DETEKSI DAN KLASIFIKASI KONTEN PROMOSI JUDI ONLINE PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN PENDEKATAN *NATURAL LANGUAGE PROCESSING* DAN *MACHINE LEARNING*

M. Luthfi Aldi Pratama, Muhammad Emirshah Yusuf, M Rafly Ramdhani, Fathoni, Ali Ibrahim
Sistem Informasi, Universitas Sriwijaya
Jl. Srijaya Negara, Bukit Lama, Kec. Ilir Barat I, Kota Palembang, 30139, Indonesia
mluthfialdi@gmail.com

ABSTRAK

Konten promosi judi online pada kolom komentar YouTube menjadi salah satu bentuk penyalahgunaan media sosial yang sulit dikendalikan. Penyebaran dilakukan dengan variasi teks sehingga menyulitkan proses identifikasi secara langsung. Penelitian ini bertujuan untuk mengembangkan pendekatan otomatis dalam mendeteksi dan mengklasifikasikan konten tersebut. Pendekatan yang digunakan mengombinasikan Natural Language Processing dan Machine Learning. Data yang digunakan berjumlah 10.230 komentar yang terdiri dari kategori judi online dan non-judi. Representasi teks dibentuk menggunakan TF-IDF, kemudian dilakukan proses klasifikasi menggunakan algoritma Support Vector Machine dan Random Forest. Hasil evaluasi menunjukkan bahwa Support Vector Machine menghasilkan performa yang lebih baik dengan akurasi 97,16%, precision 97,67%, recall 95,67%, dan F1-score 96,66%, sedangkan Random Forest memperoleh akurasi sebesar 96,72%. Hasil tersebut menunjukkan bahwa Support Vector Machine lebih efektif dalam menangani karakteristik data pada penelitian ini.

Kata kunci : *Natural Language Processing, judi online, Support Vector Machine, Random Forest, YouTube*

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah menempatkan media sosial sebagai sarana utama dalam interaksi digital dan pertukaran informasi secara global. YouTube, sebagai salah satu platform berbagi video terbesar, memberikan ruang bagi pengguna untuk berpartisipasi aktif melalui fitur kolom komentar yang memungkinkan terjadinya komunikasi dua arah. Namun, keterbukaan akses ini juga membawa tantangan baru dalam pengelolaan konten digital, di mana platform sering kali disalahgunakan oleh pihak tidak bertanggung jawab untuk menyebarkan materi yang melanggar hukum. Fenomena ini menciptakan tekanan pada integritas platform media sosial dalam menjaga kualitas interaksi antar pengguna dari berbagai gangguan konten ilegal. [1]

Salah satu permasalahan krusial yang muncul pada ekosistem YouTube adalah maraknya penyebaran promosi judi online yang dilakukan secara masif melalui kolom komentar. Para pelaku umumnya memanfaatkan akun *bot* untuk menyebarkan tautan situs perjudian pada video-video yang memiliki jumlah penonton tinggi guna menjangkau audiens secara luas dalam waktu singkat. Aktivitas ini tidak hanya mengganggu estetika dan kenyamanan interaksi pengguna, tetapi juga menjadi sarana infiltrasi konten berbahaya yang sulit dikendalikan hanya melalui pemantauan manual. Pola penyebaran yang agresif ini menuntut adanya sistem proteksi yang lebih canggih pada setiap platform interaksi digital. [2]

Aktivitas judi online membawa dampak sosial dan ekonomi yang sangat merusak bagi berbagai lapisan masyarakat di Indonesia. Secara ekonomi, perjudian menyebabkan kerugian finansial yang signifikan bagi individu dan keluarga, yang sering kali berujung pada peningkatan angka kriminalitas serta utang piutang. Secara sosial, paparan konten perjudian yang mudah diakses oleh generasi muda dapat memicu adiksi dan normalisasi terhadap perilaku ilegal yang bertentangan dengan norma masyarakat. Oleh karena itu, penanganan terhadap promosi judi online di ruang digital menjadi langkah darurat dalam melindungi ketahanan ekonomi dan mentalitas masyarakat dari pengaruh negatif industri perjudian. [3]

