

# ANALISIS SENTIMEN TERHADAP KAMPANYE ANTI-JUDI ONLINE MENGUNAKAN SUPPORT VECTOR MACHINE

<sup>1</sup>Christina Endah Karunia Tangka, <sup>2</sup>Luther A. Latumakulita\*, <sup>3</sup>Siska Ayu Widiana, <sup>4</sup>Christie E. J. C. Montolalu, <sup>5</sup>Eliasta Ketaren, <sup>6</sup>Wisard Kalengkongan, <sup>7</sup>Rillya Arundaa

<sup>1234567</sup>Program Studi Sistem Informasi, Universitas Sam Ratulangi

<sup>1</sup>[christinatangka04@gmail.com](mailto:christinatangka04@gmail.com), <sup>2</sup>[latumakulitala@unsrat.ac.id](mailto:latumakulitala@unsrat.ac.id)\*, <sup>3</sup>[siskaginting@unsrat.ac.id](mailto:siskaginting@unsrat.ac.id),  
<sup>4</sup>[christelly@unsrat.ac.id](mailto:christelly@unsrat.ac.id), <sup>5</sup>[eliasketaren@unsrat.ac.id](mailto:eliasketaren@unsrat.ac.id), <sup>6</sup>[wisard.kalengkongan@unsrat.ac.id](mailto:wisard.kalengkongan@unsrat.ac.id), <sup>7</sup>[rill@unsrat.ac.id](mailto:rill@unsrat.ac.id)

## Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kampanye anti-judi online di TikTok menggunakan pendekatan TF-IDF dan algoritma Support Vector Machine (SVM). Kumpulan data ini terdiri dari komentar pengguna yang dikumpulkan dari TikTok, yang diproses melalui beberapa tahap persiapan seperti pembersihan teks, tokenisasi, penghapusan *stopword*, dan *stemming*. Untuk melabeli data, digunakan pendekatan hibrida yang mengintegrasikan teknik berbasis aturan dengan model IndoBERTweet yang telah dilatih sebelumnya. Fitur teks kemudian diekstraksi melalui TF-IDF dan dimasukkan ke dalam klasifikasi SVM kernel linier. Model tersebut menunjukkan akurasi keseluruhan sebesar 81,7% dalam klasifikasi sentimen. Distribusi sentimen menunjukkan bahwa 72,7% komentar diklasifikasikan sebagai negatif dan 27,3% sebagai positif. Model menunjukkan kinerja yang lebih baik pada sampel negatif dibandingkan dengan yang positif, perbedaan ini kemungkinan disebabkan oleh ketidakseimbangan kelas dan bahasa informal serta beragam yang terdapat dalam komentar. Secara keseluruhan, penelitian ini memberikan gambaran tentang bagaimana sentimen publik tersebar seputar kampanye anti-judi online di media sosial TikTok, dan menunjukkan bahwa menggabungkan pendekatan berbasis aturan dan model yang telah dilatih sebelumnya merupakan strategi yang layak untuk melabeli teks pendek dan informal dalam tugas analisis sentimen.

**Kata kunci:** SVM, Kampanye anti-judi online, Hybrid labeling, Analisis sentimen

## Abstract

*This study aims to analyze public sentiment toward anti-online gambling campaigns on TikTok using the TF-IDF approach and the Support Vector Machine (SVM) algorithm. The dataset consists of user comments collected from TikTok, which were processed through several preparation stages, including text cleaning, tokenization, stopword removal, and stemming. To label the data, a hybrid approach was used that integrates rule-based techniques with a pre-trained IndoBERTweet model. Text features were then extracted via TF-IDF and fed into a linear kernel SVM classifier. The model demonstrated an overall accuracy of 81.7% in sentiment classification. The sentiment distribution showed that 72.7% of comments were classified as negative and 27.3% as positive. The model performed better on the negative sample compared to the positive one; this difference is likely due to class imbalance and the informal and varied language present in the comments. Overall, this study provides an overview of how public sentiment is distributed regarding the anti-online gambling campaign on the social media platform TikTok, and demonstrates that combining rule-based approaches with pre-trained models is a viable strategy for labeling short and informal text in sentiment analysis tasks.*

**Keywords:** SVM, Anti-online gambling campaign, Hybrid labeling, Sentiment Analysis

## PENDAHULUAN

Media sosial kini menjadi bagian integral dalam kehidupan masyarakat modern karena dinilai cepat, mudah, dan efektif dalam menyebarkan maupun mencari informasi. Pemerintah, organisasi masyarakat, hingga lembaga swasta memanfaatkan platform seperti Instagram, Facebook, dan Tiktok untuk menyampaikan pesan publik secara luas dan interaktif (Munfida et al., 2023). Namun, kemudahan tersebut juga membuka peluang berkembangnya aktivitas negatif, salah satunya praktik perjudian online. Judi online berdampak serius terhadap kesehatan mental, seperti memicu kecanduan, depresi, kecemasan, hingga terganggunya hubungan sosial (Sari et al., 2024). Laporan PPAK (2024), mencatat perputaran dana judi online di Indonesia mencapai Rp 359,8 triliun, dengan mayoritas pemain berasal dari kelompok berpenghasilan rendah, sehingga menimbulkan masalah sosial yang signifikan.

