

Implementasi Kecerdasan Buatan dalam Deteksi *Cybercrime*: Komparasi Model *Naive Bayes* dan *SVM* pada Pola Komentar Judi Online

Ardiansyah¹, Abdul Muin Nasution², Muhammad Syahputra Novelan³

^{1,3} Program Studi Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

² Randwick International Research and Analysis Institute, Medan, Indonesia

Article Info

Article history:

Received Desember 1, 2025

Revised Desember 15, 2025

Accepted Desember 30, 2025

Kata Kunci:

Kecerdasan Buatan,
Machine Learning,
Naive Bayes,
Support Vector Machine,
Judi Online

Keywords:

Artificial Intelligence,
Machine Learning,
Naive Bayes,
Support Vector Machine,
Judi Online

ABSTRAK

Maraknya promosi judi online di platform media sosial seperti YouTube telah menjadi ancaman serius dalam kategori *cybercrime* di Indonesia. Pola komentar yang bervariasi dan penggunaan bahasa non-formal menyulitkan identifikasi konten secara manual. Penelitian ini bertujuan untuk mengimplementasikan teknologi Kecerdasan Buatan (AI) melalui pendekatan Machine Learning untuk mendeteksi secara otomatis pola komentar judi online. Dua algoritma populer, yaitu Naive Bayes (NB) dan Support Vector Machine (SVM), digunakan dan dibandingkan kinerjanya untuk menentukan model klasifikasi terbaik. Data penelitian diekstraksi dari komentar YouTube berbahasa Indonesia, yang kemudian melewati tahap pra-pemrosesan teks meliputi case folding, tokenization, stopword removal, dan stemming. Fitur teks ditransformasikan menjadi bentuk numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Hasil penelitian menunjukkan perbandingan kinerja kedua algoritma berdasarkan metrik akurasi, precision, recall, dan f1-score. Temuan ini diharapkan dapat memberikan kontribusi bagi pengembangan sistem filtrasi konten negatif otomatis guna memperkuat keamanan siber di ekosistem digital Indonesia.

ABSTRACT

The proliferation of online gambling promotions on social media platforms such as YouTube has become a serious threat in the category of cybercrime in Indonesia. Varied comment patterns and the use of informal language make it difficult to identify content manually. This study aims to implement Artificial Intelligence (AI) technology through a Machine Learning approach to automatically detect online gambling comment patterns. Two popular algorithms, Naive Bayes (NB) and Support Vector Machine (SVM), were used and their performance was compared to determine the best classification model. The research data was extracted from Indonesian-language YouTube comments, which then underwent text pre-processing stages including case folding, tokenization, stopword removal, and stemming. Text features were transformed into numerical form using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The results of the study show a comparison of the performance of the two algorithms based on accuracy, precision, recall, and f1-score metrics. These findings are expected to contribute to the development of an automatic negative content filtration system to strengthen cyber security in Indonesia's digital ecosystem.



Corresponding Author:

Abdul Muin Nasution
Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara
Medan, Indonesia
Email: muinnasution@gmail.com

1. PENDAHULUAN

Keamanan ekosistem digital di Indonesia saat ini menghadapi tantangan yang semakin kompleks, baik dari sisi teknis infrastruktur maupun dari sisi konten informasi yang beredar secara masif di ruang siber. Perkembangan teknologi internet dan media sosial telah membawa dampak positif dalam memperluas akses informasi dan komunikasi, namun di sisi lain juga membuka peluang bagi munculnya berbagai bentuk kejahatan siber. Salah satu ancaman yang semakin nyata adalah maraknya promosi judi online yang tersebar luas melalui kolom komentar pada platform berbagi video seperti YouTube. Fenomena ini merupakan bentuk kejahatan siber yang memanfaatkan celah pada interaksi sosial digital dan rendahnya kesadaran keamanan informasi di kalangan pengguna. Sejalan dengan hal tersebut, setiap komputer yang terhubung ke jaringan komputer berskala luas memiliki potensi kerentanan terhadap berbagai bentuk penyalahgunaan data, manipulasi informasi, serta gangguan sistem oleh pihak yang tidak bertanggung jawab [1].

