

Penerapan Algoritma Random Forest dengan Pendekatan Hybrid Feature Engineering untuk Klasifikasi Spam Judi Online

Fitra Salam S. Nagalay¹, Fathurrahman Kurniawan Ikhsan^{2*}, Desi Rahma Aryanti³, Edwar Rosman⁴

¹ Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Wira Buana, Lampung, Indonesia

² Jurusan Teknologi Informasi, S1 Tr Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Lampung

³ Fakultas Sains dan Teknologi, Teknologi Informasi, Universitas Wira Buana, Lampung, Indonesia

⁴ Teknologi Informasi, D3 Teknik Komputer PSDKU Solok Selatan, Politeknik Negeri Padang, Padang, Indonesia

¹ fitrasalamsn@gmail.com, ^{2*} fathurrahman@polinela.ac.id, ³ desirahmaaryanti@gmail.com, ⁴ Edwar@pnp.ac.id

ABSTRACT – The proliferation of illegal online gambling spam in YouTube comments has become a serious problem that is difficult to address with conventional filters. This research aims to build an effective automated detection system using machine learning. The proposed method implements the Random Forest algorithm with an innovative Hybrid Feature Engineering approach. This approach not only analyzes lexical features from text using TF-IDF but also extracts 11 anomaly and behavioral features to capture spammer evasion tactics. To address the class imbalance issue in the dataset of 2,708 manually labeled comments, the SMOTE technique was applied to the training data. The model evaluation on the test data demonstrated very high performance, achieving an accuracy of 94.65% and an F1-Score of 0.89 for the "Judol" class. Feature importance analysis confirmed that engineered anomaly features, such as the brand-number pattern, significantly contributed to the model's success. This study proves that a hybrid approach can create a robust and accurate classification model to combat online gambling promotional content.

Keywords: Hybrid Feature Engineering; Online Gambling; Machine Learning; Text Classification; Random Forest.

ABSTRAK – Penyebaran spam judi online ilegal di kolom komentar YouTube telah menjadi masalah serius yang sulit diatasi oleh filter konvensional. Penelitian ini bertujuan untuk membangun sebuah sistem deteksi otomatis yang efektif menggunakan machine learning. Metode yang diusulkan menerapkan algoritma Random Forest dengan pendekatan Hybrid Feature Engineering yang inovatif. Pendekatan ini tidak hanya menganalisis fitur leksikal dari teks menggunakan TF-IDF, tetapi juga mengekstrak 11 fitur anomali dan perilaku untuk menangkap taktik penyamaran spammer. Untuk mengatasi masalah ketidakseimbangan kelas pada dataset yang terdiri dari 2.708 komentar berlabel manual, teknik SMOTE diterapkan pada data latih. Hasil evaluasi model pada data uji menunjukkan performa yang sangat tinggi, dengan akurasi mencapai 94.65% dan F1-Score sebesar 0.89 untuk kelas "Judol". Analisis feature importance mengonfirmasi bahwa fitur rekayasa anomali, seperti pola merek-angka, secara signifikan berkontribusi pada keberhasilan model. Penelitian ini membuktikan bahwa pendekatan hibrida mampu menciptakan model klasifikasi yang robust dan akurat untuk memerangi konten promosi judi online.

Kata Kunci: Hybrid Feature Engineering; Judi Online; Klasifikasi Teks; Machine Learning; Random Forest.

1. PENDAHULUAN

Perkembangan pesat dalam penggunaan internet yang ada di Indonesia telah mengalami peningkatan dalam penggunaan media sosial yang diakses dari *smartphone* menempatkan Indonesia peringkat tiga terbesar setelah negara China [1]. Hal tersebut telah menjadikan Youtube sebagai salah satu platform media sosial paling banyak penggunanya didunia, bahkan dikalangan masyarakat Indonesia sendiri. Lebih dari 10 hingga 20 juta video yang di upload ke dalam Youtube setiap minitnya [2]. Namun, popularitas ini dieksplorasi oleh pihak-pihak tidak bertanggung jawab untuk melakukan promosi ilegal, salah satunya adalah judi online. Dampak dari promosi judi online tidak bisa

dipandang sebelah mata, hal tersebut berdampak buruk pada aspek, mulai dari ekonomi, sosial, hingga psikologis [3].

Perjudian merupakan aktivitas mempertaruhan sejumlah uang, jika menang maka akan memperoleh seluruh uang taruhan tersebut, dan jika kalah maka semua yang dipertaruhan akan hilang [4]. Fenomena ini menjadi semakin meresahkan karena para pelaku promosi judol tidak lagi menyebarkan spam secara acak, melainkan menargetkan komunitas-komunitas tertentu dengan tingkat engagement yang tinggi. Komunitas penggemar permainan, misalnya, menjadi target utama karena demografi audiens yang muda dan aktif berinteraksi di kolom komentar dan live chat [5]. Para pelaku menggunakan berbagai taktik linguistik dan teknis yang



