

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan media sosial telah meningkatkan intensitas interaksi digital masyarakat, khususnya melalui fitur komentar. Salah satu platform dengan tingkat interaksi yang tinggi adalah YouTube, yang pada Januari 2025 memiliki lebih dari 2,53 miliar pengguna aktif bulanan secara global. Indonesia sendiri menyumbang sekitar 143 juta pengguna aktif dan termasuk dalam lima besar negara dengan jumlah pengguna terbanyak di dunia [1]. Tingginya jumlah pengguna tersebut menjadikan kolom komentar sebagai ruang interaksi yang sangat aktif, di mana jutaan komentar dipublikasikan setiap harinya pada berbagai kategori video.

Namun, tingginya aktivitas pada kolom komentar juga membuka peluang terjadinya penyalahgunaan, salah satunya berupa spam promosi judi online. Komentar tersebut umumnya berisi ajakan, tautan, maupun penyebutan nama situs tertentu yang disebarluaskan secara masif pada berbagai video [2]. Aktivitas ini tidak hanya mengganggu kenyamanan pengguna, tetapi juga berpotensi menimbulkan dampak sosial dan ekonomi yang merugikan. Penyebaran promosi perjudian secara luas dapat menjangkau kelompok usia rentan serta meningkatkan paparan terhadap praktik perjudian ilegal [3].

Hal ini sejalan dengan kondisi di Indonesia, di mana judi online telah berkembang menjadi permasalahan serius yang melibatkan jutaan masyarakat serta menimbulkan kerugian ekonomi dalam skala besar. Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK) mencatat bahwa pada tahun 2023 terdapat sekitar 4 juta pemain judi online dengan total perputaran dana mencapai Rp327 triliun. Pada tahun 2024, jumlah tersebut meningkat signifikan menjadi sekitar 8,8 juta orang, termasuk sekitar 80 ribu anak di bawah usia 10 tahun, dengan nilai transaksi yang diperkirakan mencapai Rp900 triliun. Tingginya angka keterlibatan serta besarnya perputaran dana tersebut menunjukkan bahwa praktik judi online telah berada pada kondisi yang mengkhawatirkan dan bahkan dapat dikategorikan sebagai situasi

“darurat”, sehingga memerlukan perhatian serius dalam upaya pencegahan dan penanganannya [4].

Permasalahan tersebut menunjukkan bahwa diperlukan suatu mekanisme penyaringan otomatis yang mampu mendeteksi serta memisahkan komentar spam dan non-spam secara akurat. Karakteristik kolom komentar YouTube yang bersifat terbuka dan tidak terstruktur menyebabkan data teks yang dihasilkan memiliki tingkat variasi bahasa yang tinggi, termasuk penggunaan bahasa informal, singkatan, simbol, hingga manipulasi karakter untuk menghindari deteksi system [5]. Hal ini menjadi tantangan tersendiri dalam proses klasifikasi teks, sehingga diperlukan pendekatan komputasi yang mampu memahami serta mengolah teks secara efektif.

Salah satu pendekatan yang dapat digunakan adalah Natural Language Processing (NLP) berbasis machine learning. Pendekatan ini mampu mengolah data teks tidak terstruktur menjadi representasi numerik melalui tahapan preprocessing dan ekstraksi fitur, seperti Term Frequency-Inverse Document Frequency (TF-IDF) [6]. Dengan demikian, sistem dapat melakukan proses klasifikasi secara otomatis terhadap komentar yang termasuk kategori spam maupun non-spam.

Dua algoritma yang umum digunakan dalam klasifikasi teks dan memiliki tingkat efektivitas yang tinggi adalah Naïve Bayes (NB) dan Support Vector Machine (SVM). Naïve Bayes merupakan algoritma berbasis probabilistik yang sederhana, cepat, dan efektif dalam menangani data berdimensi tinggi seperti teks [7]. Sementara itu, Support Vector Machine bekerja dengan mencari *hyperplane* optimal untuk memisahkan dua kelas data dan dikenal memiliki performa yang baik pada data teks yang kompleks dan berdimensi tinggi [8].

