

BAB I PENDAHULUAN

1.1 Latar Belakang

Perjudian online di Indonesia telah berkembang menjadi fenomena sosial yang semakin mengkhawatirkan, meskipun secara yuridis telah diatur dalam Pasal 303 Kitab Undang-Undang Hukum Pidana (KUHP) serta Pasal 27 ayat (2) juncto Pasal 45 ayat (2) Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) [1]. Namun, kendati telah memiliki landasan hukum yang jelas, praktik perjudian daring tetap menunjukkan tren peningkatan signifikan dari tahun ke tahun. Fenomena ini tidak lepas dari pesatnya perkembangan teknologi informasi, meluasnya akses internet di berbagai lapisan masyarakat [2], serta maraknya metode promosi terselubung yang memanfaatkan berbagai platform digital.

Berdasarkan laporan Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK) [3], nilai transaksi perjudian online di Indonesia melonjak tajam dari Rp2 triliun pada tahun 2017 menjadi Rp190 triliun sepanjang tahun 2022, dengan total transaksi mencapai lebih dari 157 juta kali. Lonjakan tersebut menunjukkan bahwa aktivitas perjudian daring tidak hanya berlangsung secara sporadis, melainkan telah menjadi bagian dari ekosistem digital yang aktif dan sistematis. Salah satu saluran yang paling sering dimanfaatkan dalam promosi perjudian online adalah YouTube. Platform ini menyediakan fitur seperti kolom komentar dan live chat yang memungkinkan penyebaran pesan secara masif dan real-time. Promosi judi kerap disamarkan melalui penggunaan istilah slang, kode tertentu, tautan tersembunyi, serta kalimat-kalimat persuasif yang ditujukan untuk menarik perhatian pengguna muda.

Dalam konteks ini, penting untuk memahami siapa yang menjadi sasaran utama dari promosi perjudian daring. Berdasarkan survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2024, jumlah pengguna internet di Indonesia telah mencapai 221,6 juta orang. Dari jumlah tersebut, kelompok usia Generasi Z

(12–27 tahun) dan Milenial (28–43 tahun) masing-masing mencakup 34,4% dan 30,6%. Kedua kelompok ini merupakan pengguna aktif media sosial [4], termasuk YouTube, sehingga menjadi target potensial yang rentan terhadap paparan konten promosi judi online [5].

Berkaca pada masifnya penyebaran konten promosi perjudian serta tingginya potensi keterpaparan di kalangan usia produktif, maka diperlukan sebuah solusi teknologi yang mampu mengidentifikasi dan mengeliminasi konten semacam ini secara cepat dan akurat. Salah satu pendekatan yang menjanjikan dalam hal ini adalah penggunaan metode *Machine Learning (ML)*, yang memiliki kemampuan untuk mengenali pola dan mengklasifikasikan konten berdasarkan data teks secara otomatis.

Sejumlah penelitian sebelumnya menunjukkan bahwa algoritma *machine learning* seperti *Naive Bayes* dan *Support Vector Machine (SVM)* telah berhasil digunakan dalam deteksi spam dan konten ilegal [6]. Namun demikian, adanya dinamika bahasa dalam komentar YouTube seperti penggunaan *eufemisme*, singkatan, serta penyamaran pesan menjadi tantangan tersendiri yang belum sepenuhnya dapat diatasi dengan pendekatan yang ada. Oleh karena itu, diperlukan evaluasi lebih lanjut terhadap performa algoritma *machine learning* dalam mendeteksi konten promosi judi yang tersembunyi di platform digital.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa lima algoritma *machine learning* populer yakni *Naive Bayes*, *Support Vector Machine*, *Random Forest*, *K-Nearest Neighbor (kNN)*, dan *Logistic Regression* dalam mendeteksi promosi judi online pada komentar YouTube. Evaluasi dilakukan berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*. Diharapkan hasil penelitian ini dapat berkontribusi pada pengembangan sistem moderasi otomatis yang lebih efektif serta mendukung upaya perlindungan masyarakat, khususnya generasi muda, dari dampak negatif perjudian online.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini ialah Bagaimana perbandingan kinerja lima algoritma *machine learning* *Naive Bayes*, *Support Vector Machine (SVM)*, *Random Forest*, *K-Nearest Neighbor (k-NN)*, dan *Logistic Regression* dalam mendeteksi promosi perjudian online pada komentar YouTube berdasarkan metrik evaluasi *akurasi*, *precision*, *recall*, dan *F1-score*? Serta algoritma manakah yang menunjukkan performa paling efektif dalam klasifikasi tersebut?

