

BAB 1

PENDAHULUAN

1.1. Latar Belakang

YouTube merupakan salah satu platform media sosial terbesar di dunia yang memungkinkan pengguna untuk berbagi video, memberikan komentar, serta berinteraksi dengan pengguna lain [1]. Berdasarkan survei yang dilakukan oleh Statista pada April 2024, YouTube menempati peringkat kedua sebagai media sosial paling populer, dengan jumlah pengguna aktif bulanan mencapai 2,504 miliar [2]. Dengan miliaran pengguna aktif setiap bulannya, YouTube menjadi ruang diskusi yang dinamis dan terbuka bagi berbagai macam konten serta komunitas [3], [4].

Seiring dengan perkembangan platform ini, muncul berbagai tantangan terkait moderasi konten, terutama dalam menangani komentar-komentar yang bersifat spam atau mengandung unsur ilegal, seperti promosi judi online [5]. Dalam beberapa bulan terakhir, komentar yang mempromosikan judi online semakin sering ditemukan di berbagai video YouTube. Direktur *Communication and Information System Security Research Center (CISSReC)*, Pratama Persadha, menyatakan bahwa kolom komentar YouTube menjadi sasaran empuk bagi promosi judi online karena tingginya tingkat keterlibatan pengguna di platform tersebut. Para pelaku memanfaatkan *bot* untuk menyebarkan tautan atau ajakan dengan janji keuntungan besar secara otomatis dalam jumlah besar. Hal ini diperparah dengan algoritma YouTube yang mempromosikan video berdasarkan interaksi, sehingga komentar *spam* ini semakin sulit dikendalikan [6].

Fenomena ini menjadi perhatian serius karena judi online termasuk dalam kategori aktivitas ilegal di banyak negara, termasuk Indonesia [7]. Komentar yang mengandung promosi judi online di YouTube berpotensi menarik pengguna untuk mencoba layanan tersebut, yang pada akhirnya dapat meningkatkan jumlah pemain. Berdasarkan data yang diungkap Pusat Pelaporan Analisis dan Transaksi Keuangan (PPATK) pada 26 Juli 2024, Indonesia menjadi negara dengan jumlah pemain judi online tertinggi, dengan total 4 juta pemain. Dari jumlah tersebut, 2% atau sekitar 80.000 pemain berusia di bawah 10 tahun, sementara 11% atau 440.000 pemain berusia 10-20 tahun. Mayoritas pemain judi online berasal dari kelompok usia 30-

50 tahun (40% atau 1,64 juta orang) dan di atas 50 tahun (34% atau 1,35 juta orang). Selain itu, PPATK mencatat ada 168 juta transaksi judi online dengan total perputaran dana mencapai Rp 327 triliun sepanjang tahun 2023. Secara keseluruhan, sejak tahun 2017, akumulasi perputaran dana transaksi judi online di Indonesia mencapai Rp 517 triliun [8].

Tingginya angka ini mengindikasikan bahwa promosi judi online yang beredar di media sosial, termasuk YouTube, dapat berdampak signifikan terhadap peningkatan jumlah pemain, terutama di kalangan anak-anak dan remaja. Deputi Bidang Koordinasi Peningkatan Kualitas Anak, Perempuan, dan Pemuda Kemenko PMK, Woro Srihastuti Sulistyaningrum, menekankan bahwa anak yang terjerumus dalam judi online cenderung melakukan tindakan kriminalitas karena belum siap secara ekonomi, psikososial, dan mental [8]. Oleh karena itu, pengawasan terhadap media sosial, termasuk efektivitas sistem moderasi di YouTube, menjadi faktor penting dalam membatasi penyebaran konten promosi judi online yang dapat menjangkau kelompok rentan.

