan matriks

adapter berukuran kecil, sehingga jumlah parameter yang dilatih berkurang secara signifikan. Hu *et al.* (2021) menunjukkan bahwa LoRA dapat mengurangi kompleksitas parameter hingga beberapa orde besaran tanpa menurunkan akurasi model. Databricks (2023) mencatat bahwa dalam beberapa kasus, LoRA mampu memberikan performa yang lebih baik dibandingkan *fine-tuning* penuh karena dapat mengurangi risiko *catastrophic forgetting*, yaitu kondisi ketika model kehilangan pengetahuan penting yang diperoleh selama pelatihan awal.

Penelitian terkini memperkenalkan QLoRA sebagai pendekatan *fine-tuning* yang mengombinasikan kuantisasi 4-bit dengan *adapter* LoRA, sehingga memungkinkan pelatihan model bahasa berskala sangat besar pada perangkat *Graphics Processing Unit* (GPU) dengan kapasitas memori terbatas (Dettmers et al., 2023). Dettmers et al. (2023) menunjukkan bahwa QLoRA mampu mencapai performa yang sebanding dengan *fine-tuning* 16-bit penuh, namun dengan konsumsi memori yang jauh lebih rendah, sehingga menjadikannya solusi yang efisien dan praktis untuk adaptasi *Large Language Models* pada keterbatasan sumber daya komputasi.

Penelitian ini memanfaatkan model Qwen2.5-7B-Instruct yang di-*fine-tune* menggunakan QLoRA untuk dua tugas utama yang berkaitan dengan analisis konten video pendek Instagram Reels, yaitu peringkasan isi video berbasis transkrip dan klasifikasi konten, termasuk identifikasi kategori serta deteksi iklan. Pengumpulan data dilakukan secara otomatis dari konten Instagram Reels, kemudian diperluas menggunakan pendekatan *Generate–Annotate–Learn* (GAL) yang memungkinkan pembuatan dataset anotasi dalam skala besar tanpa keterlibatan *annotator* (pemberi

label) manusia secara intensif. Penerapan metode ini memiliki urgensi karena anotasi (pemberian label) manual untuk data video berskala besar memerlukan waktu dan biaya yang tinggi serta rentan terhadap inkonsistensi. Model hasil *fine-tuning* dievaluasi menggunakan metrik standar NLP, yaitu *Recall-Oriented Understudy for Gisting Evaluation (ROUGE)* untuk menilai kualitas ringkasan teks, serta akurasi dan *F1-score* untuk menilai performa klasifikasi konten dan deteksi iklan. Pendekatan yang diusulkan dalam penelitian ini diharapkan dapat memberikan kontribusi terhadap pemanfaatan LLM yang lebih efisien dan terjangkau dalam analisis konten video pendek, serta menawarkan kerangka metodologis yang dapat diterapkan dalam pengembangan sistem otomatis di bidang pemasaran digital, moderasi konten, dan analisis media sosial secara umum.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka terdapat beberapa permasalahan yang akan dibahas dalam penelitian ini:

1. Bagaimana kinerja model Qwen2.5-7B-Instruct yang di-*fine-tune* menggunakan QLoRA dalam tugas peringkasan isi video dan metadata Instagram Reels dalam Bahasa Indonesia?
2. Bagaimana kinerja model tersebut dalam klasifikasi konten video (termasuk deteksi iklan) berdasarkan teks transkrip?
3. Bagaimana hasil evaluasi model (skor *ROUGE* dan metrik klasifikasi) dapat membantu memahami kemampuan model dalam analisis konten video?

1.3 Batasan Masalah

Penelitian ini dibatasi pada analisis konten video pendek yang berasal dari 655 unggahan Instagram Reels, yang diperoleh melalui proses pengambilan data secara acak dari berbagai akun publik dengan beragam kategori. Penelitian ini secara khusus memanfaatkan model Qwen2.5-7B-Instruct, dengan proses *fine-tuning* menggunakan metode QLoRA, yang dipilih karena efisiensinya dalam melakukan adaptasi model besar pada lingkungan komputasi yang terbatas.

Ruang lingkup penelitian dibatasi pada dua tugas utama, yaitu peringkasan isi video dan klasifikasi konten, termasuk pengkategorian video serta deteksi iklan, yang seluruhnya dilakukan berdasarkan data transkrip teks dan metrik performa (seperti jumlah *likes*, *views*, durasi, dan komentar) yang diekstraksi dari video Instagram Reels. Penelitian tidak mencakup analisis multimodal berbasis audio atau visual secara langsung, sehingga fokus penelitian berada pada pemrosesan teks sebagai representasi utama konten video.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Membangun dan melatih model Qwen2.5-7B-Instruct menggunakan QLoRA untuk tugas peringkasan isi video dan klasifikasi konten Instagram Reels.
2. Mengevaluasi kinerja model pada metrik *ROUGE* (untuk ringkasan) serta nilai akurasi dan *F1-score* (untuk klasifikasi konten dan deteksi iklan).
3. Menganalisis hasil evaluasi untuk memahami kelebihan dan keterbatasan model dalam konteks aplikasi analisis video pendek berbahasa Indonesia.