

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, platform media sosial dan berbagi video seperti YouTube telah menjadi sarana utama bagi individu dan komunitas untuk berinteraksi, berbagi informasi, dan berekspresi. Sebagai platform berbagi video terbesar di dunia dengan lebih dari 1,5 miliar pengguna aktif bulanan, YouTube menarik miliaran komentar yang mencerminkan berbagai pandangan dan opini pengguna (Yang dkk., 2022). Namun, kebebasan berekspresi ini seringkali disalahgunakan, sehingga memunculkan masalah serius seperti penyebaran konten negatif, termasuk komentar yang bersifat toksik (yaitu, komentar yang kasar, tidak sopan, atau tidak masuk akal yang dapat membuat seseorang enggan berdiskusi atau berinteraksi) dan *spam* (Dehkhoda & Gunica, 2023), khususnya terkait dengan promosi aktivitas ilegal seperti judi *online* (Judol), yang seringkali dipromosikan melalui platform seperti YouTube (Wulandari & Ramadhany, 2023). Dari perspektif penelitian ini, penyalahgunaan tersebut bukan hanya masalah teknis semata, melainkan juga masalah sosial yang mengancam integritas ruang diskusi daring dan memerlukan pendekatan solutif yang komprehensif.

Komentar negatif ini merusak reputasi kreator dan mengganggu pengguna, sementara volume yang masif membuat moderasi manual tidak efektif (Dehkhoda & Gunica, 2023; Manaroinson dkk., 2024). Oleh karena itu, dibutuhkan sistem otomatis yang mampu mengelola komentar secara efisien.

YouTube telah menyediakan fitur moderasi otomatis yang menyaring komentar berpotensi tidak pantas untuk ditinjau lebih lanjut oleh kreator (YouTube Help Center, 2025). Namun, efektivitas filter otomatis ini terbatas, karena penelitian telah menunjukkan bahwa sistem semacam ini dapat dilewati melalui manipulasi teks sederhana seperti salah ketik atau penambahan spasi (Hosseini dkk., 2017). Kerentanan inilah yang dieksploitasi oleh penyebar spam di Indonesia, di mana banyak promosi judi online (judol) berhasil lolos karena penggunaan kosakata unik dan ejaan yang sengaja disamarkan. Maka dari itu, penelitian ini bertujuan untuk menutup celah tersebut dengan mengembangkan sistem penyaring tambahan yang dirancang khusus untuk mengenali pola spam khas Indonesia, guna melengkapi dan memperkuat fitur moderasi yang sudah ada.

Penelitian sebelumnya telah banyak membahas deteksi komentar negatif menggunakan berbagai metode, termasuk pendekatan berbasis aturan, *machine learning*, hingga *deep learning* yang memanfaatkan *Natural Language Processing* (NLP) untuk menganalisis teks (Androćec, 2020). Analisis literatur menunjukkan adanya tantangan dalam mendeteksi *spam* dan toksisitas baru berbahasa dan kebudayaan Indonesia, seperti variasi istilah "Judol". Selain itu, integrasi antara sistem deteksi dengan tindakan moderasi masih perlu ditingkatkan untuk memberikan solusi yang praktis bagi pengguna YouTube (Gillespie, 2020).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan aplikasi web untuk membantu kreator YouTube mengelola komentar. Aplikasi akan fokus pada deteksi dan penghapusan otomatis komentar "Judol/Spam" dan "Toxic" menggunakan analisis teks dan Perspective API, dengan harapan dapat menciptakan ekosistem YouTube yang lebih bersih dan positif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana membangun sistem deteksi komentar otomatis yang efektif untuk mengidentifikasi komentar "Judol/Spam" pada video YouTube?
2. Bagaimana memanfaatkan Perspective API untuk mendeteksi tingkat toksisitas pada komentar-komentar video YouTube?
3. Bagaimana merancang dan mengimplementasikan antarmuka aplikasi *web* yang memungkinkan pengguna melihat, memfilter, dan menghapus komentar yang terdeteksi sebagai "Judol/Spam" atau "Toxic" dari video YouTube mereka?
4. Bagaimana mengintegrasikan fungsionalitas penghapusan komentar melalui YouTube Data API v3 dalam aplikasi *web* untuk memfasilitasi moderasi otomatis?