This work is licensed under a
Creative Commons Attribution 4.0 International License

semakin canggih untuk menyamarkan pesan promosi judi online tersebut, sebuah fenomena yang dikenal sebagai obfuscation (penyamaran teks). Taktik ini meliputi penggunaan jargon, modifikasi ejaan, dan substitusi karakter menggunakan karakter Unicode non-standar untuk menghindari deteksi [6]. Pola-pola teks yang tidak wajar ini membuat sistem filter spam bawaan platform, yang seringkali bergantung pada fitur blacklist atau pencocokan kata kunci sederhana, menjadi tidak efektif [7]. Deteksi otomatis terhadap komentar spam semacam ini menjadi tantangan yang signifikan dalam bidang Natural Language Processing. Diperlukan sebuah sistem yang tidak hanya mampu memahami konten leksikal (kata-kata) dari sebuah komentar, tetapi juga mampu mengenali pola-pola perilaku dan anomali yang menjadi ciri khas dari taktik spammer [8].

Penelitian ini bertujuan untuk membangun sebuah model klasifikasi machine learning yang efektif untuk mendeteksi komentar promosi judi online. Secara spesifik, penelitian ini menerapkan algoritma Random Forest yang dilatih menggunakan pendekatan Hybrid Feature Engineering. Pendekatan ini menggabungkan fitur tekstual konvensional yang diekstrak menggunakan TF-IDF dengan serangkaian fitur anomali yang direkayasa untuk menangkap sinyal-sinyal spam yang lebih halus, sebuah pendekatan yang terbukti efektif dalam meningkatkan performa model klasifikasi teks [9]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode deteksi konten negatif yang lebih robust dan akurat di ruang digital.

2. DASAR TEORI

Beberapa studi yang telah dilakukan sebelumnya telah memberikan kontribusi dalam membentuk dasar pemikiran dan metodologi yang digunakan dalam penelitian ini. Penelitian terdahulu yang mendasari analisis yang dilakukan dalam studi ini dan berfungsi sebagai referensi guna memperkaya pemahaman. Klasifikasi teks untuk analisis sentimen dan deteksi konten negatif pada media sosial merupakan bidang penelitian yang telah banyak dieksplorasi. Berbagai algoritma machine learning telah diterapkan untuk tugas ini. Sebagai contoh, penelitian oleh Febrian dan Alita dalam menganalisis sentimen terkait kasus judi online di media sosial X (sebelumnya Twitter) berhasil mencapai akurasi sebesar 83% dengan menggunakan Support Vector Machine (SVM), yang terbukti lebih unggul dibandingkan Naïve Bayes (79%) [10]. Hal ini menunjukkan bahwa topik judi online merupakan ranah yang relevan untuk dianalisis.

Meskipun demikian, pemilihan algoritma yang optimal tetap menjadi kunci. Sebuah studi komparatif yang lebih baru oleh Agustia dan Suryono secara langsung

membandingkan tiga algoritma untuk analisis sentimen judi online [4]. Hasil penelitian mereka menunjukkan bahwa algoritma Random Forest mencapai performa terbaik dengan akurasi 78%, mengungguli Naïve Bayes dan Logistic Regression. Keunggulan Random Forest dalam menangani interaksi kompleks antar fitur menjadikannya kandidat utama untuk penelitian ini.

Namun, performa sebuah algoritma tidak hanya bergantung pada algoritmanya itu sendiri, tetapi juga sangat dipengaruhi oleh kualitas fitur yang digunakan. Penelitian oleh Nur Fauzi dkk. Secara eksplisit menunjukkan bahwa penerapan *feature engineering* dan *hyperparameter tuning* dapat secara signifikan meningkatkan akurasi dari model Random Forest [11]. Argumen ini diperkuat oleh fakta bahwa para pelaku spam judi online modern menggunakan berbagai taktik penyamaran teks (obfuscation), seperti penggunaan karakter Unicode non-standar, untuk menghindari filter sederhana [6]. Oleh karena itu, diperlukan pendekatan rekayasa fitur yang lebih dari sekadar analisis kata kunci standar.

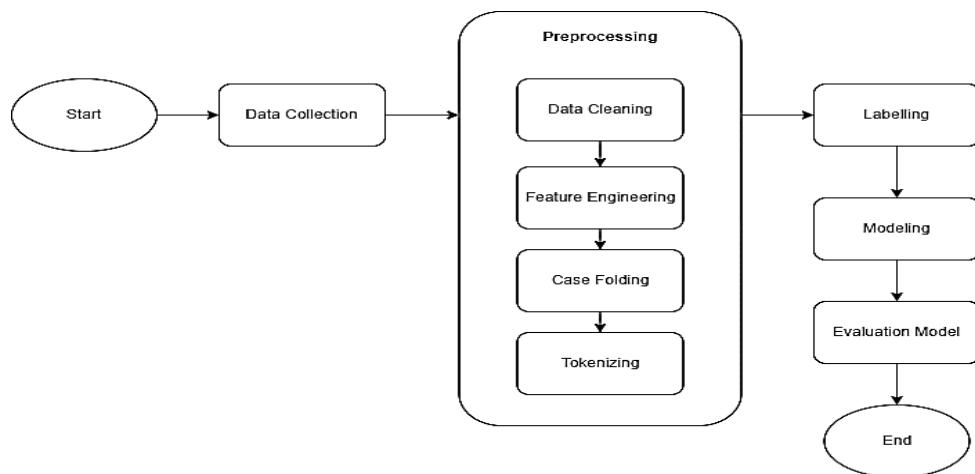
Dari beberapa penelitian tersebut, dapat disimpulkan bahwa meskipun SVM dan Naïve Bayes umum digunakan, sementara Random Forest menunjukkan potensi performa yang lebih unggul untuk domain sejenis kemudian dalam menangani masalah ketidakseimbangan kelas dan efisiensi dalam klasifikasi sentimen. Namun, untuk memaksimalkan potensinya, diperlukan pendekatan rekayasa fitur yang canggih. Banyak penelitian sebelumnya berfokus pada fitur leksikal, namun belum secara mendalam mengeksplorasi fitur-fitur anomali dan perilaku teks untuk menangkap taktik spammer.

Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan mengusulkan pendekatan *Hybrid Feature Engineering* [12]. Pendekatan ini menggabungkan kekuatan analisis leksikal (TF-IDF) dengan serangkaian fitur anomali yang dirancang khusus untuk mendeteksi pola promosi judi online. Dengan mengintegrasikan pendekatan ini dengan algoritma Random Forest yang terbukti kuat dan teknik SMOTE untuk menangani data tidak seimbang [13], penelitian ini diharapkan dapat menghasilkan model klasifikasi yang lebih akurat dan robust.

3. METODOLOGI

Proses penelitian ini dilakukan dengan beberapa tahapan utama, dimulai dari pengumpulan data hingga evaluasi model. Rangkaian tahapan penelitian ini ditampilkan pada Gambar 1.





Gambar 1. Tahapan Penelitian

1. Data Collection

Dataset yang digunakan dalam penelitian ini merupakan data yang belum pernah dipublikasikan sebelumnya dan dikumpulkan secara mandiri. Pengumpulan data dilakukan dengan teknik *web scrapping* menggunakan skrip kode python yang memanfaatkan *library Selenium*. Target akuisisi data adalah kolom komentar dari video-video youtube berbahasa Indonesia yang teridentifikasi memiliki tingkat keterlibatan audiens yang tinggi dan menjadi sasaran promosi judi online, terutama pada konten *gaming* dan *podcast*. Total data mentah yang berhasil dihimpun adalah sebanyak 3282 komentar, yang kemudian disimpan dalam format csv untuk tahapan selanjutnya.

2. Preprocessing Data

Preprocessing data adalah fase yang krusial dimana berujuan untuk menransformasi data mentah yang tidak terseruktur menjadi dataset yang bersih, terstruktur, dan siap untuk dimodelkan. Kualitas dari tahap ini akan memberikan pengaruh positif terhadap tingkat akurasi yang didapat [14]. Pada penelitian ini, tahap *preprocessing data* meliputi beberapa langkah yaitu, *data cleaning*, *feature engineering*, *case folding*, dan *tokenizing* sebagai bagian dari vektorisasi.

a. Data cleaning

Langkah awal dalam *preprocessing* adalah *data cleaning*. Tujuan utamanya adalah untuk meningkatkan kualitas dataset dengan menghilangkan inkonsistensi pada data. Dalam penelitian ini, prosesnya difokuskan pada penghapusan data duplikat. Komentar yang identik dihapus untuk memastikan bahwa setiap entri data bersifat unik dan untuk mencegah model menjadi bias terhadap sampel data yang berulang-ulang. Langkah ini merupakan praktik standar dalam persiapan data untuk memastikan integritas dataset [15].

b. Feature engineering

Penelitian ini menerapkan pendekatan *hybrid feature engineering* untuk mengekstrak sinyal secara komprehensif. Berbeda dengan pendekatan yang standar, fitur tidak hanya diekstrak dari teks bersih, tetapi juga dari teks asli yang mengandung anomali [6].

c. Case folding

Selanjutnya untuk keperluan analisis leksikal, sebuah kolom teks bersih dibuat. Proses ini melibatkan *case folding*, yaitu standarisasi teks dengan mengubah semua huruf menjadi huruf kecil untuk mengurangi kompleksitas kosakata [4].

d. Tokenizing

Tokenisasi adalah proses memcah rangkaian teks menjadi unit-unit yang lebih kecil yang disebut token, yang biasanya berupa kata atau tanda baca. Dalam penelitian ini, tokenisasi merupakan langkah implisit yang terjadi sebagai bagian dari proses vektorisasi teks menggunakan metode TF-IDF [4]. TF-IDF akan secara otomatis melakukan tokenisasi pada data latih untuk membangun kosakata, kemudian menggunakan kosakata yang sama untuk mentranformasi data uji, memastikan konsistensi dan validitas evaluasi model [16].