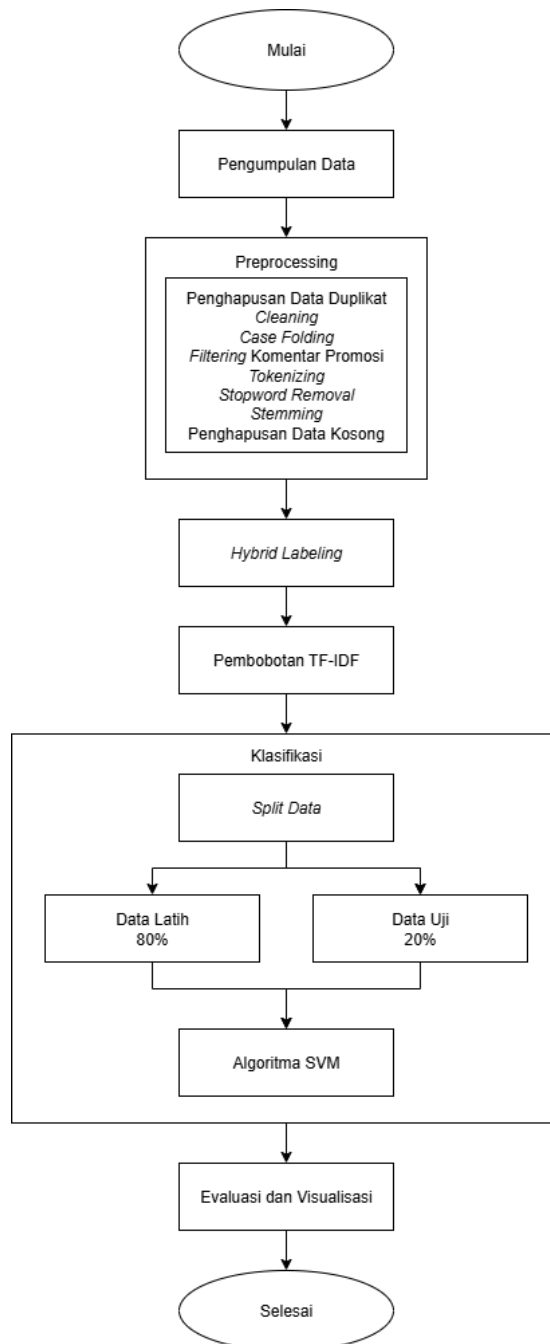
Meskipun pemerintah Indonesia telah berupaya melalui Kementerian Komunikasi dan Informatika (Komdigi) untuk memblokir ribuan situs perjudian daring, platform-platform baru terus bermunculan (Komdigi RI, 2025). Tantangan yang terus berlanjut ini menegaskan pentingnya strategi kampanye digital di media sosial sebagai langkah pelengkap. Seiring dengan meningkatnya penggunaan media sosial, kampanye digital telah menjadi saluran yang semakin relevan untuk menyebarkan pesan pencegahan dan membangun kesadaran publik tentang bahaya perjudian online (Howse et al., 2022). Kampanye yang diluncurkan Komdigi, Kampanye #LawanJudol di TikTok menyampaikan pesan anti-judi melalui berbagai format, termasuk testimoni korban, edukasi dampak negatif, dan ajakan berhenti bermain judi online (Dinas Kominfo Sanggau, 2025). Di luar program resmi tersebut, sejumlah konten kreator TikTok independen juga berperan dengan memproduksi video berlabel #StopJudol, yang mendorong audiens untuk menghindari perjudian online, sehingga memperluas jangkauan kampanye. TikTok dipilih sebagai platform untuk penelitian ini mengingat penggunaannya yang luas di Indonesia, di mana 73,5% populasi merupakan pengguna aktif berdasarkan data We Are Social 2024. Selain memiliki basis pengguna yang besar, algoritma TikTok memungkinkan konten tersebar secara luas melalui fitur *For You Page* (FYP), sehingga pesan kampanye dapat menjangkau pengguna di luar jaringan pengikut akun pembuat konten. Penelitian sebelumnya menunjukkan bahwa kampanye sosial yang disebarkan melalui TikTok efektif menjangkau audiens luas dalam waktu singkat melalui pemanfaatan tren, hashtag, dan narasi visual yang menarik (Perdana & Sulastri, 2025). Selain itu, penelitian lain menunjukkan bahwa TikTok efektif dalam meningkatkan jangkauan pesan digital dan menjangkau kelompok usia muda yang menjadi salah satu kelompok pengguna media sosial terbesar (Pratama et al., 2023). Karakteristik tersebut menjadikan TikTok relevan sebagai media kampanye digital untuk meningkatkan kesadaran masyarakat terhadap bahaya perjudian online.

Sejumlah studi terdahulu telah menganalisis sentimen terkait judi online menggunakan algoritma Naïve Bayes, SVM, dan Random Forest, dengan hasil menunjukkan keunggulan SVM dalam akurasi klasifikasi (Maulana & Yuliana, 2024; Syahputra & Wibowo, 2023; As Shidiq & Alita, 2025). Penelitian lain menunjukkan bahwa IndoBERTweet secara konsisten mengungguli algoritma tradisional seperti SVM dalam analisis sentimen teks berbahasa Indonesia, dengan akurasi dan F1-score lebih tinggi serta hasil signifikan secara statistik (Uliniansyah et al., 2025). Namun, sebagian besar penelitian sebelumnya masih berfokus pada platform Twitter/X yang memiliki karakteristik teks berbeda dengan TikTok, khususnya dari segi panjang komentar, gaya bahasa, dan tingkat informalitas. Selain itu, penelitian terkait sentimen kampanye anti-judi online di TikTok masih sangat terbatas, sehingga belum memberikan gambaran yang komprehensif mengenai persepsi publik pada platform tersebut.

Studi ini menghadirkan kontribusi baru melalui strategi pelabelan *hybrid* yang mengintegrasikan IndoBERTweet dengan rule-based keyword. Strategi ini memungkinkan deteksi sentimen yang lebih akurat pada komentar TikTok, khususnya dalam menangkap ekspresi lokal terkait judi online. Dengan dukungan algoritma SVM sebagai model klasifikasi, pendekatan ini diharapkan mampu meningkatkan kemampuan model dalam mendeteksi sentimen positif, yang selama ini menjadi tantangan dalam analisis sentimen kampanye sosial. Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan menawarkan kombinasi IndoBERTweet dan rule-based keyword sebagai strategi labeling yang lebih adaptif. Selain kontribusi metodologis, temuan ini juga memiliki implikasi praktis, karena dapat memberikan informasi kepada badan pemerintah dan komunitas digital yang lebih luas dalam menyusun strategi kampanye sosial yang lebih terarah dan berdampak, terutama yang bertujuan untuk meningkatkan kesadaran publik tentang risiko yang terkait dengan perjudian online.

