BAB I

PENDAHULUAN

1.1 *Latar Belakang*

Gaya hidup masyarakat modern tidak dapat dipisahkan dari kehadiran media sosial dan platform digital, yang menjadi sarana utama untuk berkomunikasi, berbagi informasi, serta mengekspresikan pendapat. Beragam platform menawarkan fasilitas interaksi sosial secara daring, namun YouTube menempati posisi paling dominan di antara *platform-platform* tersebut. Menurut data dari Statista, YouTube tercatat memiliki lebih dari 2,5 miliar pengguna aktif bulanan secara global pada tahun 2023, menjadikannya salah satu platform berbagi video paling populer di dunia [1]. YouTube telah berkembang menjadi sebuah media serbaguna yang melampaui fungsi awalnya sebagai platform berbagi video; selain menyediakan berbagai konten hiburan, YouTube juga berfungsi sebagai media untuk pendidikan, promosi, dan diskusi publik [2].

Fitur komunikasi antar pengguna di situs web, termasuk YouTube, umumnya difasilitasi melalui kolom komentar. Melalui fitur ini, pengguna dapat memberikan tanggapan, mengajukan pertanyaan, atau menyampaikan apresiasi. Namun, seiring dengan meningkatnya volume informasi dan interaksi di dalam kolom komentar, muncul berbagai tantangan terkait keamanan konten. Setiap platform yang memungkinkan penggunanya menghasilkan konten dan mengekspresikan opini publik rentan terhadap penyalahgunaan, termasuk dalam bentuk komentar yang bersifat negatif atau melanggar aturan. Salah satu bentuk penyalahgunaan yang kerap terjadi adalah penggunaan kolom komentar untuk menyelundupkan promosi aktivitas ilegal, seperti judi online. Awalnya praktik ini banyak ditemukan di forum-forum ilegal yang terkait dengan industri kasino, namun saat ini telah berkembang luas hingga ke media sosial. Penyebaran konten semacam ini tidak hanya melanggar Pedoman Komunitas YouTube, tetapi juga berpotensi menimbulkan dampak negatif secara moral, psikologis, dan finansial, terutama terhadap kelompok pengguna yang rentan seperti anak-anak dan remaja [3], [4]. Promosi judi online melalui komentar di YouTube sering disusun dengan

berbagai teknik persuasif dan manipulatif. Komentar-komentar semacam ini umumnya mengikuti pola tertentu, seperti penggunaan kata-kata yang menawarkan keuntungan instan, disertai dengan penggunaan emoji secara berlebihan dan penyisipan tautan eksternal yang mengarahkan pengguna ke situs perjudian ilegal [5]. Pelaku penyalahgunaan seringkali memanfaatkan teknik manipulatif seperti mengganti huruf dengan simbol, menambahkan spasi berlebihan, atau menyisipkan karakter tidak dikenal untuk mengelabui sistem deteksi otomatis [6]. Penerapan teknik *machine learning*, khususnya *deep learning*, telah terbukti mampu secara signifikan meningkatkan efektivitas dalam mendeteksi konten berbahaya [7].

Meskipun YouTube telah menerapkan sistem moderasi berbasis kecerdasan buatan untuk menyaring komentar yang tidak pantas, sejumlah besar komentar bermasalah masih berhasil lolos dari deteksi [8]. Sistem moderasi otomatis memiliki sejumlah kelemahan, khususnya dalam mendeteksi penggunaan bahasa manipulatif yang disamarkan atau tersembunyi dalam makna [9]. Keterbatasan ini semakin diperparah oleh tingginya volume komentar yang diunggah setiap detik, yang pada skala global dapat mencapai jutaan komentar per hari. Kondisi tersebut menyebabkan moderasi manual menjadi tidak efisien dan hampir mustahil untuk dilakukan secara konsisten [10]. Oleh karena itu, diperlukan pengembangan alat moderasi yang lebih cerdas dan adaptif, yang mampu memahami konteks komentar secara lebih mendalam. Salah satu pendekatan yang diharapkan dapat mengatasi permasalahan ini adalah melalui penerapan teknik pembelajaran mesin (*machine learning*), khususnya dalam penyaringan teks. Pembelajaran mesin memungkinkan sistem untuk belajar dari data historis, seperti komentar yang telah diberi label sebagai spam atau non-spam, sehingga mampu mengklasifikasikan komentar baru secara otomatis. Berbagai model supervised learning, seperti *Naive Bayes, Support Vector Machine (SVM)*, dan *Random Forest*, telah terbukti efektif dalam mendeteksi komentar spam dan konten berbahaya [11]. Metode pembelajaran mesin ini berkaitan erat dengan bidang pemrosesan bahasa alami *(Natural Language Processing/NLP)*, sebuah cabang dari kecerdasan buatan yang berfokus pada pengembangan sistem agar mampu memproses dan memahami

bahasa manusia. Dengan memanfaatkan teknik NLP, sistem dapat menganalisis teks melalui pemeriksaan terhadap pilihan kata dan struktur kalimat yang digunakan [12], [13]. Hal ini menjadi sangat penting mengingat komentar promosi terkait perjudian online seringkali disamarkan sedemikian rupa untuk menghindari deteksi oleh sistem moderasi otomatis [14]. *Natural Language Processing (NLP)* memungkinkan model untuk memahami maksud di balik sebuah komentar serta menyesuaikan analisisnya dengan berbagai variasi gaya bahasa yang digunakan oleh pengguna.