Meskipun kedua algoritma tersebut telah banyak digunakan, hasil penelitian sebelumnya menunjukkan temuan yang beragam. Beberapa penelitian menyatakan bahwa SVM memiliki tingkat akurasi yang lebih tinggi pada data teks yang kompleks, sedangkan Naïve Bayes lebih unggul dalam hal efisiensi dan stabilitas pada dataset berukuran besar [9]. Selain itu, sebagian besar penelitian hanya berfokus pada perbandingan algoritma tanpa mempertimbangkan karakteristik

husus data, seperti adanya manipulasi teks atau penggunaan pola tertentu pada komentar spam. Padahal, dalam kasus spam judi online, komentar sering kali mengandung pola spesifik seperti penyebutan nama situs, variasi penulisan (leetspeak), serta modifikasi karakter untuk menghindari deteksi.

Berdasarkan hal tersebut, terdapat celah penelitian (*research gap*) yaitu belum banyak penelitian yang secara khusus menganalisis perbandingan kinerja algoritma Naïve Bayes dan SVM pada klasifikasi komentar spam judi online di YouTube dengan mempertimbangkan karakteristik unik teks spam, serta membandingkan pendekatan berbasis *machine learning* murni dengan pendekatan yang memanfaatkan deteksi pola atau entitas tertentu dalam proses pelabelan data.

Selain itu, dataset komentar YouTube umumnya memiliki distribusi kelas yang tidak seimbang (*class imbalance*), di mana jumlah komentar non-spam lebih dominan dibandingkan komentar spam. Kondisi ini dapat mempengaruhi kinerja model, terutama dalam mendeteksi kelas minoritas [10]. Oleh karena itu, pada penelitian ini dilakukan penerapan teknik penyeimbangan data (*data balancing*) untuk mengevaluasi pengaruhnya terhadap peningkatan kinerja model klasifikasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan analisis perbandingan kinerja algoritma Naïve Bayes dan Support Vector Machine dalam klasifikasi komentar spam promosi judi online pada platform YouTube. Evaluasi kinerja dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem moderasi otomatis yang lebih efektif serta membantu menjaga kualitas interaksi digital pada platform YouTube.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini sebagai berikut :

1. Bagaimana kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan komentar spam promosi judi online pada platform YouTube?
2. Bagaimana pengaruh penerapan teknik *data balancing* terhadap kinerja

klasifikasi komentar spam judi online?

3. Algoritma manakah yang memiliki kinerja terbaik dalam mendeteksi komentar spam judi online berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka diperlukan batasan masalah sebagai berikut:

1. Penelitian ini berfokus pada proses pengolahan, analisis, dan evaluasi data komentar YouTube menggunakan lingkungan komputasi berbasis Python pada Google Colaboratory.
2. Sistem yang dikembangkan tidak diimplementasikan dalam bentuk aplikasi atau antarmuka pengguna, melainkan berupa model klasifikasi yang digunakan untuk keperluan analisis penelitian.
3. Data penelitian dibatasi pada komentar publik YouTube yang diperoleh melalui proses crawling dari beberapa video dengan tingkat interaksi tinggi pada kategori podcast, hiburan, edukasi, dan dakwah selama periode pengambilan data.
4. Tahapan *preprocessing* yang digunakan meliputi *Unicode normalization*, *case folding*, *leetspeak conversion*, *character compression*, dan *text cleaning* untuk menormalisasi variasi penulisan pada teks komentar.
5. Pelabelan data dilakukan secara otomatis menggunakan pendekatan *rule-based entity detection* berbasis daftar kata kunci terkait situs judi online serta pola *regular expression* yang fleksibel.
6. Representasi teks dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) dengan parameter yang ditentukan pada tahap eksperimen.
7. Algoritma klasifikasi yang digunakan dalam penelitian ini dibatasi pada Naïve Bayes dan Support Vector Machine (SVM).