## METODE

Penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis. Alur lengkap metodologi penelitian disajikan dalam bentuk flowchart yang ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Tahapan Penelitian

### **Pengumpulan Data**

Kumpulan data yang digunakan dalam penelitian ini diperoleh dari TikTok pada bulan September 2025, yang dikumpulkan melalui web scraping menggunakan alat *scrapping* otomatis Apify, khususnya Tiktok Comments Scraper, yang mengambil data komentar dari video TikTok mana pun menggunakan URL-nya. Data dikumpulkan dari video-video yang diberi tagar #StopJudol dan #LawanJudol, yang berfungsi sebagai sumber data utama untuk penelitian ini.

### **Preprocessing Data**

Pada tahap ini dilakukan preprocessing untuk mempersiapkan data dan meningkatkan kualitas data agar lebih terstruktur, bersih dari noise, serta siap digunakan dalam analisis sentimen. Tahapan *preprocessing data*

dalam penelitian ini meliputi penghapusan data duplikat, *cleaning*, *filtering* komentar promosi, tokenisasi, *stopword removal*, *stemming*, dan penghapusan data kosong.

- Penghapusan data duplikat yaitu menghapus komentar yang sama agar tidak terjadi pengulangan data yang dapat menimbulkan bias dalam analisis.
- *Cleaning* yaitu tahapan membersihkan teks dari elemen yang tidak relevan seperti emoji, URL, tanda baca berlebihan, mention, dan simbol yang tidak bermakna.
- *Case folding* yaitu mengubah seluruh teks menjadi huruf kecil (*lowercase*).
- *Filtering* komentar promosi yaitu menghapus komentar yang bersifat promosi atau iklan, khususnya yang mengandung kata kunci terkait perjudian seperti "slot", "bandar", atau pola angka mencurigakan, karena tidak merepresentasikan opini pengguna secara alami.
- *Tokenisasi* yaitu proses memecah teks menjadi unit kata (token) untuk memudahkan proses analisis selanjutnya.
- *Stopword removal* yaitu proses menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "yang", "di", "itu" dan sebagainya.
- *Stemming* yaitu proses mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan awalan maupun akhiran.
- Penghapusan data kosong yaitu menghapus beberapa data yang menjadi kosong atau tidak memiliki isi, setelah tahapan *cleaning* dan *filtering* dilakukan.

### **Hybrid Labeling**

Dalam tahapan ini hybrid labeling yang dilakukan merupakan gabungan dari IndoBERTtweet dan aturan *rule-based* berbasis *keyword*. Pendekatan *hybrid labeling* muncul sebagai solusi atas keterbatasan metode tunggal yaitu *rule-based* cenderung memiliki presisi tinggi pada pola eksplisit, tetapi lemah dalam menangkap konteks, sedangkan model pra-latih dapat memahami variasi bahasa namun berpotensi menghasilkan bias kelas atau distribusi label yang tidak merata (Rizkia et al., 2025).

Beberapa penelitian sebelumnya memperkuat efektivitas strategi hybrid, meskipun menggunakan kombinasi metode yang berbeda dengan penelitian ini. Penelitian (Lin & Nuha, 2023) misalnya, mengusulkan strategi *hybrid deep learning* untuk analisis sentimen teks pendek, dan hasilnya menunjukkan bahwa kombinasi metode memberikan performa lebih baik dibanding penggunaan metode tunggal. Model IndoBERTtweet dipilih karena dilatih secara khusus menggunakan data Twitter berbahasa Indonesia, sehingga lebih adaptif terhadap karakteristik teks media sosial yang singkat, informal, dan tidak baku (Koto et al., 2021). Kombinasinya dengan *rule-based labeling* berbasis kata kunci memungkinkan penanganan komentar yang memiliki indikator eksplisit, seperti promosi judi atau dukungan terhadap kampanye anti-judi.

Penelitian ini menggunakan dua kelas sentimen, yaitu positif, dan negatif. Komentar yang termasuk label positif merupakan komentar yang menunjukkan dukungan terhadap kampanye anti judi online, seperti ajakan untuk berhenti judi, penolakan terhadap praktik judi, serta pengalaman negatif yang mendorong seseorang untuk menjauhi judi. Komentar yang termasuk label negatif merupakan komentar yang tidak mendukung kampanye anti judi, termasuk komentar yang bersifat promosi judi, netral, atau tidak menunjukkan sikap penolakan terhadap judi online

Sebagaimana ditegaskan oleh (Aftab et al., 2023), analisis sentimen kerap dilakukan dalam bentuk klasifikasi biner, khususnya pada domain seperti ulasan film maupun produk, di mana dataset secara alami terstruktur ke dalam label positif dan negatif. Pendekatan ini tidak hanya menyederhanakan proses analisis, tetapi juga selaras dengan ketersediaan data serta membantu mengatasi tantangan praktis dalam pemodelan sentimen. Oleh karena itu, penelitian ini menggunakan kerangka klasifikasi biner dengan fokus pada dua label utama yaitu positif dan negatif.

### **Pembobotan TF-IDF**

Komentar yang sudah melalui tahap *cleaning* dan *labeling* kemudian dikonversi menjadi representasi vektor menggunakan TF-IDF. Pembobotan Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk mengidentifikasi kata-kata yang paling penting dalam suatu dokumen atau kumpulan dokumen. Menurut (Ramos, 2003), TF-IDF merupakan metode yang efektif untuk menentukan relevansi kata dalam teks, dengan menggabungkan frekuensi kemunculan kata di dalam dokumen dan distribusi kata tersebut di seluruh korpus. Term Frequency (TF) adalah nilai frekuensi kemunculan suatu kata dalam sebuah dokumen. Nilai TF dapat dihitung dengan menggunakan rumus:

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{jumlah kata dalam dokumen}} \quad (1)$$