Secara teknis, deteksi terhadap komentar promosi judi online menghadapi kendala besar akibat adanya teknik manipulasi teks atau *obfuscated text*. Para promotor sering kali menggunakan variasi penulisan kreatif, seperti penggunaan simbol, penggantian huruf dengan angka, hingga penyisipan karakter *non-alfanumerik* guna menghindari filter kata kunci standar. Teknik ini dirancang agar pesan tetap dapat dipahami secara visual oleh manusia, namun terbaca sebagai data acak atau noise oleh sistem keamanan tradisional. Dinamika perubahan ejaan dan penggunaan istilah slang yang terus berevolusi ini menjadikan identifikasi konten spam judi online menjadi tugas yang sangat kompleks dalam pemrosesan bahasa alami. [4]

Beberapa penelitian terdahulu telah berupaya mengatasi permasalahan spam dan konten ilegal menggunakan berbagai pendekatan *machine learning* dan *Natural Language Processing* (NLP). Implementasi algoritma klasik seperti *Naive Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* telah banyak digunakan untuk melakukan klasifikasi teks berdasarkan fitur-fitur linguistik tertentu. Studi-studi tersebut menunjukkan tingkat keberhasilan yang cukup baik dalam mendeteksi teks spam standar yang

memiliki pola bahasa formal dan konsisten. Integrasi teknik *word embedding* juga telah mulai diterapkan untuk meningkatkan pemahaman mesin terhadap representasi kata dalam dataset teks bahasa Indonesia. [5]

Meskipun terdapat berbagai perkembangan metode, sebagian besar penelitian sebelumnya masih memiliki keterbatasan dalam menangani teks yang sengaja dimanipulasi secara ekstrem. Banyak model klasifikasi konvensional yang mengalami penurunan akurasi yang signifikan saat dihadapkan pada variasi ejaan non-standar dan penggunaan karakter unik yang sering muncul dalam komentar judi online. Kesenjangan penelitian ini menunjukkan bahwa pendekatan berbasis pencocokan pola atau algoritma pembelajaran mesin sederhana belum cukup adaptif untuk mengenali konteks kalimat yang tidak terstruktur secara semantik. Keterbatasan dalam menangkap hubungan konteks antar kata menyebabkan banyak konten promosi tetap lolos dari proses penyaringan otomatis.[6]

Sebagai upaya untuk menutup celah tersebut, penelitian ini mengusulkan penggunaan model IndoBERT untuk meningkatkan akurasi deteksi dan klasifikasi promosi judi online di platform YouTube. Penggunaan model berbasis *Transformer* ini dipilih karena kemampuannya yang unggul dalam memahami konteks bahasa Indonesia secara dua arah (*bidirectional*), sehingga lebih sensitif terhadap nuansa bahasa informal dan teks yang dimanipulasi. Dengan memanfaatkan teknik *fine-tuning* pada dataset komentar riil, penelitian ini diharapkan dapat memberikan kontribusi berupa model klasifikasi yang lebih tangguh dan akurat. Hasil penelitian ini diharapkan menjadi solusi teknis yang efektif dalam menjaga kebersihan ruang komentar digital dan memitigasi penyebaran konten perjudian di internet.[7]

2. TINJAUAN PUSTAKA

Penelitian mengenai deteksi konten perjudian online pada media sosial telah banyak dilakukan dengan menggunakan berbagai pendekatan machine learning dan deep learning. Pendekatan *Convolutional Neural Network* (CNN) dan Word2Vec untuk menganalisis sentimen publik terhadap fenomena judi online. Hasil penelitian menunjukkan bahwa model CNN mampu mencapai tingkat akurasi sebesar 99,10% dalam klasifikasi sentimen terkait perjudian online.[9]

Pengembangan sistem deteksi komentar promosi judi online pada YouTube menggunakan metode *Regular Expression* dan *Fuzzy Matching*. Sistem tersebut mampu mendeteksi pola teks promosi dengan tingkat akurasi sebesar 90,85%, meskipun pendekatan berbasis aturan memiliki keterbatasan dalam mendeteksi variasi bahasa yang kompleks.[4]

Pendekatan berbasis probabilistik seperti *Naive Bayes* yang dikombinasikan dengan TF-IDF untuk melakukan filtrasi komentar spam yang berisi promosi judi online pada YouTube. Model yang dikembangkan mampu mencapai akurasi sebesar 97,1% dengan presisi sebesar 96,4% dan F1-score sebesar 96%.[5]