Kerentanan sistem informasi yang sebelumnya didominasi oleh serangan teknis pada jaringan kini telah mengalami pergeseran paradigma. Ancaman siber tidak lagi hanya berbentuk peretasan (*hacking*) atau serangan *malware*, tetapi juga berkembang menjadi manipulasi konten informasi secara sistematis dan terorganisir. Kolom komentar YouTube sering dimanfaatkan sebagai media penyebaran promosi judi online karena sifatnya yang terbuka, mudah diakses, dan mampu menjangkau audiens dalam jumlah besar. Serangan informasi semacam ini sering kali menjadi pintu masuk bagi kejahatan siber lanjutan, seperti penipuan daring, pencurian identitas, dan manipulasi psikologis (*social engineering*), yang apabila dibiarkan akan merusak kepercayaan publik serta integritas ekosistem digital nasional.

Dalam menghadapi kompleksitas permasalahan tersebut, implementasi Kecerdasan Buatan (*Artificial Intelligence / AI*) menjadi salah satu solusi strategis yang semakin mendapat perhatian. AI, khususnya melalui pendekatan *Machine Learning*, memungkinkan sistem komputer untuk belajar dari data historis dan mengenali pola tertentu tanpa harus diprogram secara eksplisit. Penerapan *Machine Learning* dalam deteksi konten ilegal atau berbahaya memberikan kemampuan kepada sistem untuk melakukan klasifikasi secara otomatis, cepat, dan konsisten. Hal ini menjadi sangat krusial mengingat pelaku promosi judi online kerap menggunakan variasi bahasa, simbol, serta istilah samaran guna menghindari sistem moderasi bawaan platform. Penelitian terkini menunjukkan bahwa pemanfaatan model klasifikasi teks berbasis pembelajaran mesin mampu meningkatkan efektivitas deteksi konten judi online secara signifikan pada platform media sosial, khususnya YouTube [2].

Penggunaan AI dalam konteks ini juga mencerminkan konsep komputer cerdas, di mana sistem tidak hanya berfungsi sebagai alat komputasi, tetapi juga mampu melakukan analisis, penalaran, dan pengambilan keputusan berbasis data. Melalui proses abstraksi data, teks mentah yang awalnya tidak terstruktur dapat diubah menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Dengan meningkatnya volume data komentar digital setiap harinya, ketergantungan pada sistem cerdas berbasis AI tidak lagi bersifat opsional, melainkan menjadi kebutuhan mendasar dalam upaya menjaga keamanan siber modern. Studi terbaru mengonfirmasi bahwa integrasi teknik *Natural Language*

Processing (NLP) dengan algoritma pembelajaran mesin mampu mengidentifikasi pola promosi judi online yang bersifat spam secara lebih akurat, bahkan ketika menggunakan bahasa informal dan slang khas Indonesia [3].

Fenomena komentar judi online di YouTube memiliki ciri khas linguistik tertentu. Contoh konkret yang sering dijumpai meliputi penggunaan kata-kata persuasif seperti “Gacor Parah”, “Slot Anti Rungkad”, “Pola Maxwin Malam Ini”, serta penawaran yang bersifat manipulatif seperti “Deposit Pulsa Tanpa Potongan” dan “Bonus New Member 100%”. Pola ini hampir selalu diikuti dengan penyematan *shortlink*, tautan mencurigakan, atau nomor WhatsApp anonim. Ancaman yang ditimbulkan oleh konten tersebut bersifat multidimensi. Selain mendorong perilaku adiktif terhadap perjudian yang berdampak pada aspek ekonomi dan kesehatan mental, tautan-tautan tersebut sering kali mengarah pada situs *phishing* atau mengandung *malware* yang dapat menginfeksi perangkat pengguna. Penelitian berbasis *deep learning* menunjukkan bahwa komentar promosi judi online memiliki pola struktur teks berulang yang dapat dikenali oleh sistem klasifikasi otomatis jika dilatih dengan data yang memadai [4].

Penelitian mengenai deteksi konten negatif menggunakan kecerdasan buatan telah berkembang pesat dalam beberapa tahun terakhir. Berbagai algoritma *Machine Learning* terbukti efektif dalam mengenali pola teks tidak terstruktur yang tersebar pada media sosial [5]. Namun demikian, tantangan yang cukup signifikan muncul pada konteks bahasa Indonesia yang tidak baku, dinamis, serta kaya akan singkatan dan istilah slang. Kompleksitas ini menuntut model AI memiliki kemampuan generalisasi yang tinggi agar mampu membedakan antara komentar organik pengguna dengan komentar promosi judi online yang dihasilkan oleh bot atau akun anonim. Oleh karena itu, tahapan *feature engineering* dan *pre-processing* teks seperti tokenisasi, stemming, dan penghapusan *stopword* menjadi faktor kunci dalam meningkatkan kinerja model klasifikasi [6].

