

# Klasifikasi Komentar Promosi Judi Online pada YouTube Menggunakan TF-IDF dan Support Vector Machine

Rinaldi Fauzan\*, Reyhan Achmad Rizal, Delima Sitanggang

Sistem Informasi, Sains dan Teknologi, Universitas Prima Indonesia, Medan, Indonesia

Email: <sup>1\*</sup> rinaldifauzann13@gmail.com, <sup>2</sup> reyhanachmadrizal@unprimdn.ac.id<sup>3</sup> delimasitanggang@unprimdn.ac.id

Email Penulis Korespondensi: rinaldifauzann13@gmail.com\*

Submitted: 03/05/2026; Accepted: 04/06/2026; Published: 30/06/2026

**Abstrak**—Penelitian ini bertujuan mengklasifikasikan komentar promosi judi online pada platform YouTube menggunakan metode *Support Vector Machine* (SVM). Data penelitian berasal dari dataset publik Kaggle yang berisi komentar pengguna YouTube terkait promosi judi online. Tahapan penelitian meliputi *preprocessing teks*, ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), serta klasifikasi menggunakan SVM *kernel linear*. Hasil pengujian menunjukkan bahwa model SVM memperoleh akurasi sebesar 95,94% dengan nilai *precision*, *recall*, dan *F1-score* yang tinggi dan seimbang. Hasil analisis juga menunjukkan bahwa mayoritas tanggapan pengguna cenderung berada pada kelas non-promosi terhadap konten judi online. Selain itu, SVM menunjukkan performa yang lebih baik dibandingkan *Naïve Bayes* pada dataset yang digunakan. Temuan ini menunjukkan bahwa kombinasi TF-IDF dan SVM efektif digunakan untuk mengidentifikasi pola tanggapan pengguna terhadap konten judi online di YouTube.

**Kata Kunci:** SVM; TF-IDF; Text Mining; Klasifikasi; Judi Online; YouTube

**Abstract**—This study aims to classify online gambling promotional comments on the YouTube platform using the Support Vector Machine (SVM) method. The research data comes from a public Kaggle dataset containing YouTube user comments related to online gambling promotions. The research stages include text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), and classification using a linear kernel SVM. Test results show that the SVM model achieved an accuracy of 95.94% with high and balanced *precision*, *recall*, and *F1-score* values. The analysis also shows that the majority of user responses tend to be in the negative class regarding online gambling content. In addition, SVM performs better compared to *Naïve Bayes* on the dataset used. These findings indicate that the combination of TF-IDF and SVM is effective for identifying patterns in user responses to online gambling content on YouTube.

**Keywords:** SVM; TF-IDF; Text Mining; Classification; Online Gambling; YouTube

## 1. PENDAHULUAN

Maraknya perkembangan teknologi informasi dan media sosial telah mengubah cara masyarakat berkomunikasi dan mengakses informasi[1]. Platform media sosial seperti YouTube tidak hanya digunakan sebagai sarana interaksi sosial, tetapi juga dimanfaatkan sebagai media promosi berbagai aktivitas, termasuk konten judi online[2]. Konten tersebut muncul dalam berbagai bentuk, seperti iklan terselubung, postingan grup, tautan eksternal, hingga kolom komentar, sehingga sulit dikendalikan dan terus menjangkau pengguna dalam jumlah besar[3]. Di Indonesia, fenomena judi online mengalami peningkatan signifikan dalam beberapa tahun terakhir seiring dengan tingginya jumlah pengguna internet dan media sosial[4]. Meskipun aktivitas judi online dilarang secara hukum, praktik promosinya tetap masif dan adaptif, terutama di YouTube yang memiliki fitur berbagi konten dan interaksi terbuka[5]. Hal ini menjadikan YouTube sebagai salah satu media utama penyebaran konten judi online yang berpotensi memengaruhi perilaku pengguna.

Dampak dari maraknya konten judi online tidak hanya bersifat ekonomi, tetapi juga sosial dan psikologis[6]. Judi online dapat memicu kecanduan, menimbulkan kerugian finansial, konflik keluarga, hingga meningkatnya tindakan kriminal. Konten promosi yang terus muncul di linimasa YouTube berisiko menormalisasi praktik judi, terutama bagi pengguna yang rentan seperti remaja dan masyarakat dengan literasi digital yang rendah[7]. Tanggapan pengguna terhadap konten judi online di YouTube sangat beragam. Sebagian pengguna menunjukkan dukungan atau ketertarikan, sebagian lain menolak dan memberikan kritik, sementara tidak sedikit pula yang bersikap netral atau hanya berinteraksi tanpa memahami dampaknya. Pola tanggapan ini mencerminkan sikap, persepsi, dan tingkat kesadaran masyarakat terhadap bahaya judi online, yang penting untuk dianalisis secara sistematis[8].

Namun, hingga saat ini, tanggapan pengguna tersebut belum banyak dianalisis menggunakan pendekatan ilmiah berbasis data. Sebagian besar penelitian masih berfokus pada aspek hukum atau dampak sosial secara umum, tanpa menggali pola respons pengguna media sosial secara langsung. Padahal, komentar dan interaksi pengguna merupakan sumber data berharga untuk memahami bagaimana konten judi online diterima dan disikapi oleh masyarakat[9]. Apabila fenomena ini tidak ditanggapi secara serius, penyebaran konten judi online di YouTube berpotensi semakin meluas dan sulit dikendalikan. Kurangnya pemahaman terhadap pola respons pengguna dapat menghambat upaya pencegahan, edukasi digital, serta pengawasan konten oleh pihak terkait. Dalam jangka panjang, kondisi ini dapat memperparah dampak negatif judi online di tengah masyarakat Indonesia[10].