Penelitian lain membandingkan performa beberapa metode analisis sentimen seperti *Transformers* (IndoBERT), VADER, dan *Naive Bayes* dalam menganalisis opini masyarakat terkait judi online. Hasil penelitian menunjukkan bahwa model berbasis transformer memiliki performa yang lebih baik dalam memahami konteks bahasa dibandingkan metode tradisional.[7]

Penggunaan algoritma *Random Forest* untuk mendeteksi promosi judi online pada media sosial dengan dataset lebih dari 10.000 komentar dan menghasilkan akurasi klasifikasi sebesar 92,97%.[10]

Pendekatan lain menggunakan pendekatan *Named Entity Recognition* (NER) dan *Support Vector Machine* (SVM) untuk mengklasifikasikan aktivitas judi online dari artikel berita digital. Penelitian tersebut menunjukkan bahwa metode machine learning dapat digunakan untuk mengidentifikasi pola aktivitas perjudian pada data teks digital.[11]

Pemanfaatan IndoBERT yang dikombinasikan dengan *One-Class* SVM mampu mendeteksi iklan judi online yang bersifat terselubung melalui pendekatan deteksi anomali pada teks.[12]

Selain itu, penelitian lain juga mengembangkan sistem deteksi promosi judi online menggunakan kombinasi *Natural Language Processing* dan *Deep Learning* berbasis Transformer yang menunjukkan tingkat akurasi tinggi dalam identifikasi teks promosi perjudian.

Berdasarkan berbagai penelitian tersebut dapat disimpulkan bahwa pendekatan *Natural Language Processing* dan *Machine Learning* memiliki potensi besar dalam mendeteksi konten berbahaya di media sosial. Namun sebagian penelitian masih berfokus pada analisis sentimen atau menggunakan satu algoritma tertentu. Oleh karena itu, penelitian ini mencoba melakukan perbandingan performa antara algoritma SVM dan *Random Forest* dalam mendeteksi konten promosi judi online pada platform YouTube.

3. METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen. Pendekatan kuantitatif diterapkan untuk mengukur kinerja model klasifikasi secara objektif melalui metrik evaluasi seperti *accuracy*, *precision*, recall, dan F1-*score*.

Metode eksperimen dilakukan dengan membangun model klasifikasi teks menggunakan pendekatan *Natural Language Processing* (NLP) serta algoritma *machine learning*, kemudian menguji performa model dalam mendeteksi konten promosi judi online pada komentar YouTube.

3.2. Sumber dan Jenis Data

Data dalam penelitian ini termasuk data sekunder yang diperoleh melalui platform Kaggle. Dataset yang

digunakan berupa kumpulan komentar pengguna pada YouTube yang telah dikelompokkan ke dalam dua kategori, yaitu komentar yang mengandung promosi judi online dan komentar yang tidak mengandung unsur tersebut. Adapun konsentrasi data yang digunakan dapat dilihat secara lengkap pada Tabel 1.

Tabel 1. Konsentrasi Data

Kelas	Label	Jumlah Data	Persentase (%)
Non-Judi	0	6021	58,85%
Judi Online	1	4209	41,15%
Total		**10230**	**100%**

Struktur dataset yang digunakan dalam penelitian ini terdiri dari beberapa atribut yang merepresentasikan informasi terkait video dan komentar pengguna. Atribut tersebut meliputi *video_id* yang berfungsi sebagai identitas unik video, *title* yang menunjukkan judul video, serta *channel_name* yang merepresentasikan nama channel YouTube. Selain itu, terdapat atribut *tanggal* yang menunjukkan waktu komentar dipublikasikan dan *author* yang merupakan nama pengguna yang memberikan komentar. Atribut utama dalam penelitian ini adalah *komentar* yang berisi teks komentar pengguna, serta *label* yang menunjukkan kelas data, yaitu komentar yang mengandung promosi judi online dan komentar non-judi.

3.3. Teknik Pengumpulan Data

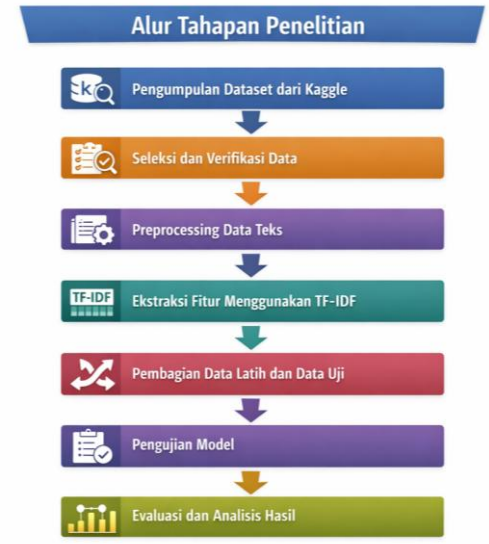
Penelitian ini memanfaatkan data sekunder yang diperoleh secara tidak langsung dari sumber yang telah tersedia, yaitu melalui platform Kaggle. Data yang digunakan merupakan data publik yang dapat diakses dan dimanfaatkan untuk keperluan analisis serta penelitian.