Dalam konteks implementasi teknologi informasi di Indonesia, pemanfaatan sistem cerdas telah diterapkan di berbagai sektor untuk mendukung efisiensi, keamanan, dan pengambilan keputusan berbasis data. Transformasi digital yang semakin cepat mendorong berbagai institusi untuk mengoptimalkan sarana dan prasarana serta menerapkan teknologi digital secara efektif dalam operasi mereka [7]. Penelitian terkini menunjukkan bahwa integrasi algoritma klasifikasi dalam sistem informasi mampu mendukung deteksi dini terhadap penyimpangan pola data dan konten berisiko tinggi pada platform digital [8]. Hal ini memperkuat urgensi penggunaan AI sebagai solusi moderasi konten otomatis, mengingat keterbatasan moderasi manual dari sisi kecepatan, konsistensi, serta kemampuan bekerja secara berkelanjutan selama 24 jam.

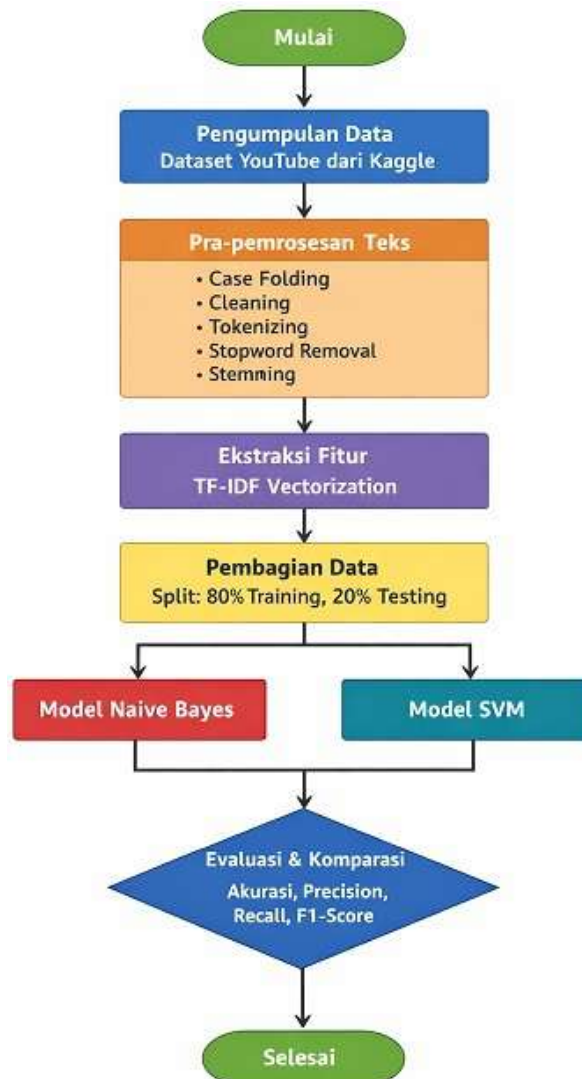
Berbagai penelitian sebelumnya juga telah membandingkan efektivitas algoritma Naive Bayes dan *Support Vector Machine* (SVM) dalam tugas klasifikasi teks, termasuk pada data berbahasa Indonesia [6]. Naive Bayes dikenal memiliki keunggulan dari sisi kesederhanaan, kecepatan pelatihan, dan efisiensi komputasi, sehingga cocok diterapkan pada sistem dengan sumber daya terbatas. Di sisi lain, SVM memiliki kemampuan yang kuat dalam menentukan *hyperplane* optimal dengan margin maksimum, sehingga lebih tangguh dalam menangani data teks yang bersifat kompleks dan non-linear. Penelitian lain menunjukkan bahwa pemilihan algoritma klasifikasi yang tepat sangat berpengaruh dalam meminimalkan kesalahan deteksi, khususnya *false positive* yang dapat merugikan pengguna normal [9].