3. Labeling Data

Pada tahap ini setiap komentar dalam dataset diberi label kategorikal “Judol” yaitu judi online atau “Normal” berdasarkan serangkaian kriteria yang telah ditentukan. Proses ini menghasilkan *ground truth* yang menjadi dasar fundamental untuk pelatihan dan evaluasi model *supervised learning*.

4. Modeling

Tahapan pemodelan ini adalah inti dari proses klasifikasi otomatis dimana model akan belajar untuk membedakan antara dua kelas judol atau normal. Pada tahapan ini dataset final yang telah diberi label dibagi menjadi data latih (80%) dan data uji (20%). Selanjutnya, label kategorikal tersebut diubah menjadi representasi



biner 1 dan 0 melalui proses *label encoding* [17]. Pada data latih, fitur teks pada kolom komentar bersih diubah menjadi matriks numerik menggunakan TF-IDF Vectorizer. Matriks hasil ini kemudian digabungkan dengan 11 fitur numerik lainnya. Teknik oversampling SMOTE diterapkan hanya pada data latih untuk menyeimbangkan jumlah sampel antara kelas “Judssol” dan “Normal”[4]. SMOTE dapat mengatasi ketidak seimbangan kelas jumlah sampel antar kelas pada dataset, terutama jumlah sampel pada kelas minoritas sangat terbatas dibandingkan kelas mayoritas [16]. Kemudian peneliti menggunakan model klasifikasi *Random Forest*. Algoritma ini dipilih berdasarkan studi literatur yang menunjukkan performa unggulnya untuk tugas klasifikasi sejenis serta kemampuannya dalam memberikan analisis dan tingkat kepentingan sebuah fitur [18].

5. Evaluation model

Tahapan terakhir dari metodologi ini adalah mengevaluasi seberapa baik performa model yang telah dilatih. Metrik evaluasi yang digunakan adalah Precision, Recall, dan F1-Score, yang merupakan standar

pengukuran yang lebih informatif dan andal dibandingkan akurasi untuk kasus klasifikasi dengan data tidak seimbang. Confusion Matrix juga dibuat untuk menganalisis secara visual tipe-tipe kesalahan yang dibuat oleh model [19].

4. HASIL DAN PEMBAHASAN

1. Dataset

Pengumpulan dataset dilakukan dengan teknik web scraping pada platform Youtube. Fokus pengumpulan adalah pada kolom komentar dari 10 video konten permainan dan podcast berbahasa indonesia. Jumlah data mentah yang berhasil dikumpulkan sebanyak 3282 komentar, yang kemudian disimpan dengan format csv. Sampel data mentah yang berhasil dikumpulkan disajikan pada Tabel 1.

Tabel 1. Sampel Dataset

No	Komentar
1	Pengen punya PC kayak ALEXIS17, biar nge-game lancar jaya.
2	1:01 semoga PROBE 855 terus jaya ya ka channel nya
3	Kayaknya kalo lo belum kenal PULAUWIJN, lo belum hidup di 2025.
4	Seperti biasa sehat selalu kang semangat ngonten dan dilancarkan segala urusannya Salam dari WETON88
5	Pengen punya PC kayak ALEXIS17, biar nge-game lancar jaya.

2. Data Preprocessing

1. Data Cleaning

Pada tahapan pembersihan data yang dilakukan yaitu menghapus data duplikat untuk meningkatkan kualitas dataset. Komentar yang identik sepenuhnya, yang bisa saja muncul akibat aktivitas spam terkoordinasi atau kesalahan saat scrapping dihilangkan untuk memastikan setiap entri data bersifat unik. Dataset yang awalnya berjumlah 3282 komentar setelah melalui proses

pembersihan duplikasi menjadi total 2.708 komentar unik.

2. Feature Engineering

Setelah dataset dibersihkan dari duplikasi, tahap selanjutnya adalah *feature engineering*. Pendekatan ini bertujuan untuk mengekstrak fitur-fitur informatif dari teks asli untuk menangkap sinyal anomali dan perilaku spammer. Sebanyak 11 fitur numerik direkayasa dari setiap komentar. Deskripsi setiap fitur yang diekstraksi disajikan secara rinci pada Tabel 2.

Tabel 2. Deskripsi Fitur Yang Diekstraksi Dari Tiap Komentar

No	Nama Fitur	Deskripsi
1	jumlah_karakter_aneh	Menghitung jumlah total karakter yang bukan merupakan huruf, angka, atau spasi (contoh: !, @, #,\$).
2	persentase_kapital	Menghitung rasio persentase huruf kapital terhadap total jumlah huruf dalam satu komentar.
3	jumlah_angka	Menghitung jumlah total digit angka yang muncul dalam komentar (0-9).
4	mengandung_font_aneh	Mendeteksi keberadaan karakter <i>full-width</i> (Contoh: D E W A) yang sering digunakan untuk menghindari filter. Bernilai 1 jika ada, dan 0 jika tidak ada.



