JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Chrome Extension Deteksi Spam Judi Online Pada Platform YouTube Menggunakan Ekstraksi Fitur dan Machine Learning

Muhammad Taufiq Ramadhan[1], Muhammad Naufal[2*], Feri Agustina[3]

[1, 2]Teknik Informatika, Fakultas Teknologi Cerdas, Universitas Dian Nuswantoro,
Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah
Email: [1]111202214552@mhs.dinus.ac.id, [2]m.naufal@dsn.dinus.ac.id, [3]feri.agustina@dsn.dinus.ac.id
Email Penulis Korespondensi: m.naufal@dsn.dinus.ac.id
Submitted **26-06-2026**; Accepted **23-10-2026**; Published **31-10-2026**

Abstrak

Penelitian mengklasifikasikan chat judi online pada fitur live chat YouTube menggunakan machine learning, karena pesan spam promosi judi sulit dimoderasi secara manual akibat volumenya yang tinggi dan bersifat real-time. Model klasifikasi teks dibangun dan dibandingkan menggunakan ekstraksi fitur Count Vectorizer dan TF-IDF yang dikombinasikan dengan Support Vector Machine (SVM) dan Random Forest pada data live chat YouTube berlabel biner "JUDOL" dan "NORMAL". Evaluasi menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan Stratified K-Fold Cross Validation. Kata seperti "slot", "gacor", "freebet", dan "wisdomtoto" menjadi indikator penting pesan judi online. Seluruh kombinasi menghasilkan akurasi di atas 99,8% pada Stratified 10-Fold Cross Validation dengan perbedaan antar-metode dalam kisaran simpangan baku. Count Vectorizer dengan SVM memperoleh rata-rata F1-score CV tertinggi (0,9990 ± 0,0010) serta akurasi dan F1-score 99,92% pada data uji independen. Model terbaik dideploy pada Chrome Extension yang terhubung dengan REST API lokal, dengan latensi inferensi rata-rata 0,46 ms per pesan, throughput 2.164 pesan per detik, dan latensi end-to-end 3,19 ms pada lingkungan 1 vCPU. Sistem tetap stabil hingga 200 pesan per detik sehingga mendukung penyaringan spam pada siaran langsung bervolume tinggi. Namun, repetisi template spam pada dataset dapat memengaruhi tingginya performa, sehingga generalisasi terhadap pola spam baru perlu diuji lebih lanjut.

Kata Kunci: Judi Online ; Live Chat Youtube; Klasifikasi Teks; Count Vectorizer; TF-IDF

Abstract

This study classifies online gambling ("judol") chats in YouTube's live chat using machine learning, as gambling-promotion spam is difficult to moderate manually because of its high volume and real-time nature. Text classifiers were built and compared using Count Vectorizer and TF-IDF feature extraction combined with Support Vector Machine (SVM) and Random Forest, trained on YouTube live chat data labeled "JUDOL" or "NORMAL". Evaluation used accuracy, precision, recall, F1-score, confusion matrix, and Stratified K-Fold Cross Validation. Words such as "slot", "gacor", "freebet", and "wisdomtoto" emerged as key indicators of gambling messages. All combinations achieved accuracy above 99.8% under Stratified 10-Fold Cross Validation, with differences between methods within the standard deviation. Count Vectorizer with SVM obtained the highest mean cross-validation F1-score (0.9990 ± 0.0010) and reached 99.92% accuracy and F1-score on an independent test set. The best model was deployed in a Chrome Extension connected to a local REST API, achieving an average inference latency of 0.46 ms per message, a throughput of 2,164 messages per second, and an end-to-end latency of 3.19 ms on a 1 vCPU environment. The system remained stable at up to 200 messages per second, supporting spam filtering in high-volume live streams. However, repeated spam templates in the dataset may have inflated performance, so generalization to new spam patterns requires further testing.