Selanjutnya, dilakukan tahap seleksi data untuk memastikan bahwa dataset yang digunakan memiliki struktur yang lengkap dan sesuai dengan kebutuhan penelitian, terutama pada bagian kolom komentar dan label klasifikasi.

3.4. Alur Tahapan Penelitian

Berdasarkan Gambar 4 dan Tabel 1, dapat dianalisis beberapa temuan penting sebagai berikut

a. Pengumpulan dataset dari Kaggle
   Data dikumpulkan dari platform Kaggle yang menyediakan berbagai dataset publik yang dapat dimanfaatkan untuk keperluan penelitian dan analisis. Dataset yang digunakan dalam penelitian ini merupakan data yang telah tersedia secara terbuka dan dapat diakses oleh peneliti, sehingga memungkinkan proses pengumpulan data dilakukan secara efisien dan sistematis.
b. Seleksi dan verifikasi data
   Data yang telah diperoleh selanjutnya melalui tahap penyaringan untuk memastikan kelengkapan serta kesesuaiannya dengan kebutuhan penelitian, terutama pada kolom komentar dan label klasifikasi. Proses ini bertujuan untuk menghindari penggunaan data yang tidak relevan, tidak lengkap, maupun tidak sesuai, sehingga kualitas data yang digunakan dalam penelitian dapat tetap terjaga.



Gambar 1. Alur Tahapan Penelitian

c. *Preprocessing* data teks
   Tahap preprocessing dilakukan untuk membersihkan serta mempersiapkan data teks sebelum memasuki proses klasifikasi. Proses ini bertujuan untuk mengurangi noise, memperbaiki struktur data, dan meningkatkan kualitas data agar lebih optimal untuk dianalisis. Dengan demikian, hasil klasifikasi yang dihasilkan diharapkan menjadi lebih akurat dan relevan. Adapun tahapan preprocessing data dalam penelitian ini ditampilkan pada Gambar 2.

Gambar 2. *Preprocessing Data*

a. *Case Folding*
   Mengubah seluruh teks menjadi huruf kecil (*lowercase*) untuk menjaga konsistensi representasi kata.
b. Penghapusan URL
   Menghapus tautan atau URL menggunakan ekspresi reguler karena tidak relevan dalam proses klasifikasi.

c. Penghapusan Angka
   Menghapus karakter numerik yang tidak berkontribusi terhadap penentuan kategori komentar.
d. Penghapusan Tanda Baca
   Menghapus simbol dan tanda baca untuk menyederhanakan teks.
e. Normalisasi Spasi
   Menghilangkan spasi berlebih sehingga teks menjadi lebih rapi dan terstruktur.
   Penelitian ini tidak menggunakan proses *stemming* dan *stopword removal* untuk mempertahankan konteks asli dari teks, terutama pada komentar yang mengandung bahasa tidak baku dan variasi penulisan.
d. Ekstraksi fitur menggunakan TF-IDF
   Representasi numerik diperoleh melalui penerapan TF-IDF yang menilai pentingnya suatu kata berdasarkan distribusinya pada kumpulan dokumen. Kata yang dominan pada satu dokumen namun tidak umum pada dokumen lain akan memperoleh bobot lebih besar. Representasi ini digunakan sebagai dasar dalam proses klasifikasi.
e. Pembagian data menjadi data latih dan data uji
   Dataset dipisahkan menjadi dua bagian dengan komposisi 80% sebagai data pelatihan dan 20% sebagai data pengujian. Pemisahan ini memungkinkan evaluasi model dilakukan pada data yang tidak terlibat selama proses pembelajaran.
f. Pelatihan model menggunakan algoritma SVM dan *Random Forest*
   Dua pendekatan digunakan dalam proses pembelajaran, yaitu *Support Vector Machine* dan *Random Forest*. Keduanya diterapkan untuk melihat perbedaan performa dalam menangani klasifikasi data.
g. Pengujian model
   Evaluasi awal dilakukan dengan mengaplikasikan model terhadap data pengujian. Hasil prediksi dari tahap ini digunakan untuk melihat kemampuan model dalam menghadapi data baru.
h. Evaluasi dan analisis hasil
   Performa model dianalisis menggunakan metrik akurasi, *precision*, *recall*, dan F1-*score* untuk menilai ketepatan serta konsistensi hasil klasifikasi.