5	mengandung_timestamp	Mendeteksi keberadaan pola waktu. (Contoh: 05:27) yang sering digunakan untuk membuat spam terlihat seperti komentar otentik.
6	ratio_karakter_unik	Menghitung rasio antara jumlah karakter unik dengan total panjang komentar. Nilai yang rendah dapat mengindikasikan teks yang repetitif.
7	mengandung_pola_merek_angka	Mendeteksi keberadaan pola teks angka (Contoh: SITUS123) yang merupakan ciri khas penamaan situs judol.
8	jumlah_kata_kunci_judol	Menghitung frekuensi kemunculan kata kunci dari daftar yang telah ditentukan sebelumnya (Contoh: gacor, wd)
9	mengandung_angka_berulang	Mendeteksi keberadaan angka yang sama secara berurutan (Contoh: 88, 777) yang sering diassociasi dengan keberuntungan
10	jumlah_emoji	Menghitung jumlah total emoji yang ada di dalam komentar menggunakan <i>library emoji</i> .

Dengan mengekstraksi fitur-fitur ini, dataset yang awalnya tidak terstruktur diubah menjadi format matriks yang terstruktur.

3. Case folding

Case folding adalah proses standardisasi teks yang mengubah semua huruf menjadi huruf kecil. Proses ini

diterapkan untuk membuat kolom baru yaitu komentar_bersih yang akan digunakan untuk analisis leksikal. Tujuan utama adalah untuk memastikan kata yang sama namun dengan kapitalisasi berbeda, misalnya “Gacor” dan “gacor” dianggap sebagai satu entitas yang sama oleh model. Contoh proses ini ditunjukkan pada Tabel 3.

Tabel 3. Contoh Proses Case Folding

No.	Sebelum	Sesudah
1	0:02 Wah keran SGI 88 terbaik menang mudah.	002 wah keran sgi88 terbaik menang mudah
2	Seperti biasa sehat selalu kang semangat ngonten dan dilancarkan segala urusannya Salam dari WETON88	seperti biasa sehat selalu kang semangat ngonten dan dilancarkan segala urusannya salam dari weton88
WETON88		

4. Tokenizing

Pada proses tokenisasi ini bertujuan untuk memecah kalimat menjadi unit-unit kata. Tujuan dari proses ini

adalah untuk mempersiapkan teks agar dapat dianalisis dan diubah menjadi representasi numerik oleh model. Contoh hasil proses tokenisasi dijelaskan pada Tabel 4.

Tabel 4. Contoh Proses Tokenisasi

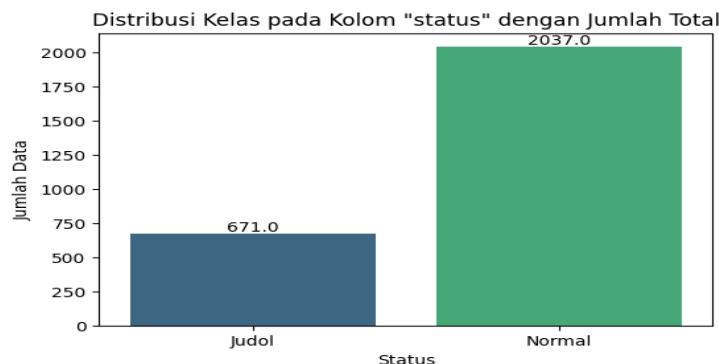
No.	Sebelum	Sesudah
1	002 wah keran sgi88 terbaik menang mudah	['002', 'wah', 'keran', 'sgi88', 'terbaik', 'menang', 'mudah']
2	seperti biasa sehat selalu kang semangat ngonten dan dilancarkan segala urusannya salam dari weton88	['seperti', 'biasa', 'sehat', 'selalu', 'kang', 'semangat', 'ngonten', 'dan', 'dilancarkan', 'segala', 'urusannya', 'salam', 'dari', 'weton88']

5. Labeling

Selanjutnya adalah proses pelabelan data. Proses ini merupakan tahap krusial karena disinilah *ground truth* untuk model *supervised learning* diciptakan. Setiap baris komentar dalam dataset diperiksa secara manual oleh peneliti untuk menentukan kategorinya. Berdasarkan serangkaian kriteria yang telah ditentukan untuk mengidentifikasi promosi, setiap komentar diberi satu

dari dua label, “Judol” untuk komentar yang terindikasi sebagai promosi judi online, dan “Normal” untuk komentar yang biasa saja. Dari total 2708 komentar, hasil pelabelan menunjukkan adanya distribusi kelas yang tidak seimbang, dimana 671 Komentar (24.8%) diklasifikasikan sebagai “Judol” dan 2037 komentar (75.2%) sebagai “Normal”. Distribusi final dari dataset berlabel ini disajikan secara visual pada Gambar 2.