8. Penelitian ini mempertimbangkan kondisi ketidakseimbangan kelas (*class imbalance*) dengan melakukan pengujian pada dataset asli (tanpa *data balancing*) serta dataset yang telah melalui teknik *data balancing* menggunakan metode *Random UnderSampling*.
9. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta didukung oleh analisis *confusion matrix*.
10. Penelitian ini tidak membahas faktor eksternal seperti perubahan algoritma moderasi YouTube, dinamika perilaku pengguna, maupun sistem deteksi spam internal pada platform YouTube.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengimplementasikan teknik Natural Language Processing (NLP) untuk melakukan pengolahan dan representasi teks komentar pada platform YouTube.
2. Menganalisis kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan komentar spam promosi judi online.
3. Menganalisis pengaruh penerapan teknik *data balancing* terhadap performa model dalam menangani ketidakseimbangan kelas (*class imbalance*).
4. Membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan metode yang paling optimal.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis, sebagai berikut:

1. Manfaat Teoritis

Penelitian ini memberikan kontribusi ilmiah dalam bidang Natural Language Processing (NLP) dan machine learning, khususnya pada penerapan

algoritma Naïve Bayes dan Support Vector Machine dalam klasifikasi teks komentar media sosial berbahasa Indonesia.

Penelitian ini juga memberikan gambaran empiris mengenai pengaruh kondisi ketidakseimbangan kelas (*class imbalance*) serta penerapan teknik data balancing terhadap performa model klasifikasi. Hasil penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang mengkaji deteksi spam berbasis teks pada platform media sosial.

2. Manfaat Praktis

Secara praktis, penelitian ini dapat menjadi acuan dalam pengembangan sistem deteksi komentar spam berbasis teks, khususnya yang berkaitan dengan promosi judi online pada platform YouTube. Hasil analisis performa model dapat membantu pengembang sistem atau peneliti dalam memilih pendekatan klasifikasi yang lebih sesuai berdasarkan kebutuhan, apakah menekankan *precision* tinggi atau *recall* yang lebih optimal.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Bab ini menjelaskan alasan dilakukannya penelitian, ruang lingkup permasalahan, serta arah dan tujuan penelitian yang akan dicapai.

BAB II TINJAUAN PUSTAKA, Bab ini berisi kajian teori dan penelitian terdahulu yang relevan dengan topik penelitian. Pembahasan mencakup konsep *Natural Language Processing* (NLP), algoritma Naïve Bayes dan Support Vector Machine (SVM), teknik preprocessing teks, konsep data balancing, serta metode evaluasi klasifikasi. Selain itu, bab ini juga menguraikan penelitian terdahulu sebagai landasan teoritis yang mendukung penelitian.

BAB III METODOLOGI PENELITIAN, Bab ini menjelaskan metode yang digunakan dalam penelitian, meliputi tahapan penelitian, sumber dan teknik pengumpulan data, proses *preprocessing* teks, pelabelan data menggunakan

pendekatan *rule-based entity detection*, penerapan teknik *data balancing*, ekstraksi fitur menggunakan TF-IDF, serta perancangan model klasifikasi menggunakan algoritma Naïve Bayes dan SVM. Selain itu, dijelaskan pula metode evaluasi yang digunakan untuk mengukur kinerja model.

BAB IV HASIL DAN PEMBAHASAN, Bab ini menyajikan hasil implementasi algoritma Naïve Bayes dan SVM terhadap dataset komentar YouTube yang telah diproses. Pembahasan mencakup analisis kinerja model, perbandingan hasil pada berbagai skenario eksperimen (dengan dan tanpa *data balancing*), serta interpretasi hasil berdasarkan metrik evaluasi yang digunakan.

BAB V PENUTUP, Bab ini berisi kesimpulan yang diperoleh berdasarkan hasil penelitian serta saran yang dapat digunakan sebagai bahan pengembangan lebih lanjut dalam penelitian terkait klasifikasi teks dan deteksi spam pada media sosial.