1.3 Ruang Lingkup

Ruang lingkup penelitian ini dibatasi pada hal-hal berikut:

1. Sistem akan berfokus pada deteksi komentar pada video *YouTube* yang dimiliki atau dikelola oleh pengguna yang telah *login* dengan akun *Google/YouTube* mereka.
2. Deteksi "Judol/Spam" akan dilakukan menggunakan pendekatan berbasis pola kata kunci dan frasa yang relevan dengan promosi judi *online* atau *spam* generik.
3. Deteksi "Toxic" akan memanfaatkan *Perspective API* dari Google untuk menganalisis tingkat *toksisitas* komentar. Atribut lain dari *Perspective API* seperti *Insult* dan *Profanity* akan digunakan sebagai penanda tambahan untuk kategori "Ditandai" jika ambang batas *toksisitas* utama tidak tercapai.
4. Penelitian ini tidak mencakup pengembangan model *machine learning* atau *deep learning* kustom untuk deteksi *toksisitas*, melainkan bergantung pada *Perspective API* sebagai layanan eksternal.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun sistem deteksi komentar otomatis yang mampu mengidentifikasi komentar "Judol/Spam" pada video *YouTube*.
2. Mengimplementasikan integrasi dengan *Perspective API* untuk melakukan deteksi tingkat *toksisitas* pada komentar video *YouTube*.
3. Mengembangkan aplikasi *web* dengan antarmuka yang intuitif untuk memungkinkan pengguna melihat, memfilter, dan melakukan penghapusan komentar yang terdeteksi sebagai "Judol/Spam" atau "Toxic" dari video

YouTube mereka.

4. Mewujudkan fungsionalitas penghapusan komentar secara efektif melalui penggunaan *YouTube Data API v3* untuk mendukung moderasi konten.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Pembuat Konten *YouTube*: Menyediakan alat bantu yang efisien untuk moderasi komentar, mengurangi beban kerja manual dalam membersihkan kolom komentar dari konten negatif, dan membantu menjaga reputasi saluran.
2. Bagi Pengguna *YouTube*: Meningkatkan pengalaman berinteraksi pada platform dengan menciptakan lingkungan komentar yang lebih bersih, positif, dan bebas dari *spam* serta ujaran kebencian.
3. Bagi Pengembangan IPTEK: Menambah referensi dan studi kasus mengenai implementasi API pihak ketiga (*YouTube Data API* dan *Perspective API*) dalam pembangunan aplikasi *web* untuk moderasi konten.
4. Bagi Peneliti Selanjutnya: Memberikan dasar dan referensi untuk pengembangan lebih lanjut dalam bidang deteksi komentar negatif, khususnya dengan fokus pada konteks lokal dan integrasi fitur moderasi.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menguraikan latar belakang masalah yang melandasi penelitian, perumusan masalah yang akan dijawab, ruang lingkup yang membatasi area studi, tujuan yang ingin dicapai, manfaat yang diharapkan dari penelitian, serta sistematika penulisan laporan secara keseluruhan.

BAB II TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini berisi kajian literatur dari penelitian-penelitian sebelumnya yang relevan serta teori-teori dasar yang mendukung konsep dan metodologi yang digunakan dalam penelitian. Ini mencakup tinjauan tentang deteksi komentar, *YouTube Data API*, *Perspective API*, serta konsep-konsep terkait pengembangan *web*.

BAB III METODE PENELITIAN

Bab ini menjelaskan secara rinci pendekatan, bahan, peralatan, serta prosedur pengumpulan data dan perancangan sistem yang digunakan.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dari perancangan sistem yang telah dilakukan, termasuk potongan kode atau konfigurasi penting. Selanjutnya, bab ini akan membahas hasil uji coba sistem, analisis temuan, dan interpretasi data yang diperoleh, dikaitkan dengan tinjauan pustaka dan dasar teori.

BAB V PENUTUP

Bab ini berisi simpulan dari keseluruhan hasil penelitian dan memberikan saran untuk pengembangan di masa depan.