Beberapa penelitian sebelumnya telah menerapkan algoritma machine learning seperti *Naïve Bayes*, SVM, serta pendekatan Word2Vec dalam analisis sentimen media sosial[11]. Penelitian oleh Hermawan dkk. menggunakan

algoritma SVM pada data Twitter dan menunjukkan tingkat akurasi yang cukup baik. Penelitian Saputra dkk. meningkatkan akurasi analisis sentimen YouTube melalui normalisasi bahasa tidak baku. Selain itu, Angelo dkk. membandingkan beberapa algoritma machine learning dalam mendeteksi komentar promosi judi online di YouTube. Meskipun demikian, penelitian sebelumnya masih berfokus pada deteksi komentar atau analisis sentimen secara umum dan belum secara spesifik mengidentifikasi pola tanggapan pengguna terhadap konten judi online menggunakan kombinasi TF-IDF dan SVM pada dataset publik YouTube[12]. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan penelitian tersebut dengan menerapkan metode klasifikasi berbasis TF-IDF dan SVM guna mengidentifikasi pola tanggapan pengguna secara lebih objektif dan terukur.

Diperlukan metode yang mampu mengolah data dalam jumlah besar dan tidak terstruktur, seperti komentar pengguna YouTube. Pendekatan data mining dengan teknik *text mining* menjadi solusi yang tepat untuk mengekstraksi informasi penting dari data teks[13]. Salah satu metode yang efektif dalam mengelompokkan data berdasarkan kemiripan karakteristik adalah algoritma SVM. Dengan menerapkan algoritma SVM, penelitian ini diharapkan mampu mengidentifikasi pola tanggapan pengguna terhadap konten judi online secara objektif dan terukur. Temuan dari penelitian ini tidak sekadar memperkaya khazanah ilmiah dalam ranah data mining dan kajian media sosial, melainkan juga berpotensi menjadi landasan strategis dalam mendukung kegiatan pengawasan, peningkatan edukasi masyarakat, serta penanganan terhadap maraknya penyebaran praktik judi online di Indonesia.

## 2. METODOLOGI PENELITIAN

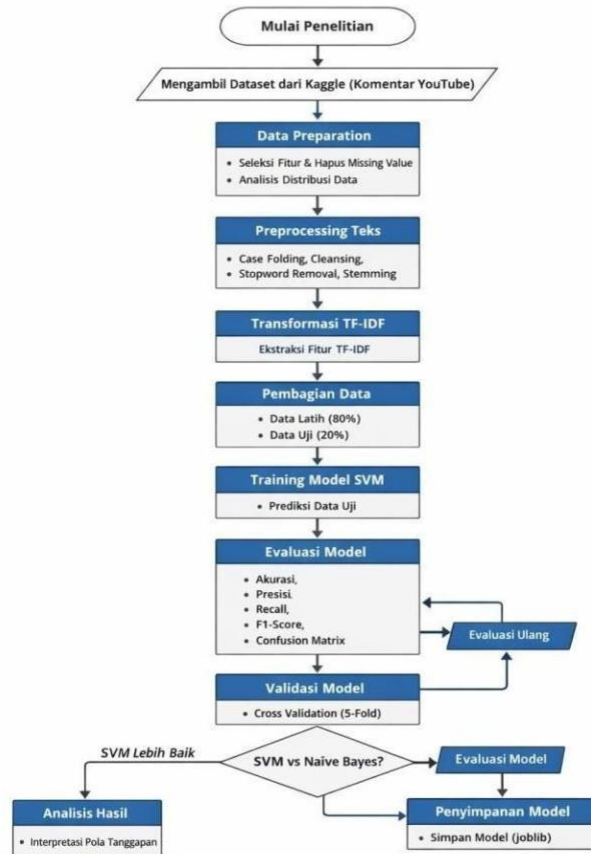
### 2.1 Alur Penelitian

Penelitian ini memiliki tahapan terstruktur yang mengacu pada kerangka kerja *text mining* dan klasifikasi dokumen. Secara garis besar, alur penelitian dimulai dari proses akuisisi data, di mana peneliti mengunduh *dataset* publik berisi komentar pengguna YouTube terkait promosi judi online dari platform Kaggle[14].

Tahap berikutnya adalah persiapan data (*data preparation*), yang mencakup proses seleksi fitur dengan memilih atribut komentar dan label sebagai variabel utama, penanganan nilai kosong, serta eksplorasi awal distribusi kelas melalui visualisasi untuk memahami keseimbangan data[15]. Setelah data siap, dilakukan tahap pemrosesan awal teks (*preprocessing*) yang meliputi *case folding*, pembersihan simbol dan karakter non-alfabet, penghapusan kata-kata umum (*stopword removal*), serta perubahan kata berimbuhan ke bentuk dasarnya (*stemming*) menggunakan pustaka Sastrawi guna menyeragamkan dan mereduksi dimensi data.

Selanjutnya, data yang telah bersih dibagi menjadi data latih (80%) dan data uji (20%) dengan metode stratified random sampling agar proporsi label tetap terjaga. Ekstraksi fitur dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dengan kombinasi unigram dan bigram untuk merepresentasikan teks dalam bentuk vektor numerik[16]. Pada tahap inti, algoritma *Support Vector Machine* (SVM) dengan kernel linear dilatih menggunakan data latih, kemudian diuji performanya pada data uji. Evaluasi model dilakukan dengan menghitung metrik akurasi, *precision*, *recall*, dan F1-score, serta disajikan dalam bentuk *confusion matrix*. Untuk memastikan konsistensi performa, dilakukan juga validasi silang (*cross-validation*) lima lipatan.

Tahap akhir dari alur penelitian ini adalah penyimpanan model terlatih menggunakan pustaka joblib agar dapat digunakan kembali, serta perbandingan kinerja SVM dengan algoritma *Naïve Bayes* sebagai pembanding. Keseluruhan alur ini dirancang untuk menjawab tujuan penelitian, yaitu mengidentifikasi pola tanggapan pengguna terhadap konten judi online di media sosial YouTube secara objektif dan terukur.