Meskipun YouTube telah menerapkan sistem moderasi untuk mendeteksi dan memfilter *spam*, sistem ini belum secara spesifik dirancang untuk mengidentifikasi komentar promosi judi online. Salah satu kendala utama adalah keterbatasan sistem moderasi otomatis yang lebih banyak mengandalkan deteksi kata kunci. Para pelaku *spam* dengan mudah menghindari deteksi ini dengan menggunakan variasi ejaan, simbol, atau frasa yang sedikit dimodifikasi. Selain itu, moderasi manual tidak selalu memungkinkan bagi kanal dengan jumlah komentar yang sangat banyak, sehingga banyak komentar promosi judi online tetap bertahan dalam waktu lama sebelum terhapus.

Untuk mengatasi keterbatasan ini, diperlukan pendekatan yang lebih canggih untuk mengatasi keterbatasan sistem moderasi yang ada. Salah satu solusi yang menjanjikan adalah penggunaan teknologi kecerdasan buatan (AI) dan pemrosesan bahasa alami (NLP), khususnya model bahasa seperti IndoBERT yang dirancang untuk bahasa Indonesia [9]. Dengan mengadaptasi model ini, sistem moderasi dapat lebih efektif dalam mengenali pola dalam komentar promosi judi online, termasuk variasi ejaan dan frasa yang digunakan untuk menghindari deteksi otomatis.

Sejumlah penelitian telah menunjukkan efektivitas IndoBERT dalam tugas klasifikasi teks berbahasa Indonesia. Penelitian yang dilakukan oleh Nurjoko dan Agus Rahardi menunjukkan bahwa IndoBERT memiliki akurasi 72% dalam mengklasifikasikan sentimen kekerasan verbal di Twitter, yang membuktikan kemampuannya dalam mengenali berbagai bentuk komunikasi, termasuk variasi frasa atau ejaan [9]. Sementara itu, penelitian oleh Muhammad Basil Musyaffa Amin et al. menunjukkan bahwa IndoBERT dapat mendeteksi *spam* dalam teks berbahasa Indonesia dengan akurasi 98%, membuktikan efektivitasnya dalam memilah pesan otomatis [10]. Dengan melakukan *fine-tuning*, IndoBERT dapat dimanfaatkan untuk meningkatkan akurasi dalam mendeteksi dan mengklasifikasikan komentar promosi judi online di YouTube. Hasil klasifikasi ini kemudian dapat digunakan sebagai dasar bagi sistem untuk mengambil tindakan moderasi yang sesuai.

Berdasarkan permasalahan dan solusi yang telah diuraikan, penelitian ini bertujuan untuk merancang sistem informasi berbasis website yang memanfaatkan *fine-tuning* model IndoBERT dan integrasi YouTube Data API untuk klasifikasi dan moderasi komentar promosi judi online. Sistem ini diharapkan dapat membantu kreator konten dalam memoderasi komentar secara otomatis, sehingga mengurangi paparan konten ilegal kepada pengguna, terutama anak-anak dan remaja. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan teknologi NLP untuk bahasa Indonesia, serta menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan moderasi konten digital.

1.2. Rumusan Masalah

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

1. Bagaimana mengimplementasikan *fine-tuning* IndoBERT untuk mengklasifikasikan komentar promosi judi online pada video YouTube?
2. Bagaimana performa *fine-tuning* IndoBERT dalam mengklasifikasikan komentar promosi judi online berdasarkan metrik evaluasi seperti *confusion matrix*, akurasi, presisi, *recall*, *F1-score*, dan AUC-ROC?
3. Bagaimana merancang sistem informasi berbasis website yang memanfaatkan *fine-tuning* model IndoBERT dan integrasi YouTube Data

API untuk memberikan fungsionalitas moderasi yang optimal serta kontrol penuh bagi pengguna dalam mengelola komentar?