Namun, nilai TF ini tidak memberikan informasi tentang pentingnya kata tersebut dalam keseluruhan kumpulan dokumen. Kata-kata umum seperti kata hubung sering muncul dengan frekuensi tinggi, tetapi tidak memiliki makna signifikan. Oleh karena itu, digunakan Inverse Document Frequency (IDF) untuk memberikan bobot pada kata-kata yang jarang muncul di seluruh dokumen (Romadhony et al., 2024). Rumus IDF adalah:

$$IDF = \log \frac{N}{n} \quad (2)$$

N adalah jumlah dokumen dalam kumpulan dokumen dan n adalah jumlah dokumen yang mengandung kata tersebut. Setelah nilai TF dan IDF diperoleh, skor TF-IDF dihitung dengan mengalikan keduanya. Kata-kata dengan nilai TF-IDF tinggi dianggap penting dan berkontribusi lebih besar dalam menentukan topik dokumen. Pembobotan TF-IDF diterapkan menggunakan Scikit-learn dengan beberapa parameter yang telah dikonfigurasi. Parameter *max\_features*= 5000, membatasi ruang fitur hanya pada 5.000 kata yang paling bermakna, sementara *ngram\_range*=(1,2) memastikan bahwa baik kata tunggal maupun kombinasi dua kata tercakup untuk mempertahankan makna kontekstual. Pengaturan *sublinear\_tf*=True menerapkan transformasi logaritmik pada frekuensi kata, mencegah kata-kata yang sangat sering muncul memengaruhi bobot secara tidak proporsional. Selain itu, *min\_df*=5 menyaring kata-kata yang jarang muncul, dan *max\_df*=0.7 menghilangkan kata-kata yang muncul terlalu luas di seluruh dokumen. Pengaturan ini dipilih secara sengaja untuk mencapai keseimbangan antara kompleksitas fitur dan relevansi, menghasilkan representasi vektor numerik yang efisien dan bermakna. Vektor TF-IDF yang dihasilkan kemudian digunakan sebagai input untuk model klasifikasi SVM.

### Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah algoritma supervised learning yang digunakan untuk melakukan klasifikasi dan regresi dengan cara membangun sebuah hyperplane yang memisahkan data berdasarkan kelasnya. Hyperplane terbaik ditentukan dari jarak pemisah (margin) yang paling besar antara data dari masing-masing kelas, sehingga mampu menghasilkan batas klasifikasi yang optimal (Sadasivam, 2021). SVM efektif ketika diterapkan pada data berdimensi tinggi, seperti representasi teks dalam analisis sentimen. Sebagai contoh dalam penelitian (Imanuddin et al., 2023) menggunakan algoritma SVM dengan membagi sentimen mereka ke dalam 2 kelas yaitu positif dan negatif, dan akurasi yang mencapai 91%, dan penelitian (Wulandari et al., 2025) juga menggunakan algoritma SVM dan memperoleh akurasi sebesar 79%.

Algoritma Support Vector Machine (SVM) dengan kernel linear digunakan dalam penelitian ini sebagai metode klasifikasi dalam penelitian ini. Model SVM dilatih menggunakan data latih sebesar 80% dan diuji dengan data uji sebesar 20% melalui metode stratified sampling agar proporsi label tetap seimbang pada kedua subset. Data uji tersebut kemudian digunakan untuk mengukur performa model yang telah dibuat.

### Evaluasi dan Visualisasi

Tahapan terakhir evaluasi dan visualisasi dilakukan untuk mengetahui kemampuan model SVM yang digunakan dan untuk mengetahui bagaimana persepsi masyarakat terhadap kampanye anti judi online di Tiktok. Evaluasi performa Support Vector Machine (SVM) dalam mengklasifikasikan sentimen masyarakat terhadap kampanye anti judi online dilakukan menggunakan *confusion matrix* untuk melihat distribusi prediksi benar dan salah pada setiap kelas sentimen. Matriks ini menampilkan jumlah prediksi benar maupun salah dalam bentuk tabel, yang dibagi ke dalam empat komponen utama: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) (Fahmy, 2022).

Tabel 1. *Confusion matrix*

|            |       | True Values         |                     |
|------------|-------|---------------------|---------------------|
|            |       | True                | False               |
| Prediction | True  | True Positive (TP)  | False Positive (FP) |
|            | False | False Negative (FN) | True Negative (TN)  |

- True Positive (TP): jumlah data yang diklasifikasikan benar sebagai kelas positif
- True Negative (TN): jumlah data yang diklasifikasikan benar sebagai kelas negatif
- False Positive (FP): jumlah data yang salah diklasifikasikan sebagai kelas positif
- False Negative (FN): jumlah data yang salah diklasifikasikan sebagai kelas negatif

. Dari keempat nilai tersebut, beberapa ukuran evaluasi dapat dihitung untuk menilai seberapa baik model dalam melakukan klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

*Accuracy* adalah persentase prediksi yang benar (baik positif maupun negatif) dari seluruh jumlah data yang diuji. Dengan rumus sebagai berikut:

$$Accuracy = \frac{TP+TN}{N} \quad (3)$$

*Precision* adalah seberapa banyak dari semua prediksi positif yang benar- benar merupakan data positif. Dengan rumus sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

*Recall (Sensitivity)* adalah seberapa banyak data positif yang berhasil dikenali oleh model dari semua data yang sebenarnya positif. Dengan rumus sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

*F1-Score* adalah rata-rata harmonis dari nilai precision dan recall. Metrik ini digunakan untuk memberikan penilaian yang seimbang antara keduanya, terutama ketika terdapat ketidakseimbangan data antar kelas. Dengan rumus sebagai berikut:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

Visualisasi distribusi sentimen positif, dan negatif ditampilkan dalam bentuk tabel dan grafik pie chart. Tahapan ini memberikan gambaran terkait persepsi masyarakat terhadap kampanye anti-judi online dan kinerja SVM dalam klasifikasi sentimen kampanye anti-judi online.