Gambar 1. Alur Penelitian

## 2.2 Jenis Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan analisis yang berorientasi pada pengolahan data untuk mempelajari pola respons pengguna terhadap konten judi online pada platform media sosial YouTube. Pemilihan pendekatan kuantitatif didasarkan pada karakteristik penelitian yang memanfaatkan data teks dalam jumlah besar yang perlu diolah secara komputasional. Komentar pengguna yang menjadi sumber data terlebih dahulu diproses menggunakan teknik *text mining* sehingga teks yang semula tidak terstruktur dapat diubah menjadi representasi numerik. Selanjutnya, data yang telah direpresentasikan dalam bentuk angka dianalisis menggunakan algoritma *Support Vector Machine* (SVM) guna melakukan proses klasifikasi terhadap tanggapan pengguna. Penerapan metode tersebut memungkinkan proses analisis dilakukan secara lebih objektif, terukur, dan sistematis sehingga mampu mendukung pencapaian tujuan penelitian yang telah ditetapkan[17].

## 2.3 Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh komentar pengguna YouTube yang terdapat pada dataset publik "Judi Online: Indonesian Online Gambling Promotion at YouTube" yang diperoleh dari platform Kaggle. Dataset tersebut berisi komentar pengguna yang berkaitan dengan konten promosi judi online dan digunakan sebagai sumber data utama dalam penelitian ini. sehingga relevan untuk dianalisis dalam konteks perilaku pengguna media sosial di Indonesia[18].

Teknik pengambilan sampel menggunakan Purposive Sampling, yaitu dengan memilih data komentar yang memenuhi kriteria tertentu guna memastikan data yang dianalisis relevan dan berkualitas. Kriteria sampel yang digunakan dalam penelitian ini meliputi:

- Komentar berasal dari konten YouTube yang secara eksplisit atau implisit mengandung unsur promosi judi online.
- Komentar berbentuk teks dan dapat diproses melalui tahapan *text mining* untuk keperluan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

## 2.4 Instrumen Penelitian

Penelitian ini menggunakan dataset publik yang diperoleh dari platform Kaggle sebagai instrumen penelitian. Dataset yang digunakan berjudul "Judi Online: Indonesian Online Gambling Promotion at YouTube" dan berisi kumpulan komentar pengguna YouTube yang telah tersusun dalam format terstruktur. Pendekatan ini sesuai dengan desain penelitian kuantitatif berbasis data mining, di mana data teks diolah dan dikonversi ke bentuk numerik melalui tahapan

*text mining* untuk dianalisis menggunakan algoritma *Support Vector Machine* (SVM)[19].

Deskripsi Instrumen:

- Instrumen penelitian berupa dataset komentar pengguna YouTube yang diperoleh dari platform Kaggle dan telah tersedia dalam format data terstruktur.
- Data komentar diproses melalui tahapan *preprocessing* teks, meliputi pembersihan data, tokenisasi, penghapusan stopword, dan pembobotan kata (TF-IDF), sehingga menghasilkan data numerik yang dapat digunakan dalam proses klasifikasi SVM.
- Instrumen ini dirancang dengan memperhatikan aspek etika penelitian, khususnya kerahasiaan data dan privasi pengguna, dengan hanya menggunakan data yang bersifat publik dan tidak menampilkan identitas individu dalam hasil analisis.

## 2.5 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan metode dokumentasi dengan memanfaatkan dataset publik yang tersedia pada platform Kaggle. Dataset yang digunakan adalah "Judi Online: Indonesian Online Gambling Promotion at YouTube" yang berisi komentar pengguna YouTube terkait konten promosi judi online.

Data diperoleh dengan mengunduh dataset dari Kaggle, kemudian dilakukan proses seleksi atribut yang relevan untuk kebutuhan penelitian. Atribut yang digunakan meliputi kolom komentar sebagai fitur utama dan kolom label sebagai target klasifikasi. Setelah proses seleksi data selesai, dataset dipersiapkan melalui tahapan pembersihan data, pemeriksaan nilai kosong, serta validasi struktur data sebelum memasuki tahap *preprocessing* dan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

## 2.6 Teknik Analisis Data

Teknik analisis data dalam penelitian ini dilakukan menggunakan Google Colab sebagai lingkungan komputasi berbasis cloud untuk mengolah dan menganalisis data komentar pengguna YouTube. Google Colab dipilih karena mendukung pemrograman Python serta menyediakan berbagai library *machine learning* yang diperlukan untuk penerapan metode *Support Vector Machine* (SVM) secara efisien dan terukur[20]. Proses analisis data dilakukan secara bertahap agar menghasilkan model klasifikasi yang akurat dan dapat dipertanggungjawabkan secara ilmiah dengan tahap sebagai berikut:

- Pengumpulan dan Input Data**  
Data berupa komentar pengguna YouTube pada konten promosi judi online dikumpulkan dan dimasukkan ke dalam Google Colab dalam format *dataset* yang terstruktur, seperti file CSV, untuk memudahkan proses pengolahan dan analisis data.
- Eksplorasi Data Awal**  
Dilakukan pemeriksaan awal terhadap *dataset* untuk mengetahui jumlah data, struktur komentar, serta mengidentifikasi data yang kosong, duplikat, atau tidak relevan. Tahap ini bertujuan untuk memastikan kualitas data sebelum dilakukan pengolahan lebih lanjut.
- Preprocessing* Data Teks**  
Data komentar dibersihkan melalui beberapa tahapan, meliputi penghapusan simbol, angka, dan karakter khusus, proses tokenisasi, penghapusan stopword, serta normalisasi kata. Tahap ini dilakukan untuk mengurangi noise dan meningkatkan kualitas data teks.
- Transformasi Data ke Bentuk Numerik**  
Data teks yang telah di-*preprocessing* diubah ke dalam bentuk numerik menggunakan metode pembobotan kata TF-IDF, sehingga setiap komentar direpresentasikan dalam bentuk vektor yang dapat diproses oleh algoritma SVM.
- Pelabelan Data**  
Dataset yang digunakan dalam penelitian ini merupakan dataset klasifikasi biner yang terdiri atas dua kelas, yaitu label 0 dan label 1. Label 0 merepresentasikan komentar yang tidak mengandung unsur promosi judi online (non-promosi), sedangkan label 1 merepresentasikan komentar yang mengandung unsur promosi atau dukungan terhadap aktivitas judi online.
- Pembagian Data Latih dan Data Uji**  
*Dataset* dibagi menjadi data latih dan data uji dengan proporsi tertentu untuk melatih model SVM dan menguji performa model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.
- Pembangunan dan Pelatihan Model SVM**  
Model *Support Vector Machine* dibangun dengan menentukan parameter dan jenis kernel yang sesuai, kemudian dilatih menggunakan data latih untuk menghasilkan model klasifikasi yang optimal.
- Pengujian Model**  
Model SVM yang telah dilatih digunakan untuk mengklasifikasikan data uji guna mengetahui kemampuan model dalam mengidentifikasi pola tanggapan pengguna terhadap konten judi online.
- Evaluasi Hasil Klasifikasi**

- Kinerja model dievaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* untuk menilai efektivitas metode SVM dalam penelitian ini.
- j. Analisis dan Interpretasi Hasil  
 Hasil klasifikasi dianalisis dan diinterpretasikan untuk mengidentifikasi pola tanggapan pengguna YouTube terhadap konten judi online sebagai dasar penarikan kesimpulan penelitian.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan dan Akuisisi Data

Data yang digunakan dalam penelitian ini bersumber dari *dataset* publik "Judi Online: Indonesian Online Gambling Promotion at YouTube" di platform Kaggle. *Dataset* ini hanya berisi satu file utama, yaitu `data_komentar_dengan_prediksi.csv`, yang mencakup data lengkap mulai dari identitas video YouTube hingga hasil prediksi akhir.

#### 3.2 Data Preparation

Proses ini dimulai dengan pemurnian data melalui seleksi fitur. Peneliti hanya memilih atribut komentar sebagai fitur utama dan atribut label sebagai target klasifikasi, sesuai fokus penelitian. Langkah ini memastikan model hanya memproses informasi esensial, tanpa terbebani data identitas video atau metadata lain yang tidak memengaruhi pola tanggapan.

Setelah kolom relevan terkumpul, missing values ditangani dengan menghapus baris yang mengandung nilai kosong. Pendekatan ini krusial dalam siklus pengembangan model *machine learning* untuk menjaga integritas dan kualitas *dataset*, sehingga menghindari kesalahan teknis atau bias saat pelatihan.

Setelah *dataset* bersih secara struktural, dilakukan analisis statistik deskriptif melalui visualisasi distribusi label dengan grafik batang. Analisis ini memetakan sebaran pola tanggapan pengguna, baik konten terkait judi online maupun komentar umum lainnya. Visualisasi memberikan gambaran awal tentang karakteristik data serta mendeteksi potensi ketidakseimbangan kelas, di mana satu label mungkin mendominasi yang lain

video_id	title	channel_name	tanggal	author	komentar	label	komentar_clean	predicted_label	
0	f8UEkmYXzZA	SKAKMAT AHOK	Pandji Pragiwaksono	1.745.412.234.123.970	TerranceNoelle-09i	Makin yakin abis baca review lain tentang 🙌🙌SG...	1	makin yakin abis baca review lain tentang SG188 .	0
1	Xi8K0_-kbHC	GAK NYANGKA BISA BEGINII PENGENDARA DIJALAN SA...	Jejelogy	1.739.601.493.342.000	deraatexplorerriders.2113	Paling suka model H2 🙌🙌	0	suka model h2	0
2	nZoNbiwP2ZE	Akhirnya Selesai Subaru Crosstrek Family Drift...	Garasi Drift	1.739.772.479.808.090	risqokurniadi7208	Mobilnya udah hancur 😞	0	mobilnya udah hancur	0
3	QpXckzQlnXg	Review Mobil Drift Seharga Super Car   BRZ V8 ...	Garasi Drift	1.738.825.556.100.780	LorrianeDotson	MANUT88: benar2 bikin aku jadi sultan	1	MANUT88: benar2 bikin sultan	1
4	nZoNbiwP2ZE	Akhirnya Selesai Subaru Crosstrek Family Drift...	Garasi Drift	1.739.858.865.953.920	Elpoco7365	Semoga lekas recover mobilnya mas Dipo	0	semoga lekas recover mobilnya mas dipo	0

Gambar 2. Data Cleansing

#### 3.3 Preprocessing Teks

Tahap *preprocessing* teks merupakan fase krusial untuk mengubah data komentar mentah yang tidak terstruktur menjadi format bersih dan standar, sehingga algoritma *Support Vector Machine* (SVM) dapat menginterpretasikannya secara akurat.