**Gambar 2.** Distribusi Kelas

6. Evaluasi Model

Dataset yang telah disiapkan dan dilabeli pada tahap sebelumnya digunakan untuk melatih dan menguji model klasifikasi. Proses dimulai dengan membagi dataset tersebut menjadi data latih 80% dan data uji 20%. Selanjutnya dalam mengatasi masalah ketidakseimbangan

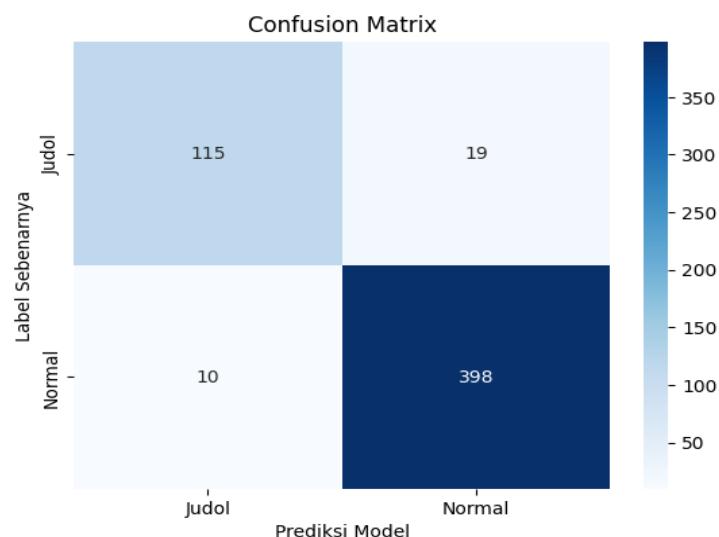
kelas pada data latih, teknik *oversampling* SMOTE diterapkan. Kemudian, model *Random Forest* dilatih menggunakan data latih yang telah seimbang tersebut. Kinerja model kemudian dievaluasi secara kuantitatif dan kualitatif menggunakan data uji. Hasil evaluasi disajikan dalam laporan klasifikasi pada Tabel 5 berikut.

Tabel 5. Hasil Klasifikasi Model Random Forest

Matrik Evaluasi	Nilai
Accuracy	94.65%
Precision	92%
Recall	86%
F1-Score	89%

Berdasarkan tabel 5, model menunjukkan performa yang sangat kuat dengan akurasi keseluruhan mencapai 94.65%. Fokus utama pada kelas minoritas, yaitu "Judol", menunjukkan f1-score sebesar 89%, yang mengidikasikan keseimbangan yang sangat baik antara precision 92% dan

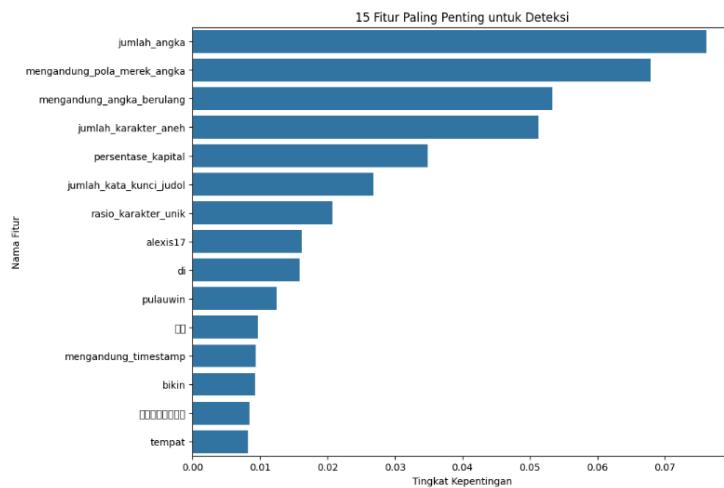
recall 86%. Nilai recall yang tinggi tersebut menandakan bahwa model berhasil mengidentifikasi sebagian besar komentar judol yang ada, sementara nilai precision yang tinggi memastikan bahwa prediksi judol yang dibuat oleh model sangat dapat diandalka



Gambar 3. Confusion Matrix

Confusion matrix pada gambar 3 memvisualisasikan hasil dari 134 komentar "Judol" yang sebenarnya, model berhasil menangkap 115 *true positive* dan hanya melewatkannya 19 komentar atau *false negative*. Kemudian model juga sangat akurat dalam mengidentifikasi komentar "Normal", dengan hanya 10 kesalahan atau *false positive*.

Selanjutnya untuk memahami faktor-faktor yang mendorong performa tinggi, dilakukan analisis terhadap tingkat kepentingan fitur yang diekstrak dari model *random forest*. Gambar 4 menampilkan 15 fitur teratas yang paling berpengaruh dalam proses pengambilan keputusan model.