3.5. Pembagian Data

Dataset yang telah melalui tahap *preprocessing* dan ekstraksi fitur kemudian dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian ini dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan dalam proses pembangunan model, sedangkan data uji dimanfaatkan untuk mengevaluasi kinerja model terhadap data yang belum pernah digunakan sebelumnya.

3.6. Support Vector Machine (SVM)

Hasil pengujian menggunakan algoritma pertama menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat kinerja yang cukup baik. Nilai evaluasi yang diperoleh mencerminkan kemampuan model dalam mengenali pola pada data yang digunakan. Performa ini dipengaruhi oleh karakteristik data serta proses pelatihan yang telah dilakukan sebelumnya. Menurut Vladimir Vapnik, SVM bertujuan untuk menemukan fungsi pemisah yang paling optimal antar kelas dengan cara memaksimalkan margin.

3.7. Random Forest

Hasil yang diperoleh dari algoritma kedua memperlihatkan adanya perbedaan performa dibandingkan model sebelumnya. Perbedaan tersebut dapat diamati dari nilai metrik evaluasi yang dihasilkan. Faktor seperti metode pembelajaran dan karakteristik algoritma turut memengaruhi kemampuan model dalam melakukan klasifikasi secara optimal. Metode ini pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 sebagai salah satu teknik *ensemble learning* untuk meningkatkan akurasi prediksi.

3.8. Evaluasi Model

Evaluasi kinerja model dilakukan dengan memanfaatkan confusion matrix untuk melihat kesesuaian antara hasil prediksi dan data aktual. Melalui matriks ini, dapat diketahui distribusi prediksi benar dan salah pada masing-masing kelas.

Penilaian performa selanjutnya menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Accuracy menggambarkan tingkat ketepatan keseluruhan model, precision menunjukkan keakuratan prediksi pada kelas tertentu, sedangkan recall merepresentasikan kemampuan model dalam menemukan seluruh data yang relevan. F1-score digunakan sebagai ukuran gabungan yang menyeimbangkan nilai precision dan recall.

Metrik tersebut digunakan sebagai dasar dalam membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest untuk menentukan model dengan performa yang paling optimal.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengujian Model

Tabel 2. Konsentrasi Data

Model	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	***F1-Score (%)***
Support Vector Machine (SVM)	97,16	97,67	95,67	96,66
Random Forest	96,72	96,98	95,33	96,15

Berdasarkan data pengujian pada Tabel 2, model SVM memperoleh nilai *accuracy* sebesar 97,16%, *precision* sebesar 97,67%, *recall* sebesar 95,67%, dan F1-score sebesar 96,66%. Sementara itu, model *Random Forest* memperoleh *accuracy* sebesar 96,72%, *precision* sebesar 96,98%, *recall* sebesar 95,33%, dan F1-score sebesar 96,15%.

Hasil ini menunjukkan bahwa kedua model memiliki performa yang sangat baik dalam mengklasifikasikan data teks, dengan tingkat akurasi di atas 96%.

4.2. Komparasi Performa Model

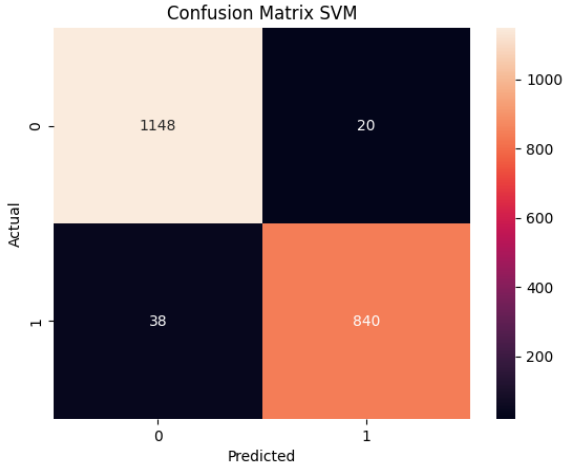
Berdasarkan hasil evaluasi, model SVM menunjukkan performa yang lebih unggul dibandingkan *Random Forest* pada seluruh metrik evaluasi. Meskipun selisih performa tidak terlalu signifikan, SVM menunjukkan konsistensi yang lebih baik dalam melakukan klasifikasi. Keunggulan ini terutama terlihat pada nilai F1-score yang lebih tinggi, yang menunjukkan keseimbangan antara *precision* dan *recall*. Dalam konteks deteksi judi online, keseimbangan ini sangat penting untuk meminimalkan kesalahan klasifikasi baik pada data positif maupun negatif.