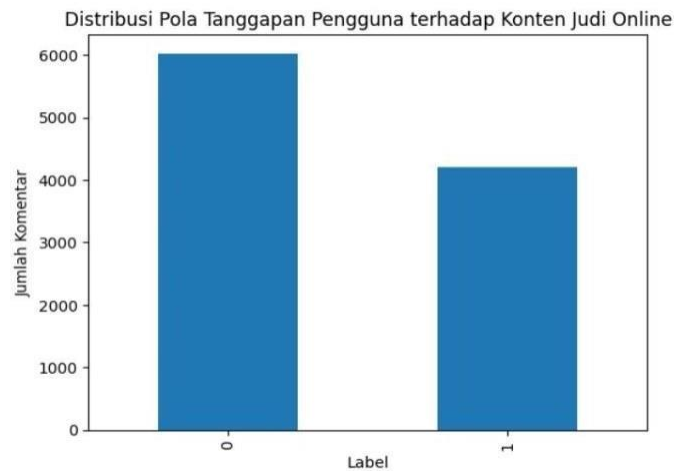
Proses dimulai dengan *case folding*, yaitu mengubah semua huruf menjadi lowercase untuk menyeragamkan data dan menghindari duplikasi akibat perbedaan kapitalisasi. Secara bersamaan, dilakukan *cleansing* dengan menghapus elemen non-informatif seperti URL, simbol, angka, karakter nonalfabet, serta spasi berlebih.

Selanjutnya, *stopword removal* menggunakan pustaka NLTK bahasa Indonesia untuk menghilangkan kata fungsional (seperti kata sambung atau depan) yang frekuensinya tinggi tapi tidak bermakna semantik dalam konteks judi online.

Tahap akhir adalah *stemming* dengan library Sastrawi, yang memetakan kata berimbuhan ke bentuk dasar. Ini meminimalkan variasi linguistik, mereduksi dimensi fitur, dan menyatukan kata dengan akar sama menjadi satu entitas. Hasilnya adalah teks bersih yang konsisten, representatif, dan siap untuk ekstraksi fitur numerik pada model klasifikasi.

index	komentar	clean_text
0	Makin yakin abis baca review lain tentang 🙌🙌SG188.	abis baca review
1	Paling suka model H2 🙌🙌	suka model h
2	Mobilnya udah hancur 😞	mobil udah hancur
3	MANUT88: benar2 bikin aku jadi sultan	bikin sultan
4	Semoga lekas recover mobilnya mas Dipo	moga lekas recover mobil mas dipo

**Gambar 3. Preprocessing Teks**



**Gambar 4. Preprocessing Data**

Gambar diatas adalah hasil dari proses *preprocessing* teks yang menunjukkan bahwa dua kategori tanggapan pengguna di Indonesia, dengan label "0" dan "1" pada sumbu X. Kategori "0" mendominasi sekitar 6000 tanggapan, jauh lebih tinggi dibanding "1" yang mencapai 4000, mengindikasikan pola mayoritas pengguna mungkin respons negatif terhadap konten judi online yang marak di media sosial.

### 3.4 Ekstraksi Fitur Pembagian Data

Setelah pembersihan, data memasuki fase pembagian *dataset* dan transformasi fitur. *Dataset* dibagi menjadi data latih (80%) dan data uji (20%) menggunakan *stratified random sampling*. Metode ini memastikan proporsi label (komentar judi online vs. netral) tetap konsisten di kedua bagian, mencegah bias, sehingga model belajar dari komposisi seimbang dan dievaluasi secara objektif. Tahap berikutnya adalah konversi teks ke numerik melalui *Term Frequency–Inverse Document Frequency* (TF-IDF). Teks bersih diubah menjadi vektor yang mencerminkan kepentingan kata dalam dokumen, karena algoritma *machine learning* hanya memproses data angka.

Ekstraksi fitur menggunakan kombinasi *unigram* (kata tunggal) dan *bigram* (dua kata berurutan), seperti "jackpot besar" atau "situs terpercaya". *Bigram* menangkap konteks frasa spesifik yang lebih bermakna daripada kata tunggal. TF-IDF memberi bobot tinggi pada kata/frasa unik yang jarang secara keseluruhan tapi sering pada label tertentu, sehingga meningkatkan presisi model SVM dalam membedakan pola tanggapan.

### 3.5 Implementasi Algoritma Support Vector Machine (SVM)

Setelah proses ekstraksi fitur selesai, model *Support Vector Machine* (SVM) dengan kernel linear dilatih menggunakan data latih. Pemilihan SVM didasarkan pada kemampuannya dalam menangani data teks berdimensi tinggi serta menghasilkan batas pemisah kelas yang optimal. Model yang telah dilatih kemudian digunakan untuk memprediksi label pada data uji. Kinerja model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan F1-score yang disajikan dalam *classification report*. Selain itu, *confusion matrix* divisualisasikan untuk memberikan gambaran detail mengenai jumlah prediksi yang benar dan salah pada setiap kelas.

### 3.6 Evaluasi dan Validasi Model

Untuk memastikan keandalan model, dilakukan evaluasi tambahan menggunakan teknik *cross-validation* sebanyak lima fold pada data latih. Metode ini bertujuan menilai kestabilan performa model pada beberapa bagian data yang berbeda sehingga hasil evaluasi tidak hanya bergantung pada satu proses pembagian data. Nilai rata-rata akurasi dari *cross-validation* digunakan sebagai indikator efektivitas metode SVM dalam mengidentifikasi pola tanggapan pengguna terhadap konten judi online

```

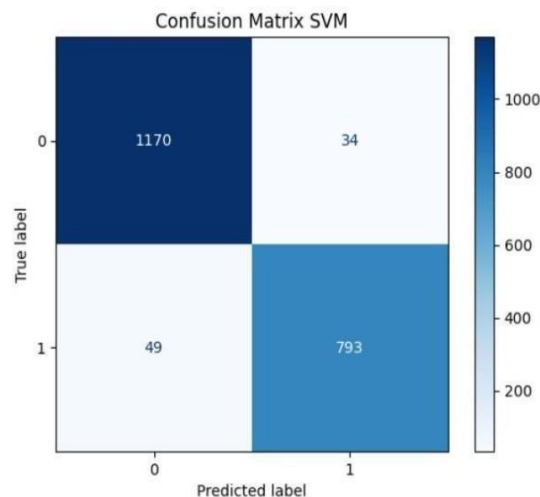
Akurasi Model: 0.9594530400782014
Classification Report:

```

	precision	recall	f1-score	support
0	0.96	0.97	0.97	1204
1	0.96	0.94	0.95	842
accuracy			0.96	2046
macro avg	0.96	0.96	0.96	2046
weighted avg	0.96	0.96	0.96	2046

**Gambar 5.** Hasil Classification Report

Berdasarkan hasil pengujian, gambar tersebut menunjukkan performa model yang sangat impresif. Algoritma *Support Vector Machine* (SVM) berhasil mengklasifikasikan data dengan tingkat akurasi mencapai 95,94%, yang menandakan bahwa model mampu membedakan pola tanggapan pengguna (antara kelas 0 dan 1) dengan sangat akurat. Nilai *precision* dan *recall* yang berada di angka 0,94 hingga 0,97 menunjukkan keseimbangan yang luar biasa dalam mendeteksi setiap label. Model tidak hanya mampu mengidentifikasi pola yang benar secara tepat, tetapi juga sangat minim dalam melakukan kesalahan klasifikasi. Hal ini diperkuat oleh nilai *F1-score* yang konsisten tinggi di kedua kelas, mengonfirmasi bahwa algoritma SVM memiliki ketahanan yang baik terhadap distribusi data yang ada, dengan total dukungan data uji sebanyak 2.046 sampel.



**Gambar 6.** Confusion Matrix

Visualisasi diatas menggambarkan performa detail dari model dalam mengklasifikasikan tanggapan pengguna, di mana model menunjukkan kemampuan prediksi yang sangat kuat pada kedua label yang diuji. Pada label 0, model berhasil memprediksi dengan benar sebanyak 1.170 data (True Negative) dan hanya mengalami kesalahan sebanyak 34 data, sementara pada label 1, model berhasil mengidentifikasi 793 data secara akurat (True Positive) dengan tingkat meleset yang rendah, yaitu hanya 49 data. Dominasi warna biru tua pada diagonal utama dari kiri atas ke kanan bawah secara visual menegaskan bahwa sebagian besar data telah terklasifikasi dengan tepat, yang membuktikan bahwa algoritma SVM memiliki tingkat presisi dan keandalan yang tinggi dalam membedakan pola-pola tanggapan dalam penelitian ini.

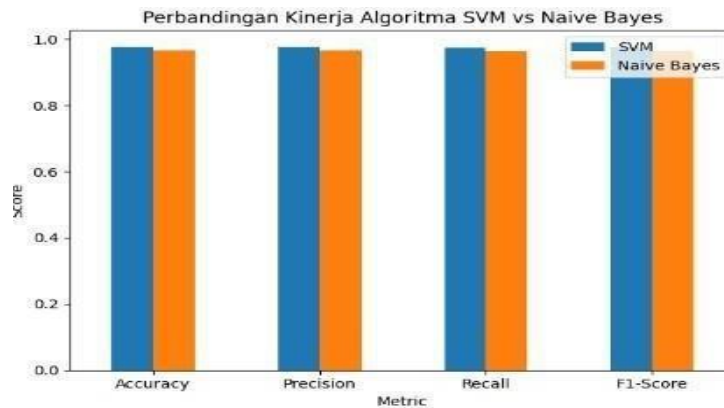
Visualisasi diatas menggambarkan performa detail dari model dalam mengklasifikasikan tanggapan pengguna, di mana model menunjukkan kemampuan prediksi yang sangat kuat pada kedua label yang diuji. Pada label 0, model berhasil memprediksi dengan benar sebanyak 1.170 data (True Negative) dan hanya mengalami kesalahan sebanyak 34 data, sementara pada label 1, model berhasil mengidentifikasi 793 data secara akurat (True Positive) dengan tingkat meleset yang rendah, yaitu hanya 49 data. Dominasi warna biru tua pada diagonal utama dari kiri atas ke kanan bawah secara visual menegaskan bahwa sebagian besar data telah terklasifikasi dengan tepat, yang membuktikan bahwa algoritma SVM memiliki tingkat presisi dan keandalan yang tinggi dalam membedakan pola-pola tanggapan dalam penelitian ini.

### 3.7 Perbandingan Dengan Algoritma *Naïve Bayes*

Berdasarkan hasil eksperimen perbandingan algoritma, model *Support Vector Machine* (SVM) memperoleh nilai akurasi sebesar 97.70%, sedangkan algoritma *Naïve Bayes* memperoleh akurasi sebesar 96.58%. Selain itu, nilai *precision*, *recall*, dan *F1-score* pada SVM juga lebih tinggi dibandingkan *Naïve Bayes*. Hasil ini menunjukkan bahwa algoritma SVM memiliki kemampuan yang lebih baik dalam mengklasifikasikan komentar pada *dataset* penelitian ini karena mampu memisahkan data dalam ruang fitur TF-IDF yang berdimensi tinggi secara lebih optimal.

Perlu diperhatikan bahwa nilai akurasi sebesar 95,94% yang diperoleh pada evaluasi dan validasi model merupakan hasil evaluasi menggunakan metode *cross-validation* lima fold untuk mengukur kestabilan model SVM. Sementara itu, nilai akurasi sebesar 97,70% pada perbandingan ini diperoleh dari eksperimen perbandingan kinerja algoritma SVM dan *Naïve Bayes* pada skenario pengujian yang berbeda. Oleh karena itu, perbedaan nilai akurasi tersebut disebabkan oleh perbedaan metode evaluasi yang digunakan.