Volume data komentar YouTube yang sangat besar dan terus bertambah setiap waktu menjadikan pengawasan manual tidak lagi memadai. Oleh karena itu, solusi cerdas berbasis klasifikasi otomatis menjadi pendekatan yang paling realistis dan berkelanjutan [10]. Evaluasi kinerja algoritma klasifikasi modern menunjukkan bahwa kombinasi teknik *pre-processing* yang tepat dengan model pembelajaran mesin yang sesuai mampu meningkatkan nilai *recall*, sehingga semakin sedikit komentar judi online yang lolos dari sistem deteksi [11]. Temuan ini menegaskan bahwa penelitian komparatif terhadap algoritma klasifikasi bukan sekadar kajian akademis, melainkan bagian penting dari strategi perlindungan ruang siber nasional.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada implementasi dan komparasi algoritma Naive Bayes dan SVM dalam mendeteksi pola komentar judi online pada platform YouTube. Kedua algoritma ini dipilih karena mewakili pendekatan klasik *Machine Learning* yang masih banyak digunakan dan relevan untuk klasifikasi teks. Dengan memanfaatkan bahasa pemrograman Python serta pustaka pendukung seperti Scikit-Learn, NLTK, dan Sastrawi, penelitian ini diharapkan mampu menghasilkan sistem deteksi otomatis yang akurat, efisien, dan dapat diimplementasikan secara nyata. Hasil dari penelitian ini diharapkan dapat menjadi referensi penting dalam pengembangan strategi moderasi konten berbasis *Machine Learning* di masa depan, guna menciptakan ruang digital Indonesia yang lebih aman, sehat, dan produktif bagi seluruh lapisan masyarakat.

2. METODE

Metodologi penelitian ini dirancang untuk membandingkan efektivitas dua algoritma *Machine Learning* dalam mendeteksi ancaman konten ilegal. Alur penelitian ini dilakukan secara sistematis yang meliputi tahap pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, klasifikasi, dan evaluasi hasil [12]. Selain itu, pendekatan klasifikasi berbasis data mining telah terbukti efektif dalam memproses karakteristik data untuk menghasilkan sistem penentuan kualitas yang akurat pada berbagai domain aplikasi [13]. Secara visual, tahapan penelitian ini disajikan dalam Gambar 1.



Gambar 1. Diagram Alir Metodologi Penelitian Klasifikasi Komentar Judi Online

2.1. Dataset dan Sumber Data

Data penelitian ini bersumber dari dataset publik pada platform Kaggle berjudul "*Komentar Judi Online*" [14]. Dataset ini berisi kumpulan teks komentar dari platform YouTube dalam format .csv yang telah diberi label biner: "Judi" dan "Bukan Judi". Data ini mencerminkan karakteristik nyata dari gangguan informasi di ruang siber Indonesia.

2.2. Lingkungan Pengembangan (Development Environment)

Seluruh tahapan eksperimen, mulai dari pengolahan data hingga pengujian model, diimplementasikan menggunakan bahasa pemrograman Python melalui platform berbasis *cloud* Google Colab (colab.research.google.com). Pustaka utama yang digunakan meliputi Pandas untuk manipulasi data, Scikit-Learn untuk pemodelan algoritma, serta Sastrawi untuk pengolahan bahasa alami. Pemilihan Python didasarkan pada kedudukannya sebagai bahasa pemrograman utama dalam pengembangan sistem komputer cerdas, yang memungkinkan implementasi algoritma machine learning secara efisien dan mampu menangani komputasi data besar dengan presisi tinggi [15]. Platform Google Colab juga mempermudah pembangunan dan pengujian model machine learning serta langkah-langkah pra-pemrosesan teks, termasuk penerapan algoritma seperti Naive Bayes dan SVM, melalui lingkungan Python berbasis cloud yang efisien dan tanpa memerlukan instalasi lokal [16]

2.3. Tahapan Pra-pemrosesan Teks (Text Preprocessing)

Untuk menjamin kualitas data, dilakukan pra-pemrosesan teks guna menghilangkan *noise*. Proses ini dijalankan secara bertahap penerapannya menggunakan compiler online kode python yakni colab.research.google.com dan setiap hasilnya diekspor ke dalam file .csv yang disimpan pada drive.google.com untuk validasi integritas data:

- 1) Case Folding: Mengubah seluruh teks menjadi huruf kecil (Output: *judi_online_casefolding.csv*).
- 2) Cleaning: Menghapus karakter non-alfabet seperti angka, tanda baca, simbol, dan URL. (Output: *judi_online_cleaned.csv*).
- 3) Tokenizing: Memecah kalimat menjadi unit kata tunggal (Output: *judi_online_tokenized.csv*).
- 4) Stopword Removal: Menghilangkan kata-kata umum yang tidak bermakna (Output: *judi_online_no_stopwords.csv*).
- 5) Stemming: Mengubah kata berimbuhan menjadi kata dasar menggunakan algoritma Sastrawi (Output: *judi_online_stemmed.csv*).