## HASIL DAN PEMBAHASAN

### Hasil Pengumpulan Data

Komentar-komentar dikumpulkan dari total 9 video TikTok dengan menggunakan tagar #StopJudol dan #LawanJudol sebagai kriteria pencarian, yang menghasilkan total 10.931 komentar berbahasa Indonesia. Dari jumlah tersebut, 3.300 komentar diambil dari video yang diberi tag #LawanJudol, sedangkan sisanya berasal dari konten yang diberi tag #StopJudol. Kumpulan data yang dikumpulkan mencakup berbagai atribut seperti teks, *diggCount*, *replyCommentTotal*, *createTimeISO*, *uniqueID*, *videoWebUrl*, *uid*, *cid*, dan *avatarThumbnail*. Namun, untuk melindungi privasi pengguna, tidak semua atribut ini digunakan dalam analisis. Selama tahap prapemrosesan, hanya teks komentar yang disimpan sebagai data inti untuk analisis sentimen, sedangkan atribut lain dikecualikan karena atribut tersebut tidak relevan dengan tujuan penelitian.

### Hasil Preprocessing

Tahapan *preprocessing data* dilakukan untuk meningkatkan kualitas data sebelum dianalisis. Proses ini melakukan penghapusan data duplikat, *cleaning*, *case folding*, *filtering* komentar promosi, tokenisasi, *stopword removal*, *stemming*, dan penghapusan data kosong. Hasil *preprocessing* ditunjukkan pada Tabel 2.

Tabel 2. Hasil Preprocessing

| No. | Text  | Preprocessed                     |
|-----|---|----------------------------------|
| 1.  | 6. Jangan kumpul dengan lingkungan yang main judi | jangan kumpul lingkung main judi |
| 2.  | blom nyoba dan gamau nyoba                        | blom nyoba gamau nyoba           |
| 3.  | betul,sedih KLO suami masih maen judol            | sedih klo suami maen judol       |
| 4.  | maxs win baru berhenti                            | maxs win baru henti              |
| 5.  | yakin berhenti                                    | yakin henti                      |

Setelah seluruh tahapan preprocessing dilakukan, jumlah data mengalami perubahan. Jumlah data yang awalnya sebanyak 10.931 komentar berkurang menjadi 6.643 komentar (23,2%). Berdasarkan Tabel 1, tahapan *preprocessing* berhasil mengurangi *noise* dalam data, seperti kata tidak baku, simbol, dan komentar yang tidak relevan, sehingga kualitas datanya menjadi baik akan memberikan pengaruh positif terhadap performa model klasifikasi.

### Hasil Hybrid Labeling

Setelah tahapan *preprocessing*, pendekatan *hybrid* yang menggabungkan IndoBERTweet dan *rule-based* berbasis *keyword* diterapkan pada data komentar. Hasil pelabelan ini mengelompokkan setiap komentar ke dalam salah satu dari dua kelas sentimen: positif atau negatif. Hasil dari proses pelabelan tersebut ditunjukkan pada Gambar 2, sedangkan distribusi label dirangkum dalam Tabel 3.

| preprocessed   | label_keyword | label_bert | label_hybrid |
|--|---------------|------------|--------------|
| semangat mau fahlan  | negatif       | positif    | positif      |
| semangat udah semangat lebih baik  | negatif       | positif    | positif      |
| semangat kalah tiba main pagi zjh  | positif       | negatif    | positif      |
| semangat   | negatif       | positif    | positif      |
| bisusilah ikut mega kkg gods   | negatif       | positif    | positif      |
| bilangin   | negatif       | negatif    | negatif      |
| bener mau tag teman main enak  | negatif       | negatif    | negatif      |
| adik sepuh jadi kaya mungkin karna judul jadi admin kamboja sejak pergi keluarga | negatif       | negatif    | negatif      |
| gimana nyadarinya yaa  | negatif       | negatif    | negatif      |
| huka   | negatif       | negatif    | negatif      |

Gambar 2. Hasil Pelabelan Data

Gambar 2 menampilkan hasil akhir proses pelabelan sentimen pada data komentar dalam video tagar #LawanJudol dan #StopJudol di Tiktok setelah melalui seluruh tahapan pra-pemrosesan. Tabel terdiri dari empat kolom utama, yaitu *preprocessed*, *label\_keyword*, *label\_bert*, dan *label\_hybrid*. Kolom *preprocessed* berisi teks komentar Tiktok yang sudah melalui tahap *preprocessing*, kolom *label\_keyword* berisi hasil pelabelan menggunakan pendekatan *rule-based* berbasis *keyword*, *label\_bert* berisi hasil dari model IndoBERTweet, dan *label\_hybrid* berisi hasil akhir dari kombinasi antara *rule-based* dan IndoBERTweet.

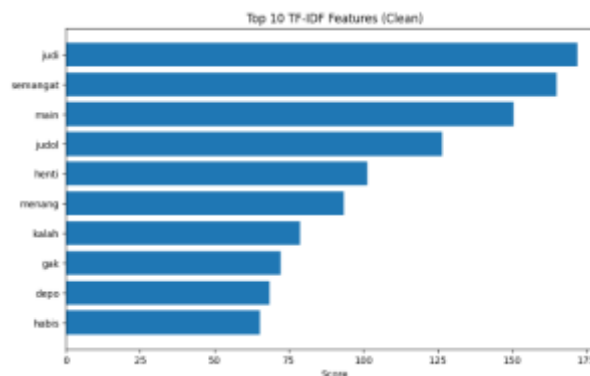
Tabel 3. Distribusi Sentimen

| Sentimen | Jumlah | Persentase |
|----------|--------|------------|
| Positif  | 1813   | 27,3%      |
| Negatif  | 4830   | 72,7%      |
| Total    | 6643   | 100%       |

Hasil pelabelan menunjukkan bahwa mayoritas komentar sebanyak 72,7% bersentimen negatif. Dominasi sentimen negatif mengindikasikan bahwa masih terdapat pengguna yang menunjukkan sikap permisif terhadap aktivitas perjudian online, baik melalui dukungan langsung, candaan, maupun pengalaman pribadi terkait kemenangan dan aktivitas bermain judi. Kondisi ini juga dapat dipengaruhi oleh karakteristik media sosial TikTok yang memungkinkan berbagai kelompok pengguna berinteraksi secara terbuka pada konten kampanye, termasuk pengguna yang tidak sepakat dengan pesan yang disampaikan. Meskipun demikian, jumlah komentar positif sebanyak 27,3% menunjukkan bahwa kampanye tetap memperoleh dukungan dari sebagian pengguna yang menyuarakan ajakan berhenti berjudi dan kesadaran terhadap dampak negatif perjudian online.