Ketika *dataset* tersebut dijalankan dengan menggunakan dua algoritma yang berbeda maka menghasilkan data analisis perbandingan sebagai berikut:



**Gambar 7.** Perbandingan Kinerja Algoritma SVM dan *Naive Bayes*

**Tabel 1.** Hasil Perbandingan Kinerja

Algoritma	Accuracy	Precision	Recall	F1-Score
SVM	0.9770	0.9776	0.9749	0.9762
<i>Naive Bayes</i>	0.9658	0.9659	0.9634	0.9646

Berdasarkan hasil pengujian menggunakan *dataset* yang sama, diperoleh bahwa algoritma *Support Vector Machine* (SVM) memiliki kinerja yang sedikit lebih unggul dibandingkan dengan algoritma *Naive Bayes* dalam melakukan klasifikasi data komentar. Hal ini terlihat dari nilai *accuracy* yang diperoleh SVM sebesar 97,70%, sedangkan *Naive Bayes* memperoleh 96,58%. Selain itu, nilai *precision* pada SVM mencapai 97,76%, lebih tinggi dibandingkan *Naive Bayes* yang memperoleh 96,59%. Pada metrik *recall*, SVM juga menunjukkan performa yang lebih baik dengan nilai 97,49%, sementara *Naive Bayes* memperoleh 96,34%. Begitu pula pada nilai F1-score, SVM mencapai 97,62%, sedangkan *Naive Bayes* memperoleh 96,46%. Perbedaan nilai pada setiap metrik evaluasi tersebut membuktikan bahwa algoritma (SVM) memiliki kemampuan yang lebih baik dalam mempelajari pola dari fitur teks yang dihasilkan oleh proses TF-IDF, sehingga mampu melakukan klasifikasi komentar dengan tingkat akurasi dan konsistensi yang lebih tinggi dibandingkan algoritma *Naive Bayes* pada *dataset* penelitian ini.

### 3.8 Deployment dan penyimpanan Model

Pada tahap akhir, program menyediakan fungsi prediksi untuk menguji komentar baru yang belum pernah dilihat oleh model. Model SVM dan TF-IDF vectorizer kemudian disimpan ke dalam file menggunakan library *joblib*. Penyimpanan ini memungkinkan model digunakan kembali tanpa perlu melakukan proses pelatihan ulang, sehingga dapat dimanfaatkan untuk pengujian lanjutan atau pengembangan sistem analisis komentar di masa mendatang.

### 3.9 Analisis Error

Berdasarkan *confusion matrix* pada Gambar 6, model SVM masih menghasilkan kesalahan klasifikasi sebesar 34 data pada kelas 0 (false positive) dan 49 data pada kelas 1 (false negative). Meskipun jumlah ini relatif kecil dibandingkan total data uji (2.046 sampel), keberadaan error tersebut mengindikasikan adanya keterbatasan model dalam membedakan komentar yang memiliki kemiripan fitur kebahasaan antar kedua kelas. Untuk mengidentifikasi penyebab spesifik kesalahan, diperlukan penelusuran lebih lanjut terhadap sampel komentar yang salah klasifikasi guna memahami pola bahasa, istilah ambigu, atau konteks implisit yang tidak dapat ditangkap secara optimal oleh ekstraksi fitur TFIDF dan kernel linear SVM.

## 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa tanggapan pengguna terhadap konten judi online di platform YouTube menunjukkan variasi yang cukup beragam, namun secara umum didominasi oleh tanggapan bernada negatif. Hal ini mengindikasikan adanya kesadaran sebagian pengguna terhadap dampak buruk yang ditimbulkan oleh konten perjudian daring. Penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) untuk melakukan proses klasifikasi komentar pengguna melalui tahapan text mining yang meliputi preprocessing teks, pembobotan kata menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), serta proses klasifikasi menggunakan model SVM. Hasil pengujian menunjukkan bahwa model yang dibangun memiliki performa

yang sangat baik dengan tingkat akurasi di atas 95%, serta menghasilkan nilai *precision*, *recall*, dan *F1-score* yang tinggi dan relatif seimbang. Selain itu, hasil perbandingan dengan algoritma *Naïve Bayes* menunjukkan bahwa SVM memiliki performa yang lebih optimal dalam mengolah data teks berdimensi tinggi pada dataset penelitian ini. Dengan demikian, metode yang digunakan dinilai cukup andal dalam mengklasifikasikan tanggapan pengguna di media sosial serta berpotensi dimanfaatkan sebagai dasar dalam kegiatan pemantauan maupun pengendalian penyebaran konten judi online di ruang digital. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada penggunaan dataset yang berasal dari satu platform media sosial dan klasifikasi yang dilakukan masih terbatas pada kategori tertentu. Oleh karena itu, penelitian selanjutnya dapat dikembangkan dengan memanfaatkan dataset yang lebih besar dan lebih beragam dari berbagai platform media sosial, menerapkan metode yang lebih kompleks seperti *deep learning*, meningkatkan proses *preprocessing* untuk menangani bahasa tidak baku dan emoji, serta mengembangkan klasifikasi multikelas maupun sistem analisis secara *real-time* agar hasil penelitian dapat dimanfaatkan secara lebih luas.