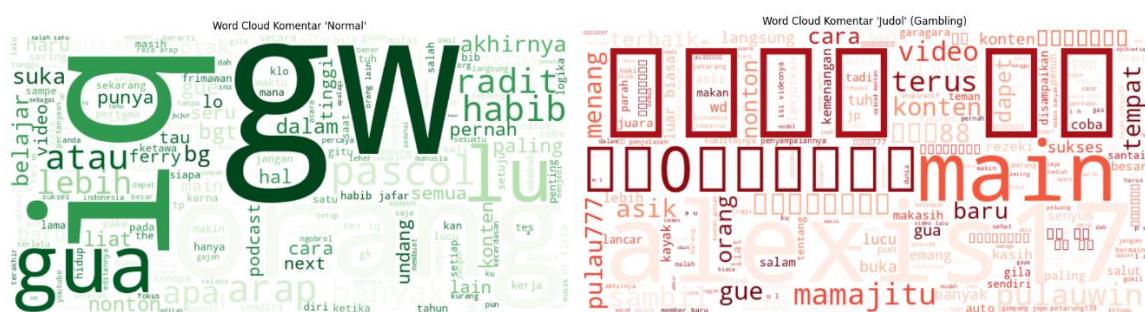


Gambar 4. Grafik tingkat kepentingan fitur tertinggi

Hasil analisis pada gambar 4 memberikan wawasan yang sangat berharga. Terlihat jelas bahwa fitur-fitur yang direkayasa secara manual mendominasi peringkat atas, yang mengonfirmasi efektivitas pendekat *hybrid feature engineering*. Fitur mengandung_pola_merek_angka muncul sebagai prediktor paling kuat. Hal ini menunjukkan bahwa pola penamaan situs judi yang menggabungkan nama merek dengan angka seperti SITUSJUDOL88 adalah ciri khas yang paling mudah dikenali dan paling diandalkan oleh model untuk mengidentifikasi komentar spam.

Fitur-fitur anomali dan perilaku lainnya seperti rasio_karakter_unik, jumlah_karakter_aneh, dan presentase_kapital juga menunjukkan kontribusi yang sangat signifikan. Ini membuktikan hipotesis penelitian bahwa bagaimana sebuah komentar ditulis dengan penggunaan karakter non standar, kapitalisasi berlebihan, dan struktur yang tidak wajar seringkali menjadi sinyal yang lebih kuat daripada kata kunci yang spesifik.

Sebagai bukti kualitatif pendukung, Gambar 5 menyajikan perbandingan *word cloud* antara komentar “Normal” dan “Judol”.



Gambar 5. Perbandingan word cloud komentar Normal (kiri) dan Judol (kanan).

Perbandingan visual pada gambar tersebut mengonfirmasi temuan kuantitatif sebelumnya. *Word cloud* "Judol" (kanan) didominasi oleh nama-nama merek, jargon promosi yang relevan, bahkan karakter-karakter unik, sementara *word cloud* "Normal" (kiri) berisi kata-kata yang sesuai dengan konteks interaksi yang wajar. Visualisasi ini secara efektif merangkum perbedaan fundamental dalam kosakata yang digunakan oleh kedua kelas.

5. KESIMPULAN

Penelitian ini berhasil membangun dan mengevaluasi sebuah model *machine learning*. Evaluasi kuantitatif menunjukkan bahwa model Random Forest yang diusulkan memiliki kinerja yang sangat kuat dengan tingkat akurasi mencapai 94.65%. fokus utama pada kelas "Judol" menunjukkan hasil yang sangat baik, dimana model berhasil mencapai f1-score sebesar 89%, yang mengindikasikan keseimbangan yang solid antara precision sebesar 92% dan recal 86%.

Performa tinggi ini didorong oleh efektivitas pendekatan *Hybrid Feature Engineering*, yang divalidasi melalui analisis *feature importance*. Hasil analisis menunjukkan bahwa fitur-fitur rekayasa anomali dan perlilaku secara konsisten mendominasi peringkat atas sebagai prediktor terkuat. Fitur mengandung_pola_merek_angka menjadi faktor paling signifikan. Temuan tersebut membuktikan bahwa cara sebuah komentar ditulis seringkali menjadi simbol yang lebih kuat daripada kata-kata spesifik yang digunakannya.