Keywords: Online Gambling; Youtube Live Chat; Text Classification; Count Vectorizer; TF-IDF

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi di era transformasi digital telah mengubah paradigma interaksi sosial masyarakat secara masif. Platform berbasis media sosial kini tidak hanya berfungsi sebagai media konsumsi konten yang pasif, melainkan telah berevolusi menjadi ruang komunikasi sinkronus yang dinamis[1]. YouTube merupakan salah satu platform media sosial dengan jumlah pengguna terbesar di dunia, serta menyediakan fitur live streaming yang terintegrasi dengan live chat sehingga memungkinkan interaksi real-time antara kreator dan audiens[2]. Namun, tingginya volume dan kecepatan pertukaran pesan dalam ekosistem digital ini memicu tantangan baru berupa kerentanan terhadap penyalahgunaan platform oleh pihak-pihak yang tidak bertanggung jawab. Salah satu anomali komunikasi yang kian marak dan menjadi ancaman serius pada kolom live chat YouTube adalah penyebaran promosi judi online secara ilegal melalui mekanisme pesan spam. Fenomena judi online telah berkembang menjadi permasalahan sosial dan ekonomi yang serius di Indonesia karena berpotensi menimbulkan kerugian finansial, gangguan kesehatan mental, serta berbagai konsekuensi sosial yang secara signifikan memengaruhi masyarakat, khususnya kelompok usia muda [3], [4]. Para pelaku penyebar spam memanfaatkan karakteristik live chat yang serba cepat dengan menyisipkan kata-kata kunci khusus (leksikon khas) seperti "slot", "gacor", "maxwin", dan "scatter" untuk menarik perhatian audiens secara agresif. Penetrasi pesan ilegal yang tidak terkendali ini tidak hanya merusak kenyamanan pengguna, tetapi juga mempersulit pengawasan konten digital karena metode penyebaran yang adaptif dan masif.

Upaya moderasi konten yang sepenuhnya mengandalkan manusia semakin sulit diterapkan pada platform digital berskala besar karena keterbatasan kapasitas moderator dalam memproses volume pesan yang terus meningkat dan berlangsung secara real-time. Kondisi ini menyebabkan pendekatan moderasi manual menjadi kurang efisien serta sulit diskalakan untuk memenuhi kebutuhan pengawasan konten modern [5], [6]. Fenomena serupa juga ditemukan pada berbagai platform digital yang menghasilkan data teks dalam jumlah besar, di mana kapasitas pemrosesan manual sering

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1465
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

kali tidak mampu mengimbangi volume data yang masuk sehingga diperlukan metode otomatis untuk membantu proses pengelompokan dan pengelolaan informasi [7]. Oleh karena itu, implementasi sistem moderasi otomatis berbasis pembelajaran mesin (machine learning) dengan pendekatan klasifikasi teks (text classification) menjadi sebuah urgensi teknologi yang mutlak. Dalam proses klasifikasi teks, tahapan ekstraksi fitur (feature extraction) merupakan komponen fundamental yang berfungsi mentransformasikan data teks yang bersifat tidak terstruktur menjadi representasi numerik atau vektor fitur yang dapat diproses oleh algoritma machine learning. Kualitas representasi fitur yang dihasilkan akan sangat memengaruhi performa model klasifikasi yang dibangun [8]. Seiring dengan perkembangan pemrosesan bahasa alami, terdapat berbagai metode ekstraksi fitur yang dapat diimplementasikan, mulai dari pendekatan tradisional seperti N-grams, hingga representasi vektor padat (dense vector) yang mampu menangkap konteks semantik seperti Word2Vec, GloVe, dan FastText. Meskipun metode-metode lanjutan tersebut menawarkan pemahaman konteks yang lebih baik, pendekatan berbasis frekuensi kata masih menjadi pilihan utama dalam banyak kasus karena kesederhanaan dan efektivitasnya. Dua metode ekstraksi fitur yang paling sering digunakan dalam pemrosesan bahasa alami adalah Count Vectorizer dan Term Frequency-Inverse Document Frequency (TF-IDF)[9], [10]. Pemilihan di antara kedua metode ini menjadi aspek krusial karena perbedaan karakteristik representasi data tersebut akan berdampak langsung pada akurasi dan efisiensi komputasi model saat menangani data real-time.