### Hasil Pembobotan TF-IDF

Tahapan selanjutnya adalah pembobotan fitur menggunakan TF-IDF untuk mengidentifikasi kata-kata yang paling berpengaruh dalam setiap kelas sentimen. Kata dengan bobot tinggi menunjukkan kata yang paling berpengaruh dalam dataset. Gambar 3 menunjukkan kata dengan tf-idf tertinggi.



Gambar 3. Kata dengan TF-IDF Tertinggi

Gambar 3 menunjukkan sepuluh kata dengan bobot TF-IDF tertinggi yang merepresentasikan karakteristik utama data komentar Tiktok terkait kampanye anti-judi online. Kata-kata seperti “judi”, “judol”, “depo”, “menang”, dan “kalah” mendominasi, yang menunjukkan bahwa diskusi pengguna banyak berfokus pada aktivitas dan pengalaman terkait perjudian online. Di sisi lain, munculnya kata seperti “semangat” dan “henti” mencerminkan adanya dukungan terhadap kampanye anti-judi, terutama dalam bentuk motivasi untuk berhenti dan ajakan perubahan perilaku. Hal ini menunjukkan bahwa pesan kampanye telah terserap dan direfleksikan dalam komentar pengguna.

Kata “gak” muncul sebagai salah satu fitur dominan karena sering digunakan dalam konteks negasi, seperti “gak mau judi” atau “gak main lagi”, yang mengindikasikan sikap penolakan terhadap praktik perjudian. Temuan ini menunjukkan bahwa percakapan pengguna di Tiktok tidak hanya didominasi oleh istilah terkait perjudian, tetapi juga mengandung elemen dukungan terhadap kampanye anti-judi, sehingga mencerminkan dinamika persepsi publik yang terbentuk di media sosial.

Meskipun demikian, dominasi kata-kata yang berkaitan dengan aktivitas perjudian menunjukkan bahwa perhatian sebagian besar pengguna masih tertuju pada praktik perjudian itu sendiri dibandingkan pesan pencegahan yang disampaikan dalam kampanye. Hal ini mengindikasikan bahwa konten kampanye anti-judi tidak hanya menjangkau pengguna yang mendukung upaya pencegahan perjudian, tetapi juga pengguna yang membahas atau memiliki pengalaman terkait perjudian online. Akibatnya, kolom komentar menjadi ruang interaksi berbagai pandangan, baik yang mendukung maupun yang tidak sejalan dengan tujuan kampanye. Kondisi tersebut dapat menjelaskan dominasi sentimen negatif pada hasil pelabelan, karena sebagian pengguna lebih banyak membahas aktivitas dan pengalaman terkait perjudian dibandingkan memberikan dukungan terhadap pesan kampanye anti-judi online.

### Hasil Klasifikasi SVM

Dalam tahapan ini, jumlah fitur yang digunakan sebanyak 1580 fitur, yang diperoleh dari proses ekstraksi fitur menggunakan metode TF-IDF dengan representasi unigram dan bigram. Fitur-fitur tersebut digunakan sebagai input dalam proses klasifikasi menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear. Evaluasi performa model dilakukan menggunakan metrik *precision*, *recall*, *f1-score* dan *accuracy*. Hasil evaluasi model ditunjukkan pada Gambar 4, sedangkan distribusi prediksi dalam bentuk *confusion matrix* ditampilkan pada Gambar 5.

```

    • Training SVM Linear
    Classification Report:
              precision    recall  f1-score   support

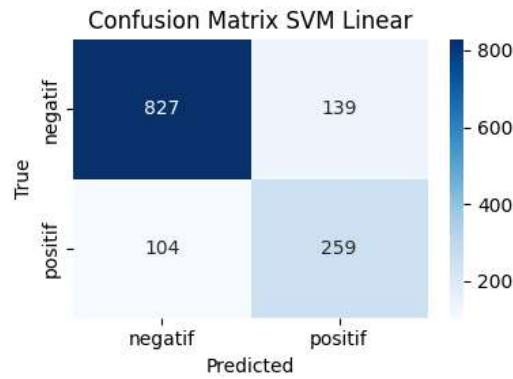
    negatif   0.888    0.856    0.872     966
    positif   0.651    0.713    0.681     363

    accuracy          0.817    1329
    macro avg         0.770    0.785    0.776    1329
    weighted avg      0.823    0.817    0.820    1329
  
```

Gambar 4. Hasil Akurasi SVM

Berdasarkan hasil pengujian, model SVM dengan kernel linear menghasilkan nilai akurasi sebesar 81,7% yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan cukup baik. Pada kelas negatif, model memperoleh nilai *precision* sebesar 0.888, *recall* sebesar 0.856, dan *f1-score* sebesar 0.872, yang menunjukkan performa tinggi dalam mengenali komentar yang tidak mendukung kampanye anti-judi online.

Sementara itu, pada kelas positif, model memperoleh nilai *precision* sebesar 0.651, *recall* sebesar 0.713, dan *f1-score* sebesar 0.681. Nilai *recall* yang relatif lebih tinggi dibandingkan *precision* menunjukkan bahwa model cukup mampu mendeteksi komentar yang mendukung kampanye anti-judi, meskipun masih terdapat kesalahan berupa *false positive*, yaitu komentar yang sebenarnya tidak mendukung tetapi diklasifikasikan sebagai positif.



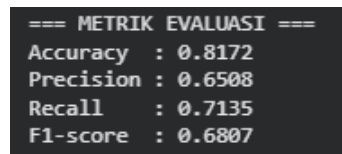
Gambar 5. Confusion Matrix

Hasil *confusion matrix* menunjukkan bahwa model SVM mampu mengklasifikasikan komentar negatif dengan baik, ditunjukkan oleh 827 prediksi benar dan 139 kesalahan klasifikasi. Pada kelas positif, model berhasil mengenali 259 komentar dengan benar, namun masih terdapat 104 komentar positif yang salah diklasifikasikan sebagai negatif, serta 139 komentar negatif yang salah diklasifikasikan sebagai positif.