Temuan kuantitatif ini diperkuat lebih lanjut secara kualitatif melalui visualisasi perbandingan *word cloud*. Komentar berlabel "Judol" secara jelas didominasi oleh nama-nama merek seperti alexis17, weton88 dan jargon promosi (cair, rezeki, menang). Hal ini menunjukkan perbedaan yang sangat kontras komentar berlabel "Normal" yang didominasi oleh kata-kata yang relevan dengan konteks video (keren, game, undang). Kombinasi dari hasil kuantitatif dan kualitatif ini secara meyakinkan menunjukkan bahwa metodologi yang diusulkan berhasil membangun sebuah model yang nilai akurasinya besar, tetapi juga mampu mengidentifikasi pola-pola spam yang kompleks.enomoran angka. Jika diperlukan, dapat juga berikan saran untuk penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] E. Borgia, "The in[1] H. Junawan and N. Laugu, "Eksistensi Media Sosial , Youtube , Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia 1 Hendra A . PENDAHULUAN Perkembangan teknologi media yang sangat pesat serta dengan kemajuan teknologi-teknologi yang semakin har," *Baitul Ulum J. Ilmu Perpust. dan Inf.*, vol. 4, no. 1, pp. 41–57, 2020.
- [2] F. Prasetya, S. S. Embuningtiyas, and D. Andriyanto, "Youtube Sebagai Media Pembelajaran Di Indonesia," *J. Pendidik. Teknol. Inf.*, vol. 5, no. 2, pp. 192–202, 2022, doi: 10.37792/jukanti.v5i2.600.
- [3] F. J. E. Sipayung and C. A. Handoyo, "Dampak Dalam Mempromosikan Iklan Judi Online (Studi Kasus Iklan Judi Online Indonesia)," *Jurnal Intelek Dan Cendikian Nusantara*, vol. 1, no. 3, pp. 4548–4553, 2024.
- [4] I. T. Jadidah *et al.*, "Analisis maraknya judi online di Masyarakat," *Jurnal Ilmu Sosial Dan Budaya Indonesia*, vol. 1, no. 1, pp. 20–27, 2023.
- [5] Universitas Gadjah Mada, "Judi Online Makin Marak di Kalangan Anak Muda, Pakar UGM Sarankan Perlunya Edukasi Literasi Keuangan," <https://ugm.ac.id/id/berita/judi-online-makin-marak-di-kalangan-anak-muda-pakar-ugm-sarankan-perlunya-edukasi-literasi-keuangan/> (accessed Aug. 01, 2025).
- [6] A. Uchendu, T. Le, and D. Lee, "Attribution and Obfuscation of Neural Text Authorship: A Data Mining Perspective," *ACM SIGKDD Explor. Newsl.*, vol. 25, no. 1, pp. 1–18, 2023, doi: 10.1145/3606274.3606276.
- [7] A. A. Abdo *et al.*, "AI-based Spam Detection Techniques for Online Social Networks: Challenges and Opportunities," *J. Internet Serv. Inf. Secur.*, vol. 13, no. 3, pp. 78–103, 2023, doi: 10.58346/JISIS.2023.I3.006.
- [8] A. Muhammin, I. A. Taufik, and D. D. Daniswara, "Pendeteksian Spam pada E-mail menggunakan Pendekatan Natural Language Processing," *Pros. Semin. Nas. Sains Data*, vol. 3, no. 1, pp. 116–121, 2023, doi: 10.33005/senada.v3i1.90.
- [9] N. P. R. Adiati *et al.*, "Comparative study of predictive models for hoax and disinformation detection in indonesian news," *Int. J. Adv. Intell. Informatics*, vol. 10, no. 3, pp. 504–516, 2024, doi: 10.26555/ijain.v10i3.878.
- [10] M. F. As Shidiq and D. Alita, "Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data Dari Media Sosial X Pendekatan Naive Bayes Dan Svm," *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 24–35, 2025, doi: 10.47080/simika.v8i1.3624.
- [11] N. P. Nur Fauzi, S. Khomsah, and A. D. Putra Wicaksono, "Penerapan Feature Engineering dan Hyperparameter Tuning untuk Meningkatkan Akurasi Model Random Forest pada Klasifikasi



- Risiko Kredit," J. Teknol. Inf. dan Ilmu Komput., vol. 12, no. 2, pp. 251–262, 2025, doi: 10.25126/jtiik.2025128472.
- [12] N. Nyoman Eny Perimawati, R. Rudolf Huizen, D. Pramana Hostiadi, and M. Sistem Informasi, "Analisa Pengaruh Pre-Procesing Data Untuk Model Deteksi Akun Palsu Pada Media Sosial," vol. 2, no. 1, p. 2025, 2025.
- [13] R. Cherekar, "Automated Data Cleaning: AI Methods for Enhancing Data Quality and Consistency," Int. J. Emerg. Trends Comzzzput. Sci. Inf. Technol., vol. 5, no. 1, pp. 31–40, 2024, doi: 10.63282/3050-9246.ijetcsit-v5i1p104.
- [14] Dwi Nanda Agustia and Ryan Randy Suryono, "Comparison of Naïve Bayes, Random Forest, and Logistic Regression Algorithms for Sentiment Analysis Online Gambling," INOVTEK Polbeng - Seri Inform., vol. 10, no. 1, pp. 284–295, 2025, doi: 10.35314/prk93630.
- [15] U. S. Shanthamallu and A. Spanias, "Introduction to Machine Learning," Synth. Lect. Signal Process., vol. 01, pp. 1–8, 2022, doi: 10.1007/978-3-031-03758-0_1.
- [16] M. Guntara and F. D. Astuti, "Komparasi Kinerja Label-Encoding dengan One-Hot-Encoding pada Algoritma K-Nearest Neighbor menggunakan Himpunan Data Campuran," JIKO (Jurnal Inform. dan Komputer), vol. 9, no. 2, p. 352, 2025, doi: 10.26798/jiko.v9i2.1605.
- [17] C. M. F. Andriani and D. Susilaningrum, "Klasifikasi Waiting Time for Pilot di Pelabuhan Tanjung Perak Menggunakan Metode Regresi Logistik - Synthetic Minority Oversampling Technique (SMOTE)," J. Sains dan Seni ITS, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.109844.
- [18] C. Azzaria, E. Daniati, and A. Ristyawan, "Peningkatan Akurasi Deteksi Liver Disease melalui Hyperparameter Tuning pada Algoritma Random Forest," vol. 4, no. 2, pp. 139–147, 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.