Berdasarkan kesenjangan dan kebutuhan tersebut, penelitian ini bertujuan untuk menguji, membangun, serta melakukan analisis perbandingan (komparasi) performa model klasifikasi teks otomatis dalam mendeteksi chat judi online pada fitur live chat YouTube. Eksperimen ini secara spesifik membandingkan efektivitas penggunaan metode Count Vectorizer dan TF-IDF sebagai teknik ekstraksi fitur, Masing-masing representasi fitur teks kemudian dapat dipadukan dengan berbagai algoritma machine learning untuk membangun model klasifikasi. Algoritma populer meliputi Support Vector Machine (SVM) dan Random Forest, yang dikenal tangguh dalam menangani data berdimensi tinggi. SVM mencari hyperplane optimal untuk memisahkan kelas data, sedangkan Random Forest (RF) menggunakan ensemble pohon keputusan untuk mengeksploitasi interaksi fitur yang kompleks dan meningkatkan stabilitas prediksi. Selain itu, algoritma lain juga dapat digunakan sesuai karakteristik dataset dan tujuan analisis. Misalnya, Logistic Regression menawarkan pendekatan probabilistik yang sederhana namun efektif untuk klasifikasi biner. k-Nearest Neighbors (k-NN) mengandalkan kedekatan sampel dalam ruang fitur untuk menentukan kelas, cocok untuk dataset dengan pola lokal yang jelas. Naive Bayes memanfaatkan asumsi independensi antar fitur untuk menghitung probabilitas kelas, sering digunakan pada teks karena efisien dan stabil terhadap dimensi tinggi. Algoritma berbasis neural network, seperti Multi-Layer Perceptron (MLP), mampu menangkap pola non-linear kompleks antar fitur. Pemilihan algoritma yang tepat, atau kombinasi beberapa algoritma, sangat menentukan akurasi, generalisasi, dan efisiensi komputasi model dalam menangani data teks real-time. Dataset yang digunakan dalam eksperimen ini bersumber dari data mentah kolom chat YouTube yang telah melalui proses pembersihan komprehensif dan dikategorikan secara biner ke dalam label "JUDOL" dan "NORMAL". Guna menjamin objektivitas, stabilitas, dan validitas hasil evaluasi kinerja model, penelitian ini menerapkan metode Stratified K-Fold Cross Validation. Sebagai luaran praktis dari penelitian ini, kombinasi model dan ekstraksi fitur dengan performa terbaik diintegrasikan ke dalam sebuah prototype ekstensi peramban yang berfungsi melakukan penyaringan dan penyembunyian pesan judi online secara real-time langsung pada antarmuka YouTube. Sebagian besar penelitian terdahulu mengenai deteksi pesan spam judi online terfokus pada analisis teks statis pada media sosial seperti X/Twitter, SMS, atau ulasan produk secara offline, tanpa mempertimbangkan karakteristik obrolan langsung (live chat) yang sarat manipulasi teks sengaja (*obfuscation*), seperti pemisahan huruf antar-spasi dan substitusi angka leet. Selain itu, riset sebelumnya jarang menjangkau integrasi sistem langsung pada antarmuka pengguna akhir (*end-user level*). Kebaruan (*novelty*) dari penelitian ini terletak pada penggabungan modul pra-pemrosesan *de-obfuscation* khusus leksikon judi online dengan komparasi Count Vectorizer dan TF-IDF, serta perancangan purwarupa ekstensi peramban (Chrome Extension) berbasis REST API yang mampu melakukan penyaringan dan penyorotan konten spam secara real-time langsung pada elemen DOM live chat YouTube.