## REFERENCES

- [1] Y. Handrianto, O. Pahlevi, and M. H. Winarsa, "Pemanfaatan Algoritma Support Vector Machine dalam Analisis Sentimen Ulasan Pengguna Aplikasi POLRI Presisi," vol. 7, no. 2, 2025, doi: doi.org/10.31294/infotech.v7i2.11378.
- [2] K. A. Fathor, F. P. Gani, and M. Z. Saleh, "Khresna Adityo Fathor, Fauzan Putra Gani, and Mohamad Zein Saleh, 'Fenomena Iklan Judi Online Pada Platform Digital Generasi Z Di Indonesia', optimal, vol. 4, no. 1, pp. 184–189, Mar. 2024.," vol. 4, no. 1, 2024.
- [3] M. X. Tf-idf, "D. R. Arrayyan, R. G. . Guntara, dan M. R. . Nugraha, 'Deteksi Komentar Spam Judi Online Berbahasa Indonesia Menggunakan XGBoost dan TF-IDF', Jurnal Algoritma, vol. 22, no. 2, hlm. 2066–2075, Nov 2025.," pp. 2066–2075, 2025, doi: 10.33364/algoritma/v.22-2.3012.
- [4] D. A. Zulfikar *et al.*, "D. A. Zulfikar, Sufianto Ahmad Arrafi, Elvin Mauladan, Ramdhan Shurai Hafiz, Syabid Al Basith, and Wahyudya Triwidodo, 'ANALISIS SENTIMEN PUBLIK TERHADAP PENANGGULANGAN JUDI ONLINE MENGGUNAKAN METODE NAIVE BAYES DAN SVM DI INDONESIA', infotech, vol. 11, n," vol. 11, no. 2, pp. 480–488, 2025, doi: doi.org/10.31949/infotech.v11i2.16795.
- [5] J. Friska *et al.*, "Jumita Friska, Adita Hervani Br Barus, Alfianti Alfianti, Dianrani Anastasya Purba, Ika Novita Padang, and Jenni Romayanti Ginting, 'Pola Perilaku Pengguna Situs Judi Online dan Faktor-Faktor yang Mempengaruhi', risoma, vol. 2, no. 6, pp. 282–296, Nov. 20," no. 2016, pp. 282–296, 2024.
- [6] H. Sosial, "A. K. Sari, M. Al-Fajrih, and I. Ahdiyanti, 'Dampak Judi Online Terhadap Kesehatan Mental Dan Hubungan Sosial', Al-Hikmah, vol. 3, no. 2, pp. 31–44, Jun. 2023.," vol. 3, no. 2, pp. 31–44, 2024.
- [7] A. D. Prastiko and A. D. Wiranata, "A. Dwi Prastiko dan A. . Davy Wiranata, 'Analisis Sentimen Publik terhadap Fenomena Judi Online di Media Sosial X dengan SVM', Jurnal Pustaka AI, vol. 5, no. 2, hlm. 306–315, Agu 2025.," pp. 5–8, 2025.
- [8] A. R. Putra and A. Hendrawan, "ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE TERKAIT JUDI ONLINE MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK ( CNN )," vol. 6, no. 3, pp. 1227–1235, 2025, doi: 10.46576/djtechno.
- [9] F. Ilmu, I. Politik, U. Diponegoro, J. A. Suryo, and S. K. Pos, "M. Yusuf Akbar, T. Pradekso, and N. Surayya Ulfa, 'PENGARUH TERPAAN IKLAN JUDI ONLINE DI MEDIA SOSIAL, TINGKAT PENGAWASAN ORANG TUA, DAN INTENSITAS KOMUNIKASI PEER GROUP TERHADAP MINAT BERMAIN JUDI ONLINE,' Interaksi Online, vol. 12, no. 4, pp. 120-129, A".
- [10] N. Simanjuntak and A. H. Muhammad, "BULLETIN OF COMPUTER SCIENCE RESEARCH Analisis Perbandingan Algoritma SVM dan CNN dalam Mendeteksi Website Judi Online Berdasarkan Konten Teks," vol. 5, no. 4, pp. 361–371, 2025, doi: 10.47065/bulletincsr.v5i4.586.
- [11] O. I. Gifari, M. Adha, I. R. Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," vol. 2, no. 1, pp. 36–40, 2022.
- [12] M. Nur, R. Nidhi, A. A. Arifiyanti, and D. S. Yudha, "M. Nur Rachman Nidhi Suryono, Amalia Anjani Arifiyanti, and Dhian Satria Yudha Kartika, 'ASPECT-BASED SENTIMENT ANALYSIS PADA ULASAN APLIKASI ACCESS BY KAI MENGGUNAKAN METODE TF-IDF DAN ALGORITMA SUPPORT VECTOR MACHINE', INSTEK, vol. 10, no. 2," vol. 10, no. 2, pp. 414–423, 2025.
- [13] I. Mursidah, R. Sanjaya, B. Yulianto, D. Sweetania, and P. Sularsih, "Analisis Sentimen Pengguna pada Ulasan Game Honkai Star Rail Menggunakan Algoritma Support Vector Machine ( SVM )," vol. 14, pp. 1178–1188, 2025.
- [14] M. F. A. Shidiq and D. Alita, "KASUS JUDI ONLINE MENGGUNAKAN DATA DARI MEDIA SOSIAL X PENDEKATAN NAIVE BAYES DAN SVM," vol. 8, no. 1, pp. 24–35, 2025.
- [15] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.
- [16] M. Das, S. Kamalanathan, and P. Alphonse, "A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset," vol. 5571, pp. 0–2, 2021.
- [17] F. N. Hasan, T. Informatika, U. Muhammadiyah, and P. Hamka, "ANALISIS SENTIMEN TANGGAPAN PENGGUNA APLIKASI BALE BY BTN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE ( SVM )," vol. 4, no. 4, pp. 315–326, 2025.
- [18] M. Angelo, J. Hendrik, and A. History, "Jurnal Teknologi dan Manajemen Informatika Comparative Performance of Machine Learning Algorithms for Detecting Online Gambling Promotional Comments on Youtube Article Info ABSTRACT," vol. 11, no. 2, pp. 132–142, 2025.



- [19] A. Maulana, A. Yuliana, T. Bandung, J. Politeknik, J. Pesantren, and K. Cimahi, "ONLINE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MECHINE," vol. 12, no. 3, pp. 3706–3714, 2024.
- [20] M. H. Ali, J. Ismail, and A. Info, "Jurnal JPILKOM ( Jurnal Penelitian Ilmu Komputer ) Analisis Sentimen Aplikasi Livin dengan Algoritma Neural Network , Random Forest , dan SVM," vol. 3, no. 1, 2025.