4.3. Analisis Model Terbaik SVM

Model *Support Vector Machine* (SVM) dipilih sebagai model terbaik dalam penelitian ini karena memiliki performa tertinggi secara keseluruhan. SVM mampu mengklasifikasikan data dengan tingkat kesalahan yang relatif rendah, baik dalam bentuk *false positive* maupun *false negative*. Nilai *precision* yang tinggi menunjukkan bahwa prediksi positif yang dihasilkan sebagian besar benar, sedangkan nilai *recall* yang tinggi menunjukkan kemampuan model dalam mendeteksi sebagian besar komentar yang mengandung promosi judi online.

Dengan demikian, model SVM tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga menunjukkan keseimbangan performa yang baik.

4.4. Analisis Confusion Matrix



Gambar 3. *Confussion Matrix* SVM

Berdasarkan confusion matrix pada model SVM, diperoleh nilai sebagai berikut:

a. *True Negative* (TN): 1148
b. *False Positive* (FP): 20
c. *False Negative* (FN): 38
d. *True Positive* (TP): 840

Hasil ini menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Namun, masih terdapat kesalahan klasifikasi yang perlu diperhatikan. Kesalahan yang paling krusial adalah false negative, yaitu komentar promosi judi online yang tidak terdeteksi oleh model. Hal ini berpotensi menyebabkan konten ilegal tetap lolos dari sistem. Sementara itu, kesalahan *false positive* relatif kecil, sehingga model cukup baik dalam menghindari kesalahan klasifikasi terhadap komentar normal.

4.5. Pengaruh *Preprocessing* terhadap Model

Tahap *preprocessing* memiliki peran yang krusial dalam meningkatkan performa model klasifikasi, khususnya dalam pengolahan data teks yang cenderung tidak terstruktur dan mengandung banyak *noise*. Proses pembersihan teks, seperti penghapusan URL, angka, tanda baca, serta normalisasi spasi, secara signifikan berkontribusi dalam menyederhanakan representasi data sehingga lebih mudah diproses oleh model.

Dengan berkurangnya *noise*, distribusi kata dalam dokumen menjadi lebih konsisten, sehingga metode ekstraksi fitur seperti TF-IDF dapat menghasilkan representasi numerik yang lebih relevan. Hal ini berdampak langsung pada kemampuan model dalam mengidentifikasi pola-pola penting yang membedakan antara komentar promosi judi online dan komentar normal.

Preprocessing berperan dalam menyeleksi dan mengeliminasi fitur-fitur yang kurang relevan sehingga kompleksitas dimensi data dapat dikurangi. Kondisi ini berdampak pada meningkatnya efisiensi dalam proses pelatihan model, sekaligus menekan kemungkinan terjadinya *overfitting*. Oleh karena itu, penerapan *preprocessing* yang tepat dan terstruktur menjadi salah satu faktor penting dalam mendukung tercapainya kinerja klasifikasi yang optimal.

4.6. Analisis Kesalahan Klasifikasi (*Error Analysis*)

Tabel 3. *Misclassified Data*

Index	Teks	Actual	Predicted
33	sungguh ***D*** o r ***A*** ketagihan berat sulit berhenti	1	0
6321	terimakasih upload bg sibuk **WETON**	1	0
6127	duit malaysiarmbabickup blnja rumah satu tahun	0	1
4802	kualitas layanan di ρυłαü sangat baik	1	0

Index	Teks	Actual	Predicted
2183	gila hasilnya keren banget	0	1
4766	gue baru ngeh kalo p̤ṳl̤a̤ṳw̤i̤n̤ udah masuk semua media sosial	1	0
1432	yuhuu salam kalbar	0	1
932	alahh udah kayak paling sedunia ajaa	0	1
2886	bisa main fortnite ga bamh d	0	1
6753	🅤🅛🅤 hadiah	1	0

Kesalahan *false negative* terjadi pada komentar yang seharusnya diklasifikasikan sebagai promosi judi online, namun terdeteksi sebagai komentar normal. Contohnya adalah komentar yang menggunakan manipulasi karakter seperti *"𝘿 o r 𝘈"*, *"𝙒𝙀𝙏𝙊𝙉"*, atau *"ρυłαü"*. Penggunaan karakter khusus ini menyulitkan model dalam mengenali kata kunci yang relevan.