2.4. Ekstraksi Fitur dan Klasifikasi

Implementasi pada tahap ini dilakukan secara programatik untuk mengubah data teks yang telah bersih menjadi model prediktif. Proses ini merupakan inti dari penerapan paradigma komputer cerdas, di mana sistem melakukan pembelajaran mesin (Machine Learning) terhadap pola data secara mandiri tanpa instruksi pemrograman manual untuk setiap kata kunci. Tahapan teknisnya adalah sebagai berikut:

1. Vektorisasi TF-IDF.

Data teks dari file *judi_online_stemmed.csv* tidak dapat diproses langsung oleh algoritma klasifikasi. Oleh karena itu, dilakukan ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ini memberikan bobot numerik pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan kelangkaannya di seluruh dataset. Implementasi dilakukan menggunakan fungsi *TfidfVectorizer* dari pustaka Scikit-Learn [17].

2. Pembagian Data (Data Splitting)

Dataset dibagi menjadi dua bagian utama menggunakan fungsi *train_test_split*. Proporsi yang ditetapkan adalah 80% sebagai data latih (training set) untuk membangun pengetahuan model, dan 20% sebagai data uji (testing set) untuk mengevaluasi kemampuan generalisasi model terhadap data baru

[17]. Untuk menjaga konsistensi hasil eksperimen (reproduksibilitas), digunakan parameter `random_state=42`.

3. Pemodelan Naive Bayes dan SVM.

Setelah fitur numerik (dari langkah Vektorisasi TF-IDF) terbentuk, dilakukan proses pelatihan (*training*) terhadap dua algoritma:

a. Multinomial Naive Bayes (MNB)

Dipilih karena kemampuannya yang sangat efisien dalam menangani fitur diskrit seperti frekuensi kata dalam klasifikasi teks.

b. Support Vector Machine (SVM):

Diimplementasikan dengan *kernel linear* untuk mencari garis pemisah (*hyperplane*) maksimal yang membedakan antara komentar bermuatan judi dan bukan judi. Penggunaan SVM sangat efektif pada data dengan dimensi fitur yang tinggi seperti hasil dari TF-IDF.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Dataset dan Ekstraksi Fitur

Setelah melalui seluruh tahapan pra-pemrosesan teks (seperti *filtering*, *stemming*, dan *stopword removal*), data sebanyak 8.440 baris komentar YouTube diubah menjadi bentuk vektor numerik menggunakan metode TF-IDF. Berdasarkan output proses vektorisasi yang dibatasi pada fitur paling signifikan, dihasilkan matriks fitur dengan total 5.000 kata unik (fitur). Pembatasan ini bertujuan untuk mereduksi *noise* dan memastikan model hanya mempelajari kata-kata yang memiliki bobot informasi tinggi terhadap pola komentar judi online.

3.2. Perbandingan Kinerja Model

Evaluasi dilakukan menggunakan data uji (*testing set*) sebesar 20% dari total dataset, yakni sebanyak 1.688 komentar. Berdasarkan hasil pengujian menggunakan algoritma *Multinomial Naive Bayes* dan *Support Vector Machine* (SVM), didapatkan hasil performa yang disajikan pada Tabel 1.