Secara keseluruhan, model cenderung memiliki performa yang lebih baik dalam mengklasifikasikan sentimen negatif dibandingkan sentimen positif. Hal ini disebabkan oleh distribusi data yang tidak seimbang yaitu komentar negatif lebih dominan dibandingkan komentar positif serta variasi bahasa pada komentar positif yang lebih kompleks dan beragam.

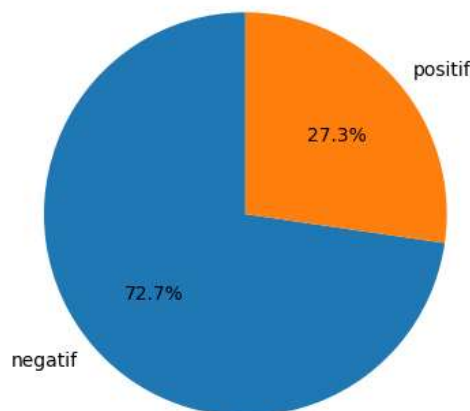
### Evaluasi dan Visualisasi Sentimen

Tahapan terakhir evaluasi dan visualisasi dilakukan untuk mengetahui kemampuan model SVM yang digunakan dan untuk mengetahui bagaimana persepsi masyarakat terhadap kampanye anti judi online di Tiktok. Evaluasi performa Support Vector Machine (SVM) dalam mengklasifikasikan sentimen masyarakat terhadap kampanye anti judi online dilakukan menggunakan parameter *accuracy*, *precision*, *recall* dan *F1-score*. Hasilnya ditunjukkan dalam Gambar 6.



Gambar 6. Metrik Evaluasi

### Distribusi Sentimen Masyarakat



Gambar 7. Distribusi Sentimen Masyarakat

Diagram lingkaran menunjukkan bahwa mayoritas komentar (72,7%) termasuk dalam kategori negatif, sedangkan 27,3% sisanya bersifat positif. Dominasi sentimen negatif mengindikasikan bahwa masih adanya resistensi terhadap pesan kampanye anti-judi online di kalangan pengguna TikTok. Tingginya proporsi komentar negatif dapat dipengaruhi oleh karakteristik media sosial yang memungkinkan pengguna menyampaikan berbagai pandangan dan pengalaman secara terbuka pada kolom komentar. Hasil analisis TF-IDF menunjukkan bahwa percakapan pengguna masih banyak berfokus pada aktivitas dan pengalaman terkait perjudian online, yang ditunjukkan oleh dominasi kata-kata seperti “judi”, “judol”, “depo”, “menang”, dan “kalah”. Selain itu, mekanisme penyebaran konten melalui fitur *For You Page* (FYP) memungkinkan kampanye menjangkau audiens yang sangat beragam, termasuk pengguna yang tidak memiliki pandangan sejalan dengan tujuan kampanye. Kondisi tersebut menyebabkan kolom komentar menjadi ruang interaksi berbagai perspektif, baik yang mendukung maupun yang tidak mendukung pesan anti-judi online, sehingga berkontribusi terhadap dominasi sentimen negatif.

Temuan ini menunjukkan bahwa tantangan kampanye anti-judi online tidak hanya terletak pada penyebaran informasi kepada masyarakat, tetapi juga pada upaya membangun kesadaran dan mengubah persepsi sebagian pengguna terhadap risiko perjudian online. Secara keseluruhan, model SVM kernel linier bekerja dengan cukup baik, terutama dalam mendeteksi komentar negatif, sementara kinerjanya pada kelas positif relatif lebih rendah karena ketidakseimbangan kelas dan variasi linguistik yang lebih besar pada komentar positif. Hasil penelitian ini menunjukkan bahwa persepsi audiens terhadap kampanye anti-judi online di TikTok masih didominasi oleh sentimen negatif, meskipun terdapat sejumlah pengguna yang memberikan dukungan terhadap pesan kampanye.

## KESIMPULAN

Penelitian ini menerapkan pendekatan TF-IDF dan Support Vector Machine (SVM) untuk menganalisis sentimen publik terhadap kampanye anti-judi online di TikTok. Model SVM dengan kernel linier menunjukkan hasil yang memuaskan, dengan tingkat akurasi sebesar 81,7% dan kemampuan yang andal dalam mengklasifikasikan sebagian besar komentar secara tepat. Dalam hal distribusi sentimen, sebagian besar komentar (72,7%) bersifat negatif, dengan hanya 27,3% yang diklasifikasikan sebagai positif, menunjukkan bahwa ekspresi dukungan terhadap kampanye anti-judi masih kalah jumlah dibandingkan komentar yang menentang atau mendukung perilaku perjudian.

Model menunjukkan performa yang lebih baik pada kelas negatif dibandingkan kelas positif. Hal ini dipengaruhi oleh ketidakseimbangan data serta kompleksitas bahasa pada komentar positif yang lebih bervariasi dan kontekstual. Meskipun demikian, penggunaan representasi TF-IDF membantu dalam mengekstraksi fitur teks pada data media sosial yang bersifat pendek dan tidak formal. Penelitian ini juga menerapkan pendekatan *hybrid labeling* yang menggabungkan metode rule-based dan model pra-latih seperti IndoBERTweet dalam proses pelabelan data memungkinkan pembuatan label yang lebih terstruktur untuk mendukung pelatihan model klasifikasi.

Secara keseluruhan, penelitian ini memberikan gambaran tentang persepsi masyarakat terhadap kampanye anti-judi online, serta menunjukkan bahwa kombinasi metode berbasis aturan dan model pra-latih dapat digunakan sebagai strategi dalam proses pelabelan data pada analisis sentimen teks media sosial.