Meskipun model menunjukkan performa yang tinggi, masih terdapat beberapa kesalahan klasifikasi yang dapat dikategorikan menjadi *false negative* dan *false positive*.

Sementara itu, kesalahan *false positive* terjadi pada komentar normal yang diklasifikasikan sebagai judi online, seperti *"gila hasilnya keren banget"* atau *"yuhuu salam kalbar"*. Hal ini disebabkan oleh kemiripan pola teks dengan data pelatihan, sehingga model memberikan prediksi yang kurang tepat.

Secara umum, kesalahan klasifikasi dipengaruhi oleh beberapa faktor, yaitu penggunaan bahasa tidak baku, variasi penulisan, konteks yang ambigu, serta keterbatasan metode TF-IDF yang belum mampu memahami makna semantik secara mendalam.

4.7. Analisis Hasil Prediksi Model

Berdasarkan hasil prediksi yang diperoleh, model menunjukkan kemampuan yang baik dalam mengklasifikasikan komentar yang mengandung promosi judi online, meskipun terdapat variasi penulisan dan manipulasi karakter. Hal ini mengindikasikan bahwa model tidak hanya bergantung pada kemunculan kata kunci secara eksplisit, tetapi juga mampu menangkap pola distribusi kata yang menjadi ciri khas dari masing-masing kelas.

Kemampuan model dalam mengidentifikasi komentar dengan karakteristik tidak baku menunjukkan bahwa representasi fitur yang dihasilkan melalui TF-IDF cukup efektif dalam menangkap informasi penting dari teks. Namun demikian, performa model masih memiliki keterbatasan dalam memahami konteks semantik secara mendalam, terutama pada teks yang bersifat ambigu atau menggunakan *obfuscated text* yang kompleks.

Di sisi lain, model juga menunjukkan kemampuan yang baik dalam mengklasifikasikan komentar non-judi, yang mengindikasikan bahwa sistem memiliki tingkat generalisasi yang cukup tinggi. Hal ini penting untuk memastikan bahwa model tidak hanya sensitif terhadap pola tertentu, tetapi juga mampu membedakan konteks secara lebih luas.

Secara keseluruhan, hasil prediksi menunjukkan bahwa pendekatan yang digunakan telah mampu menangkap karakteristik utama dari data, meskipun masih terdapat ruang pengembangan, khususnya dalam meningkatkan pemahaman konteks teks yang lebih kompleks.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa metode *Natural Language Processing* (NLP) dan *machine learning* dapat digunakan secara efektif untuk membangun sistem klasifikasi teks pada komentar media sosial. Model yang dihasilkan sudah mampu bekerja dengan baik dan memberikan hasil yang konsisten, sehingga bisa dijadikan dasar untuk pengembangan sistem moderasi konten otomatis yang lebih luas. Selain itu, pendekatan yang digunakan juga cukup fleksibel untuk diterapkan pada kasus serupa dengan karakteristik data yang berbeda.

Sistem ini masih bisa ditingkatkan dengan mencoba model yang lebih memahami konteks bahasa, serta menambahkan variasi data agar model semakin adaptif terhadap perubahan pola teks. Pengembangan juga bisa diarahkan ke implementasi langsung pada sistem *real-time*, serta menggabungkan beberapa metode agar hasil deteksi menjadi lebih kuat dan stabil, terutama dalam menghadapi teks yang semakin bervariasi dan kompleks.