1.3. Tujuan Penelitian

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

1. Mengimplementasikan *fine-tuning* IndoBERT untuk mengklasifikasikan komentar promosi judi online pada video YouTube.
2. Mengukur performa model IndoBERT yang telah di-*fine-tuning* dalam mengklasifikasikan komentar promosi judi online berdasarkan metrik evaluasi seperti *confusion matrix*, akurasi, presisi, *recall*, *F1-score*, dan AUC-ROC.
3. Merancang sistem informasi berbasis website yang memanfaatkan *fine-tuning* model IndoBERT dan integrasi YouTube Data API untuk memberikan fungsionalitas moderasi komentar yang optimal serta memungkinkan pengguna memiliki kontrol penuh dalam mengelola komentar.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam berbagai aspek, baik teoritis maupun praktis, serta bermanfaat bagi berbagai pihak yang berkepentingan, di antaranya:

1. Manfaat Teoritis
 - a. Berkontribusi dalam pengembangan teknologi NLP untuk bahasa Indonesia, khususnya dalam konteks moderasi konten digital.
 - b. Menjadi referensi bagi penelitian selanjutnya terkait penggunaan model NLP seperti IndoBERT dalam klasifikasi dan moderasi konten ilegal di platform media sosial.
 - c. Memperluas wawasan akademik terkait pemanfaatan *fine-tuning* model NLP dalam pengelolaan komentar otomatis berbasis AI.

2. Manfaat Praktis

Bagi Kreator Konten:

- a. Membantu kreator konten dalam memoderasi komentar secara otomatis, sehingga mengurangi beban moderasi manual.

- b. Memberikan kontrol penuh kepada kreator konten dalam mengelola komentar yang masuk.
- c. Bagi Pengguna Youtube:
- d. Mengurangi paparan konten ilegal, khususnya promosi judi online, bagi pengguna, terutama anak-anak dan remaja.
- e. Meningkatkan keamanan dan kenyamanan dalam berinteraksi di platform YouTube.
- f. Bagi Pengembang Teknologi:
- g. Memberikan contoh implementasi praktis dalam membangun sistem moderasi otomatis berbasis AI dan YouTube Data API.
- h. Memperluas pemahaman tentang pemanfaatan fine-tuning model NLP dalam mendeteksi komentar berisiko secara otomatis.

1.5. Batasan Masalah

Untuk memastikan penelitian ini tetap fokus dan terarah, diperlukan batasan masalah yang menentukan ruang lingkup kajian. Batasan ini bertujuan untuk memperjelas aspek-aspek yang akan dibahas serta menghindari pembahasan yang terlalu luas atau di luar konteks penelitian. Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya membahas komentar yang mempromosikan judi online di platform YouTube. Komentar dengan tema lain (seperti *hate speech*, *spam* umum, atau promosi produk) tidak termasuk dalam cakupan penelitian.
2. Sistem ini dirancang khusus untuk mengklasifikasikan dan memoderasi komentar berbahasa Indonesia. Komentar dalam bahasa lain tidak termasuk dalam cakupan penelitian.
3. Sistem ini memanfaatkan YouTube Data API untuk mengambil dan memoderasi komentar. Penelitian tidak mencakup integrasi langsung dengan platform YouTube atau modifikasi pada sistem internal YouTube.
4. Sistem yang dirancang berbasis website dan berfungsi sebagai aplikasi pihak ketiga. Penelitian tidak mencakup perancangan aplikasi mobile atau desktop.

5. Penelitian ini menggunakan model IndoBERT yang di-*fine-tuning* untuk tugas klasifikasi. Model NLP lain atau pendekatan non-NLP tidak termasuk dalam cakupan penelitian.
6. Evaluasi performa sistem hanya dilakukan berdasarkan metrik evaluasi seperti *confusion matrix*, akurasi, presisi, *recall*, *F1-score*, dan AUC-ROC.
7. Evaluasi sistem dilakukan menggunakan pendekatan pengujian *black-box* untuk mengevaluasi fungsionalitas sistem, serta *System Usability Scale* (SUS) untuk mengukur tingkat kebergunaan sistem berdasarkan persepsi pengguna.