## DAFTAR PUSTAKA

- (PPATK), P. P. dan A. T. K. (2024). *Laporan Tahunan PPATK 2024*. <https://www.ppatk.go.id/publikasi/read/255/laporan-tahunan-ppatk-tahun-2024.html>
- Aftab, F., Bazai, S. U., Marjan, S., Baloch, L., & Aslam, S. (2023). *A Comprehensive Survey on Sentiment Analysis Techniques*. 14(May), 1288–1298. <https://doi.org/10.14716/ijtech.v14i6.6632>
- As Shidiq, M. F., & Alita, D. (2025). Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data Dari Media Sosial X Pendekatan Naive Bayes Dan SVM. *Jurnal Sistem Informasi Dan Informatika (Simika)*, 8(1), 24–35. <https://doi.org/10.47080/simika.v8i1.3624>
- Fahmy, M. M. (2022). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research (ERJ)*, 6(5), 1–12. [https://erjeng.journals.ekb.eg/article\\_296718\\_30a98aac15193d04dc73ba9bc00cf046.pdf](https://erjeng.journals.ekb.eg/article_296718_30a98aac15193d04dc73ba9bc00cf046.pdf)
- Howse, E., Cullerton, K., Grunseit, A., Bohn-Goldbaum, E., Bauman, A., & Freeman, B. (2022). Measuring public opinion and acceptability of prevention policies: an integrative review and narrative synthesis of methods. *Health Research Policy and Systems*, 20(1), 1–19. <https://doi.org/10.1186/s12961-022-00829-y>
- Imanuddin, S. H., Adi, K., & Gernowo, R. (2023). Sentiment Analysis on Satusehat Application Using Support Vector Machine Method. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 5(3), 143–149. <https://doi.org/10.35882/jeemi.v5i3.304>

- Koto, F., Lau, J. H., & Baldwin, T. (2021). INDOBERTWEET : A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 10660–10668.
- Lin, C. H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00782-9>
- Maulana, A., & Yuliana, A. (2024). Analisis Sentimen Opini Publik Terkait Judi Online Pada Pengguna Aplikasi X Menggunakan Algoritma Naïve Bayes Dan Support Vector Mechine. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 12(3), 3706–3714.
- Munfida, N. S., Soetijono, I. R., & Indrayati, R. (2023). Penggunaan Media Sosial sebagai Media Kampanye dalam Penyelenggaraan Pemilihan Umum dan Pemilihan Kepala Daerah. *Jurnal Kajian Konstitusi*, 3(2), 161–181. <https://doi.org/10.19184/j.kk.v3i2.41193>
- Perdana, C., & Sulastri. (2025). EFEKTIVITAS KAMPANYE SOSIAL MELALUI TIKTOK : STUDI KASUS KAMPANYE LINGKUNGAN. *Jurnal Media Dan Komunikasi*, 2, 46–49.
- Pratama, Y. A., Sarwoprasodjo, S., & Seminar, A. U. (2023). Efektivitas Tiktok Sebagai Media Kampanye Komunikasi Viral Pencegahan Dan Penanganan Covid-19. *Jurnal Komunikasi*, 9(2), 59–68. <https://doi.org/10.30997/jk.v9i2.7786>
- Ramos, J. (2003). *Using TF-IDF to Determine Word Relevance in Document Queries*. [https://pdfs.semanticscholar.org/b3bf/6373ff41a115197cb5b30e57830c16130c2c.pdf%0A?\\_ga=1.183821606.1606390940.1482600858](https://pdfs.semanticscholar.org/b3bf/6373ff41a115197cb5b30e57830c16130c2c.pdf%0A?_ga=1.183821606.1606390940.1482600858)
- RI, K. K. dan D. (2025). *Sejak Awal Januari 2025 Komdigi Telah Menindak 43 Ribu Konten Judol*. Dinas Komunikasi, Informatika, Dan Persandian Kota Lhokseumawe. <https://kominfo.lhokseumawekota.go.id/berita/read/sejak-awal-januari-2025-komdigi-telah-menindak-43-ribu-konten-judol>
- Rizkia, A. S., Wufron, & Roji, F. F. (2025). Sentiment Analysis of Coretax : A Comparison of Manual , Transformers- Based , and Lexicon-Based Data Labeling on IndoBERT Performance. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(July), 1037–1048.
- Romadhony, A., Faraby, S. Al, Rismala, R., Wiseti, U. N., & Arifianto, A. (2024). Sentiment Analysis on a Large Indonesian Product Review Dataset. *Journal of Information Systems Engineering and Business Intelligence*, 10(1), 167–178.
- Sadasivam, D. G. S. (2021). Crop Yield Prediction using Granular SVM. *International Journal of Recent Technology and Engineering (IJRTE)*, 9(6), 85–89. <https://doi.org/10.35940/ijrte.f5417.039621>
- Sanggau, D. K. dan I. K. (2025). *TikTok Indonesia Luncurkan Kampanye #LawanJudol; Tegaskan Komitmen Lawan Aktivitas Perjudian Online bersama Komdigi RI*. Diskominfo Sanggau. <https://diskominfo.sanggau.go.id/berita/read/tiktok-indonesia-luncurkan-kampanye-lawanjudol-tegaskan-komitmen-lawan-aktivitas-perjudian-online-bersama-komdigi-ri>
- Sari, A. K., Al-Fajrih, M., & Ahdiyanti, I. (2024). Dampak Judi Online Terhadap Kesehatan Mental Dan Hubungan Sosial. *Al-Hikmah: Jurnal Dakwah Dan Komunikasi*, 3(2), 31–44. <https://doi.org/10.47945/al-hikmah.v3i2.1539>
- Syahputra, H., & Wibowo, A. (2023). Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, 9(1), 165–173. <https://doi.org/10.26555/jiteki.v9i1.25861>
- Uliniansyah, M. T., Jarin, A., & Santosa, A. (2025). *Modeling sentiment analysis of Indonesian biodiversity policy Tweets using IndoBERTweet*. 14(3), 2389–2401. <https://doi.org/10.11591/ijai.v14.i3.pp2389-2401>
- Wulandari, I., Prastya, S. E., Zulfadhilah, M., & Anshari, R. (2025). Analisis Sentimen Terhadap Tindakan Kekerasan Seksual pada Media Sosial Tiktok Menggunakan Metode Support Vector Machine (SVM). *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 8(2), 768–776. <https://doi.org/10.32672/jnkti.v8i2.8892>