Pemilihan algoritma SVM dan RF dalam studi ini didasarkan pada ketangguhan keduanya dalam menangani data berdimensi tinggi yang dihasilkan oleh proses ekstraksi fitur teks. Representasi teks menggunakan Count Vectorizer dan TF-IDF umumnya menghasilkan ruang fitur berdimensi tinggi dengan tingkat sparsitas yang tinggi. Dalam kondisi tersebut, SVM dikenal mampu menemukan hyperplane pemisah yang optimal melalui prinsip maximum margin sehingga sering digunakan sebagai algoritma klasifikasi pada berbagai tugas text mining dan natural language processing[11], [12]. Di sisi lain, RF menawarkan pendekatan ensemble learning yang terbukti tangguh terhadap overfitting serta mampu memetakan interaksi antar fitur yang kompleks [13], [14]. Dengan mengkomparasikan kedua algoritma ini terhadap dua metode ekstraksi fitur yang berbeda, penelitian ini dapat memetakan secara presisi kombinasi mana yang paling optimal bukan hanya dari segi akurasi prediktif, melainkan juga tingkat efisiensi komputasional yang menjadi syarat mutlak bagi pemrosesan data berkecepatan tinggi. Sebagai penegasan kontribusi ilmiah, penelitian ini menjembatani kesenjangan antara evaluasi teoretis machine learning dan implementasi praktisnya di lingkungan produksi. Jika sebagian besar studi pemrosesan bahasa alami berfokus murni pada optimalisasi metrik performa di lingkungan simulasi, penelitian ini memadatkan proses tersebut hingga tahap deployment. Integrasi model klasifikasi terbaik ke dalam bentuk Chrome Extension membuktikan bahwa intervensi langsung terhadap pesan spam judi online dapat dilakukan secara mandiri di sisi klien. Dengan demikian, luaran dari penelitian ini tidak sekadar memperkaya literatur komparasi metode ekstraksi fitur pada teks live chat, tetapi juga menghadirkan alat bantu proaktif dan terukur dalam menciptakan ekosistem media sosial yang lebih bersih, aman, dan kondusif bagi masyarakat.

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

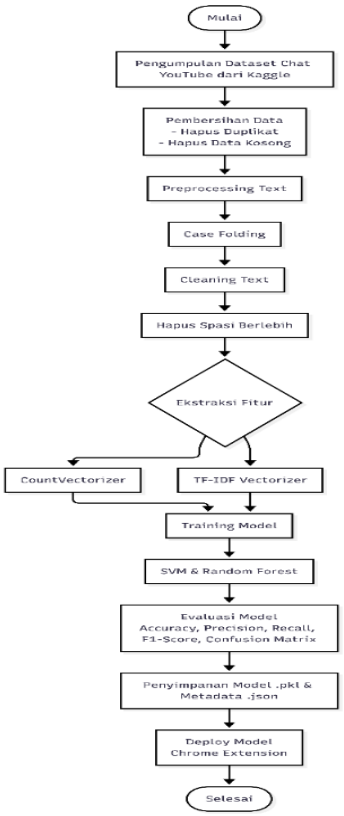
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2. METODOLOGI PENELITIAN

Bab ini menguraikan pendekatan komprehensif dan sistematis yang diterapkan untuk menjawab rumusan masalah serta mencapai tujuan penelitian, yakni membangun model klasifikasi teks untuk mendeteksi pesan judi *online*. Desain penelitian ini mengadopsi pendekatan eksperimental kuantitatif dalam ranah *machine learning*. Fokus metodologi diarahkan pada simulasi komparatif antara dua teknik ekstraksi fitur, yakni *Count Vectorizer* dan TF-IDF, yang dipadukan dengan dua algoritma klasifikasi yaitu SVM dan RF. Seluruh rangkaian metode dirancang sedemikian rupa guna memastikan keandalan, reproduktibilitas, serta validitas hasil pengujian pemodelan pada dataset *chat* YouTube.