Tabel 1. Perbandingan Akurasi Klasifikasi Terkini

Metrik Evaluasi	Multinomial Naive Bayes	Support Vector Machine (SVM)
Akurasi (%)	95,08%	96,45%
Precision (Rata-rata)	0.96	0.97
Recall (Rata-rata)	0.94	0.96
F1-Score (Rata-rata)	0.95	0.96
Jumlah Data Latih	6.752	6.752
Jumlah Data Uji	1.688	1.688
Total Fitur (Kata)	5.000	5.000

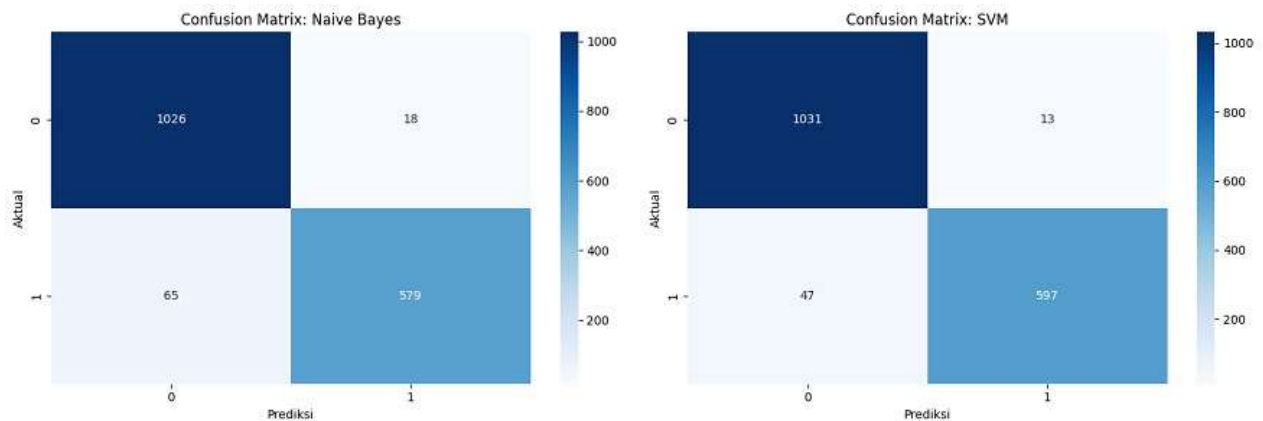
Berdasarkan Tabel 1, model Support Vector Machine (SVM) mencatatkan efektivitas tertinggi dengan akurasi sebesar 96,45%, unggul 1,37% dibandingkan *Multinomial Naive Bayes* yang memperoleh 95,08%. Nilai *Precision* yang mencapai 0.98 pada label judi (1) oleh model SVM menunjukkan bahwa model memiliki tingkat kesalahan *false positive* yang sangat rendah.

3.3. Diskusi dan Analisis Temuan

Hasil pengujian menunjukkan bahwa model SVM dengan kernel linear memberikan hasil paling optimal. Keunggulan SVM dalam menangani data teks berdimensi tinggi (5.000 fitur hasil TF-IDF) konsisten dengan teori bahwa SVM mampu memisahkan kelas secara efisien melalui hyperplane optimal di ruang vektor yang luas. SVM sangat efektif untuk tugas klasifikasi teks karena kemampuannya bekerja langsung dengan representasi numerik berdimensi tinggi tanpa memerlukan seleksi fitur agresif [17].

3.3.1. Visualisasi Confusion Matrix

Visualisasi confusion matrix diperlukan untuk mengevaluasi kinerja model klasifikasi secara lebih komprehensif karena mampu menunjukkan distribusi prediksi benar serta jenis kesalahan klasifikasi pada setiap kelas, sehingga analisis performa tidak hanya bergantung pada nilai akurasi semata [18].



Gambar 2. Visual Confusion Matrix NB dan SVM

Gambar 2 menyajikan perbandingan *Confusion Matrix* antara algoritma *Multinomial Naive Bayes* (a) dan *Support Vector Machine* (b) pada data uji sebanyak 1.688 komentar. Melalui visualisasi ini, dapat dilakukan analisis mendalam mengenai distribusi prediksi benar dan kesalahan klasifikasi yang dilakukan oleh kedua model:

1. Analisis Prediksi Benar (Diagonal Utama): Secara visual, kedua model menunjukkan dominasi angka pada diagonal kiri atas ke kanan bawah. Pada model SVM, tercatat 1.031 komentar non-judi dan 597 komentar judi berhasil diklasifikasikan dengan benar. Angka ini lebih tinggi dibandingkan Naive Bayes yang mengidentifikasi benar 1.026 komentar non-judi dan 579 komentar judi. Hal ini menegaskan bahwa SVM memiliki stabilitas yang lebih baik dalam mengenali kedua kelas secara seimbang.
2. Analisis Kesalahan Lolos (*False Negative*): Kesalahan paling krusial dalam deteksi konten ilegal adalah *False Negative* (kotak kiri bawah), yaitu komentar judi yang dianggap normal oleh sistem. Model SVM terbukti lebih unggul karena hanya meloloskan 47 komentar judi, sedangkan Naive Bayes meloloskan hingga 65 komentar. Penurunan angka *False Negative* sebesar 27,6% pada SVM menunjukkan bahwa algoritma ini lebih sensitif terhadap pola kata tersamar seperti "pulau" atau "gunungwin" yang digunakan oleh pelaku promosi.
3. Analisis Kesalahan Salah Tangkap (*False Positive*): Pada kotak kanan atas, model SVM meminimalkan kesalahan "salah tangkap" menjadi hanya 13 data, lebih rendah dibandingkan Naive Bayes (18 data). Hal ini sangat penting dalam implementasi praktis guna memastikan bahwa komentar pengguna asli tidak terhapus atau terblokir secara tidak sengaja oleh sistem moderasi otomatis.

Kesimpulan Visual: Perbedaan distribusi angka pada kedua matriks di atas menunjukkan bahwa meskipun kedua model memiliki performa tinggi, SVM dengan kernel linear lebih efektif dalam menarik garis pemisah (*hyperplane*) pada ruang fitur TF-IDF yang kompleks. Hal ini menjadikan SVM sebagai model yang lebih direkomendasikan untuk menangani karakteristik dataset komentar YouTube yang dinamis dan penuh dengan istilah eufemisme.

4. KESIMPULAN

Efektivitas Sistem Cerdas: Integrasi teknik pra-pemrosesan teks yang mendalam (Cleaning, Stopword, Stemming) dan ekstraksi fitur TF-IDF (5.000 fitur) terbukti sangat efektif dalam mendeteksi komentar judi online di platform YouTube. Perbandingan Algoritma: Model Support Vector Machine (SVM) memberikan performa terbaik dengan akurasi sebesar 96,45%, mengungguli Multinomial Naive Bayes yang memperoleh akurasi 95,08%. Hal ini menunjukkan bahwa SVM lebih unggul dalam memproses data teks berdimensi tinggi. Kemampuan Deteksi Eufemisme: Implementasi melalui Python berhasil menciptakan sistem komputer cerdas yang mampu mengenali pola kata sandi atau eufemisme seperti "pulau", "gunungwin", dan "uang jajan". Sistem terbukti sensitif dalam menangkap pola promosi ilegal meskipun pelaku menggunakan istilah samaran untuk menghindari filter platform. Keunggulan Teknis: Berdasarkan *Confusion Matrix*, SVM menunjukkan tingkat kesalahan terkecil (*False Negative* hanya 47 data), yang berarti sistem ini sangat andal untuk digunakan sebagai alat moderasi otomatis guna memitigasi penyebaran konten judi online.