DAFTAR PUSTAKA

[1] K. Iansyah, A. L. Nurlaili, dan M. M. Al Haromainy, "Pengembangan Model IndoBERT untuk Deteksi Komentar Promosi Judi Online di YouTube," *SOBAT (Jurnal Teknologi Informasi)*, vol. 7, no. 1, pp. 292–302, 2025, doi: 10.32897/sobat.2025.7.1.5256

[2] S. H. Bakhtiar and A. N. Adilah, "Fenomena Judi Online : Faktor, Dampak, Pertanggungjawaban Hukum", Innovative, vol. 4, no. 3, pp. 1016–1026, May 2024, https://doi.org/10.31004/innovative.v4i3.10547

[3] A. García-Pérez, A. Krotter, dan G. Aonso-Diego, "The impact of gambling advertising and marketing on online gambling behavior: an analysis based on Spanish data," Public Health, vol. 234, pp. 170–177, Sep. 2024, doi: 10.1016/j.puhe.2024.06.025.

[4] A. N. Cahyo, R. Hermawan, M. A. D. Wijaya, dan A. P. Sari, "Deteksi Teks Promosi Judi Online Pada Kolom Komentar YouTube dengan Metode Regex dan Fuzzy Matching," RESTIKOM: Riset Teknik Informatika dan Komputer, vol. 7, no. 2, pp. 176–187, Aug. 2025, doi: 10.52005/restikom.v7i2.441.

[5] F. J. M. Riza, R. G. Guntara, dan M. R. Nugraha, "Implementasi Algoritma Naive Bayes untuk Filtrasi Spam Komentar Judi Online pada

YouTube," *Indonesian Journal of Digital Business*, vol. 5, no. 2, pp. 411–423, Jul. 2025, doi: 10.17509/ijdb.v5i2.88641.

[6] M. F. As Shidiq dan D. Alita, "Analisis Sentimen Masyarakat terhadap Kasus Judi Online Menggunakan Data dari Media Sosial X Pendekatan Naive Bayes dan SVM," vol. 8, no. 1, pp. 24–35, 2025.

[7] Sugiyo, A. B. Susanto, dan S. Anggai, "Analisis Eksperimental Kinerja Transformers, VADER, dan Naive Bayes dalam Analisis Sentimen Teks Bahasa Indonesia (Studi Kasus Komentar Terkait Judi Online)," vol. 3, no. 1, pp. 149–167, Jul. 2025

[8] J. P. Suharsono and D. Nurahman, "Pemanfaatan YouTube Sebagai Media Peningkatan Pelayanan dan Informasi," Ganaya: Jurnal Ilmu Sosial dan Humaniora, vol. 7, no. 1, pp. 298–304, 2024.

[9] D. D. N. Cahyo, A. Prasetyo, and Y. Nugroho,

[10] "Sentiment analysis of public opinion on online gambling using CNN and Word2Vec," Jurnal Sistem Informasi dan Informatika, 2025. [Online]. Available: https://ejournal.lppm-unbaja.ac.id/index.php/jsii/article/view/3624

[11] D. A. Zulfikar and F. Fitria, "Detecting online gambling promotion on social media with Random Forest," in Proceedings of the International Conference on Business and Information Technology Systems (ICoBITS), 2025. [Online]. Available: https://icobits.ubhinus.ac.id/index.php/ICoBITS/article/view/50

[12] W. M. Darwansah, D. Setiawan, and H. Prabowo, "Classification and mapping of online gambling based on news articles using named entity recognition and support vector machine," Journal of Informatics Technology and Operations Systems, 2025. [Online]. Available: https://ejournal.uniks.ac.id/index.php/JTOS/article/view/4707

[13] M. R. Septiansah, N. Putri, and A. Kurniawan, "Analisis komparatif algoritma one-class untuk klasifikasi teks iklan judi online berbasis embedding IndoBERT," RESTIKOM Journal, 2025. [Online]. Available: https://restikom.nusaputra.ac.id/index.php/restikom/article/view/restikom.v7i2.441

[14] S. Samuel and D. P. Kristiadi, "Deteksi teks promosi judi online menggunakan kombinasi natural language processing dan deep learning," Journal of Information Technology and Digital Engineering, 2025. [Online]. Available: https://journal.eng.unila.ac.id/index.php/jitet/article/view/5187

[15] M. R. Adrian, M. P. Putra, M. H. Rafialdy, dan N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,"*Jurnal Informatika UPGRIS*, vol. 7, no. 1, pp. 36–39, 2021. https://journal.upgris.ac.id/index.php/JIU/article/view/7099/4309

[16] S. Mahmuda, "Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube,"*Jurnal Jendela Matematika*, vol. 2, no. 1, pp. 21–31, Jan. 2024. https://doi.org/10.47065/JOSYC.V3I4.2154