1.3 Batasan Masalah

Agar penelitian ini lebih fokus dan terarah, batasan masalah yang diterapkan adalah sebagai berikut

1. Dataset penelitian diperoleh dari komentar pada platform YouTube dalam periode Januari hingga Maret 2025.
2. Penelitian hanya mencakup komentar berbahasa Indonesia dan tidak melibatkan konten multimodal (gambar, video, atau tautan eksternal).
3. Penelitian dibatasi pada lima algoritma *machine learning* yaitu *Naive Bayes*, *Support Vector Machine*, *Random Forest*, *K-Nearest Neighbor*, dan *Logistic Regression* tanpa melakukan optimasi *hyperparameter* yang mendalam.
4. Evaluasi performa algoritma hanya menggunakan metrik *akurasi*, *precision*, *recall*, dan *F1-score*, tanpa mempertimbangkan aspek komputasi seperti waktu pelatihan atau efisiensi penggunaan sumber daya.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut

1. Menganalisis kemampuan lima algoritma *machine learning* *Naive Bayes*,

Support Vector Machine (SVM), *Random Forest*, *K-Nearest Neighbor (k-NN)*, dan *Logistic Regression* dalam mendeteksi konten promosi perjudian online pada komentar YouTube.

2. Membandingkan performa masing-masing algoritma berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score* dan menentukan algoritma yang paling efektif dalam melakukan klasifikasi komentar yang mengandung promosi judi online.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat teoritis dan praktis sebagai berikut.

1. Manfaat Teoritis
 - a. Memberikan wawasan tambahan tentang performa algoritma *machine learning* dalam deteksi konten berbahaya khususnya promosi judi online di platform digital.
 - b. Memperkaya literatur ilmiah mengenai penerapan *machine learning* dalam moderasi konten digital berbahasa Indonesia.
2. Manfaat Praktis
 - a. Memberikan dukungan bagi platform digital dalam mengembangkan sistem pemantauan komentar melalui rekomendasi algoritma *machine learning* yang mampu mendeteksi promosi perjudian online secara cepat, akurat, dan efisien.
 - b. Mendukung implementasi kebijakan pemerintah dan lembaga terkait dalam upaya perlindungan masyarakat, terutama generasi muda dari dampak negatif perjudian online.

1.6 Sistematika Penulisan

Sistematika Sistematika penulisan skripsi ini disusun untuk memberikan gambaran yang jelas mengenai alur pembahasan, sehingga memudahkan pembaca

dalam memahami isi dan tujuan penelitian. Adapun sistematika penulisan skripsi ini adalah sebagai berikut

1. BAB I PENDAHULUAN

Bab ini memuat uraian mengenai latar belakang yang menjelaskan alasan dilakukannya penelitian terkait maraknya promosi judi online di platform YouTube, rumusan masalah yang menjadi fokus kajian, batasan masalah yang diterapkan agar penelitian lebih terarah, tujuan penelitian yang hendak dicapai, manfaat penelitian baik secara teoritis maupun praktis, serta sistematika penulisan skripsi. Pada bab ini penulis menitikberatkan pada urgensi pengembangan sistem deteksi otomatis komentar promosi judi online dan alasan pemilihan lima algoritma machine learning sebagai objek perbandingan.

2. BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian pustaka yang meliputi penelitian-penelitian terdahulu yang relevan dengan topik penelitian, serta landasan teori yang digunakan. Pembahasan mencakup konsep judi online dan regulasinya di Indonesia, pengertian dan prinsip kerja machine learning, uraian mengenai algoritma Naive Bayes, Support Vector Machine, Random Forest, Logistic Regression, dan K-Nearest Neighbor, teknik representasi data teks menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), serta metrik evaluasi performa model klasifikasi teks seperti akurasi, precision, recall, F1-score, dan Area Under Curve (AUC).

3. BAB III METODE PENELITIAN

Bab ini menjelaskan rancangan dan tahapan penelitian secara rinci. Pokok pembahasan mencakup objek penelitian yaitu komentar YouTube yang mengandung indikasi promosi judi online, alur penelitian mulai dari pengumpulan data melalui teknik scraping, tahap preprocessing (pembersihan dan normalisasi data), pelabelan data secara manual, penyeimbangan data (undersampling), proses tokenisasi menggunakan custom regular expression, dan representasi data dengan TF-IDF. Selain itu, bab ini juga menguraikan proses pembagian data

menjadi training set dan testing set, pelatihan model menggunakan lima algoritma machine learning, evaluasi performa model, serta deskripsi perangkat keras dan lunak yang digunakan.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi model dan analisis perbandingan performa algoritma. Pembahasan meliputi hasil preprocessing dan distribusi data setelah penyeimbangan, visualisasi kata kunci menggunakan wordcloud, hasil pengujian tiap algoritma berdasarkan metrik evaluasi, serta analisis terhadap kelebihan dan kelemahan masing-masing algoritma. Pada bab ini ditunjukkan bahwa Random Forest memberikan performa terbaik, diikuti SVM, dengan pembahasan mendalam mengenai faktor-faktor yang mempengaruhi kinerja model.

5. BAB V PENUTUP

Bab ini memuat kesimpulan yang diperoleh dari penelitian, yang merangkum hasil perbandingan performa kelima algoritma dan memberikan rekomendasi algoritma yang paling efektif untuk deteksi komentar promosi judi online di YouTube. Bab ini juga memuat saran untuk penelitian selanjutnya, seperti penerapan teknik optimasi hyperparameter, dan pengujian sistem pada data real-time.