REFERENSI

- [1] A. M. Nasution, M. Zarlis, and S. Suherman, "Analysis and implementation of Honeyd as a low-interaction honeypot in enhancing security systems," *Random International Social Science Journal*, vol. 2, no. 1, pp. 124–135, Jan. 2021, doi: 10.47175/rissj.v2i1.209.
- [2] A. Widiyanto, M. Prameswari, and M. A. Latief, "Gambling comments detection on YouTube: A comparative study of tree-based boosting, LSTM, and GRU models," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 23, no. 2, pp. 144–160, Jul. 2025, doi: 10.12962/j24068535.v23i2.a1305.
- [3] D. Diffran N. C., R. Handayani, V. B. Lestari, and S. Febriani, "Sentiment analysis of public opinion on online gambling through social media using convolutional neural network," *Journal of Informatics and Telecommunication Engineering*, vol. 9, no. 1, Jul. 2025, doi: 10.31289/jite.v9i1.15024.
- [4] M. Z. Ammar, R. E. Putra, and Y. Yamasari, "Deep learning-based detection of online gambling promotion spam in Indonesian YouTube comments," *Journal of Applied Informatics and Computing*, vol. 9, no. 6, pp. 3632–3641, Dec. 2025, doi: 10.30871/jaic.v9i6.11240.
- [5] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh text preprocessing terhadap analisis sentimen komentar masyarakat pada media sosial Twitter (studi kasus: Pilpres 2019)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 406–414, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [6] M. F. Madjid, D. E. Ratnawati, and B. Rahayudi, "Sentiment analysis on app reviews using support vector machine and naïve Bayes classification," *Sinkron*, vol. 7, no. 1, pp. 556–562, Feb. 2023, doi: 10.33395/sinkron.v8i1.12161.
- [7] A. Firda and M. Y. Harahap, "Optimalisasi sarana dan prasarana di MTS PAB 5 Klambir Lima untuk meningkatkan mutu pendidikan pada era digital," *Jurnal Ilmiah Al-Hadi*, vol. 11, no. 1, pp. 133–143, Oct. 2025, doi: 10.54248/alhadi.v11i1.4966.
- [8] D. R. Arrayyan, R. G. Guntara, and M. R. Nugraha, "Deteksi komentar spam judi online berbahasa Indonesia menggunakan XGBoost dan TF-IDF," *Jurnal Algoritma*, vol. 22, no. 2, pp. 2066–2075, Nov. 2025, doi: 10.33364/algoritma/v.22-2.3012.
- [9] M. Wahyudi and M. S. Novelan, "Perbandingan kinerja K-NN dan naïve Bayes dalam tata kelola teknologi informasi untuk menganalisa tingkat kepuasan pengguna aplikasi Shopee," *Journal of Science and Social Research*, vol. 8, no. 3, pp. 3424–3431, Aug. 2025, doi: 10.54314/jssr.v8i3.3547.
- [10] N. L. Ratniathi *et al.*, "Sentiment analysis of YouTube comments on Indonesian presidential election using SVM," *Journal of Physics: Conference Series*, vol. 1469, no. 1, p. 012101, 2020, doi: 10.1088/1742-6596/1469/1/012101.
- [11] V. Nurmaylina and Y. Akbar, "Sentiment analysis of social media X users towards legislators engaged in online gambling using naïve Bayes algorithm," *International Journal of Software Engineering and Computer Science*, vol. 4, no. 3, pp. 1128–1136, 2024, doi: 10.35870/ijsecs.v4i3.3079.

- [12] A. Irianti, H. Halimah, S. Sutedi, and M. Agariana, "Integration of BERT and SVM in sentiment analysis of Twitter/X regarding Constitutional Court Decision No. 60/PUU-XXII/2024," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 2, pp. 469–482, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.4068.
- [13] P. H. Putra and M. S. Novelan, "Perancangan aplikasi penentuan kualitas sayuran berdasarkan warna menggunakan data mining," in *Proc. Seminar of Social Sciences Engineering and Humaniora (SCENARIO)*, Univ. Pembangunan Panca Budi, Medan, Indonesia, Apr. 2022.
- [14] F. Fahrudin, "Komentar judi online dataset," *Kaggle*, 2023. [Online]. Available: <https://www.kaggle.com/datasets/fahruu/komentar-judi-online>. [Accessed: 25-Dec-2025].
- [15] D. Tehare, V. Mogadpalli, N. R. Gupta, S. Takalkar, and M. Patil, "Role of Python programming language in data science and big data processing," *International Journal of Technology & Emerging Research*, vol. 1, no. 3, pp. 5–18, Jul. 2025.
- [16] S. Octarina, F. M. Puspita, E. Yuliza, and I. Indrawati, "Pendampingan penggunaan Google Colab pada pembelajaran Python dan machine learning bagi dosen matematika di Palembang," *Jurnal Pepadu*, vol. 6, no. 1, pp. 56–66, 2025, doi: 10.29303/pepadu.v6i1.6457.
- [17] V. B. Lestari and C. A. P. Hutagalung, "Evaluation of TF-IDF extraction techniques in sentiment analysis of Indonesian-language marketplaces using SVM, logistic regression, and naïve Bayes," *J-KOMA: Journal of Computer Science and Applications*, vol. 8, no. 1, 2025, doi: 10.21009/j-koma.v8i1.05.
- [18] E. Helmud, F. Fitriyani, and P. Romadiana, "Classification comparison performance of supervised machine learning random forest and decision tree algorithms using confusion matrix," *Jurnal Sistem Informasi dan Komputer (SISFOKOM)*, vol. 13, no. 1, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1985.
- [19] M. G. Pradana, "Penggunaan fitur wordcloud dan document term matrix dalam text mining," *Jurnal Ilmiah Informatika (JIF)*, vol. 8, no. 1, 2025