Dalam implementasi dunia nyata, kombinasi antara pemrosesan bahasa alami (Natural Language Processing/NLP) dan teknik pembelajaran mesin umumnya diterapkan untuk mendeteksi serta menyaring spam, ujaran kebencian, dan konten berbahaya di berbagai platform daring [15]. Teknologi ini mampu mengenali pola kompleks dalam teks, serta memahami struktur dan makna di baliknya. Studi yang dilakukan oleh Chollet menunjukkan bahwa model yang dirancang dengan baik dapat menjadikan sistem deep learning sangat efektif dalam memahami makna dan tujuan di balik kata-kata, bahkan ketika kata-kata tersebut dipelintir atau dimodifikasi. Model-model seperti Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), serta model yang lebih baru seperti BERT dan RoBERTa, telah terbukti menunjukkan kinerja yang baik dalam tugas klasifikasi teks yang membutuhkan pemahaman konteks. Pendekatan berbasis machine learning dan pemrosesan bahasa alami (NLP) sejalan dengan perkembangan di bidang keamanan siber dan etika digital. Teknologi ini berperan penting dalam mendukung regulasi digital, melindungi pengguna, serta mengatasi kompleksitas dan penyamaran dalam kejahatan digital yang terus berkembang[16], [17]. Dengan demikian, penerapan pembelajaran mesin dan pemrosesan bahasa alami dalam memilah komentar terkait perjudian online di YouTube tidak hanya berfokus pada aspek teknologinya semata [18]. Penelitian ini akan difokuskan pada pengembangan model pembelajaran mesin yang mampu menyaring komentar yang mengandung iklan perjudian online di platform YouTube. Pendekatan ini bertujuan untuk menghasilkan sistem yang lebih akurat dan adaptif

dibandingkan dengan metode tradisional, sekaligus berkontribusi terhadap pengembangan pengetahuan dan penerapannya dalam regulasi konten daring.

1.2 *Rumusan Masalah*

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

1. Metode apa yang paling efektif untuk digunakan dalam mengumpulkan dan untuk melabel komentar YouTube yang mengandung unsur promosi judi online ?
2. Algoritma *machine learning* apa yang paling efektif untuk melakukan klasifikasi komentar judi online ?
3. Sejauh apa model klasifikasi machine learning mampu mencapai akurasi dan efisiensi optimal dalam mendukung sistem moderasi otomatis komentar spam judi online di YouTube ?

1.3 *Batasan Masalah*

Untuk menjaga fokus dan ruang lingkup penelitian agar tetap terarah, maka batasan-batasan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan terbatas hanya pada 20 *channel* yang telah diseleksi dan berasal dari Indonesia serta komentar dari 50 video terakhir pada laman YouTube untuk masing-masing *channel*. 50 Video terakhir serta komentar dikumpulkan pada 9 Maret 2025, data diatas tanggal tersebut tidak termasuk dalam penelitian ini.
2. Komentar yang dianalisis hanya diklasifikasikan ke dalam dua kategori, yaitu komentar terkait *spam* promosi judi online dan komentar non-*spam*.

1.4 *Tujuan dan Manfaat Penelitian*

1.4.1 Tujuan Penelitian

Adapun tujuan daripada penelitian ini adalah sebagai berikut:

1. Mengumpulkan data komentar untuk mengklasifikasi komentar *spam* yang mengandung unsur judi *online*.
2. Mengimplementasikan teknik ekstraksi fitur beserta algoritma klasifikasi yang mampu mencapai F1-score ≥ 0,80 pada data uji, guna menghasilkan model yang stabil dan andal.
3. Mengevaluasi efektivitas model dalam mendukung sistem moderasi komentar otomatis yang ringan dan dapat dioperasikan secara real-time.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. **Manfaat Akademis:**
   Memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang *Natural Language Processing* (NLP) dan *machine learning*, melalui penerapan metode klasifikasi teks untuk mendeteksi konten komentar berbahaya seperti promosi judi online.
2. **Manfaat Praktis:**
   Menyediakan pendekatan berbasis teknologi yang dapat digunakan sebagai dasar dalam pengembangan sistem moderasi otomatis pada platform digital, guna membantu mengurangi penyebaran komentar yang melanggar aturan seperti spam judi online.
3. **Manfaat Sosial:**
   Mendukung terciptanya lingkungan digital yang lebih sehat, aman, dan produktif, dengan meminimalisir paparan konten ilegal kepada pengguna—khususnya anak-anak dan remaja—melalui peningkatan efektivitas moderasi komentar.

1.5 *Sistematika Penulisan*

Berikut sistematika penulisan pada skripsi yang telah disusun oleh penulis:

BAB I PENDAHULUAN

Bagian pendahuluan memuat uraian latar belakang, perumusan masalah, batasan penelitian, tujuan, serta manfaat studi. Pada bagian ini dijelaskan persoalan yang menjadi fokus penelitian sekaligus alasan mengapa topik tersebut penting untuk diangkat.

BAB II LANDASAN TEORI

Bab ini menguraikan landasan teoretis terkait framework, perangkat lunak, dan algoritma yang menunjang penelitian. Tinjauan teori disusun berdasarkan referensi jurnal ilmiah dan buku yang membahas metode maupun studi relevan sebelumnya.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian memuat penjelasan tentang objek studi, tahapan atau metode yang diterapkan, teknik dan alur pengumpulan data, serta definisi operasional tiap variabel penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan penerapan metodologi pada objek penelitian guna mencapai tujuan studi, mencakup uraian proses pengolahan dan analisis data hingga penyajian hasil akhir yang diperoleh.

BAB V KESIMPULAN DAN SARAN

Bab ini menyajikan simpulan penelitian dan rekomendasi yang disusun berdasarkan kendala maupun temuan analisis. Saran tersebut diharapkan bermanfaat bagi penelitian selanjutnya dengan topik atau tujuan serupa