2.1 Tahapan Penelitian

Pelaksanaan penelitian ini dirancang melalui serangkaian tahapan (*pipeline*) terstruktur yang berkesinambungan, mulai dari akuisisi data mentah hingga penyimpanan model akhir. Alur kerja komprehensif ini secara visual direpresentasikan pada diagram alir Gambar 1, dengan penjabaran proses sebagai berikut:



Gambar 1. Kerangka Kerja Penelitian

Rangkaian eksperimen dimulai dengan pengumpulan dataset obrolan YouTube dari Kaggle [15], yang kemudian diolah melalui tahap pembersihan dan *preprocessing* teks untuk menormalisasi data. Teks yang telah terstandarisasi selanjutnya diekstraksi menjadi vektor numerik menggunakan metode komparatif *Count Vectorizer* dan TF-IDF, lalu dilatih secara paralel menggunakan algoritma SVM dan RF. Performa dari setiap kombinasi model dan fitur tersebut kemudian dievaluasi menggunakan validasi silang *Stratified K-Fold* berdasarkan metrik akurasi, presisi, *recall*, *F1-score*, serta *confusion matrix* untuk menentukan pendekatan yang paling optimal. Rangkaian penelitian ini diakhiri dengan tahap penyimpanan model terbaik (.pkl) beserta metadatanya (.json) guna keperluan implementasi sistem di masa mendatang.

2.2 Pengumpulan Data

Tahap awal yang paling krusial dalam siklus pengembangan model pembelajaran mesin adalah akuisisi data yang representatif terhadap domain masalah. Pada penelitian ini, dataset yang digunakan merupakan dataset publik (public dataset) yang diperoleh melalui platform repositori Kaggle [15]. Berkas data tersebut memuat rekam riwayat pesan obrolan langsung (live chat) yang dihimpun dari siaran langsung salah satu agenda budaya tahunan berskala besar di Indonesia, yaitu acara Wayang Jogja Night Carnival #9 yang diselenggarakan dalam rangka Perayaan HUT ke-268 Kota Yogyakarta di platform YouTube. Secara struktural, dataset mentah ini memiliki format tabel terstruktur dengan total

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1467
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

volume awal sebanyak 6.350 baris observasi yang memuat beberapa atribut informasi esensial, meliputi waktu pengiriman pesan (datetime), nama identitas pengguna (author_name), isi dokumen teks asli (message), teks pasca-pembersihan awal (cleaned_message), serta target klasifikasi biner (label). Validitas ground truth dibangun melalui mekanisme pelabelan terawasi (supervised labeling) dengan verifikasi manual berbasis leksikon spesifik promosi perjudian (seperti keberadaan tautan eksternal, ajakan deposit, serta istilah khas slot/kasino online) untuk menetapkan label 1 (JUDOL), dan pesan apresiasi budaya atau komunikasi kondusif dari penonton untuk menetapkan label 0 (NORMAL). Sebelum data teks dialirkan ke dalam tahapan pemrosesan bahasa alami, dilakukan audit kualitas data yang mendeteksi keberadaan 2 (dua) baris data kosong (null values) pada kolom pesan teks. Setelah dua baris data kosong tersebut dieliminasi, dataset akhir yang digunakan secara konsisten berjumlah 6.348 baris data bersih, yang terdiri dari 3.187 data kelas JUDOL dan 3.161 data kelas NORMAL. Distribusi ini menunjukkan karakteristik sebaran kelas yang relatif seimbang (hampir seimbang), sehingga secara teoretis mengeklusi risiko bias algoritma terhadap kelas mayoritas selama proses pembelajaran terawasi [16].

2.3 Preprocessing Text

Karakteristik pesan obrolan pada fitur live chat YouTube umumnya bersifat tidak terstruktur, dipenuhi oleh variasi gaya bahasa informal, penggunaan simbol dekoratif, serta kesalahan pengetikan (typo). Tingginya noise linguistik ini berpotensi merusak matriks fitur dan menurunkan akurasi algoritma klasifikasi. Oleh karena itu, tahapan preprocessing text mutlak diperlukan untuk menstandarisasi bentuk dokumen pada kolom pesan sebelum diekstraksi menjadi vektor numerik. Rangkaian pemrosesan awal dalam penelitian ini diotomatisasi melalui tiga langkah linier sebagai berikut:

1. Case Folding & Normalisasi Substitusi Karakter: Langkah fundamental ini diawali dengan mengubah seluruh teks menjadi huruf kecil (lowercase). Selanjutnya, dilakukan penanganan terhadap teknik manipulasi karakter (LeetSpeak) yang sering digunakan spammer untuk menghindari deteksi. Sistem diprogram untuk mengonversi angka atau simbol yang menyerupai huruf kembali menjadi alfabet (0→o, 1→i, 3→e, 4→a, 5→s, 7→t, @→a, dan !→i).
2. Cleaning Text: Tahap ini bertujuan menyaring elemen data yang tidak memiliki nilai kontribusi prediktif terhadap klasifikasi judi online. Pembersihan dilakukan dengan pendekatan RegEx untuk menghapus seluruh karakter non-alfabetik, termasuk angka tersisa, simbol matematika, emoji, tanda baca, serta tautan URL eksternal, sehingga menyisakan murni deretan huruf kata demi kata.
3. Penghapusan Spasi Berlebih: Pada modus penyebaran spam judi online, pelaku sering kali menggunakan teknik obfuscation dengan menyisipkan spasi di setiap huruf dalam satu kata (misalnya "S L O T") untuk mengelabui filter deteksi platform. Tahap ini bekerja secara spesifik menggunakan RegEx untuk merekonstruksi kata-kata yang terpecah tersebut kembali utuh, sekaligus mereduksi spasi ganda menjadi spasi tunggal standar agar maknanya dapat ditangkap oleh algoritma ekstraksi fitur.

2.4 Ekstraksi Fitur dan Pembagian Data

Dataset teks yang telah distandarisasi kemudian dialirkan ke dalam dua skenario metode ekstraksi fitur yang berbeda untuk dibandingkan efektivitasnya, terdiri dari Count Vectorizer dan TF-IDF. Untuk mencegah terjadinya kebocoran data (data leakage), proses inisialisasi dan fitting pembobotan fitur (fit) dibatasi secara ketat hanya pada subset data training. Data pengujian (test set) maupun lipatan validasi (validation fold) hanya ditransformasikan (transform) menggunakan kamus kosakata dan bobot frekuensi yang telah dipelajari dari data training tanpa mengkalkulasi ulang parameter fitur dari data uji.

1. Count Vectorizer
   Sebagai implementasi dari model Bag-of-Words, Count Vectorizer bekerja dengan membangun kosakata dari seluruh korpus teks, kemudian merepresentasikan setiap dokumen sebagai vektor numerik berdasarkan frekuensi kemunculan kata. Setiap dimensi pada vektor merepresentasikan suatu term dalam kosakata, sedangkan nilai yang terkandung menunjukkan jumlah kemunculan term tersebut pada dokumen yang bersangkutan. Representasi ini memungkinkan data teks yang bersifat tidak terstruktur diolah lebih lanjut oleh algoritma klasifikasi maupun analisis teks [17].

$$X(d,t) = f(t,d) \quad (1)$$

2. TF-IDF Vectorizer
   Metode TF-IDF (*Term Frequency–Inverse Document Frequency*) memberikan pendekatan pembobotan statistik yang lebih dinamis dibandingkan representasi berbasis frekuensi murni. Selain mempertimbangkan frekuensi kemunculan

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

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

suatu kata dalam dokumen atau pesan tertentu (*Term Frequency*), metode ini juga memperhitungkan distribusi kata tersebut pada keseluruhan korpus dokumen melalui komponen *Inverse Document Frequency*. Dengan mekanisme tersebut, kata-kata yang sering muncul pada suatu dokumen tetapi relatif jarang ditemukan pada dokumen lain akan memperoleh bobot yang lebih tinggi karena dianggap memiliki kemampuan yang lebih baik dalam membedakan karakteristik suatu dokumen . Dalam konteks penelitian ini, kata-kata yang secara spesifik berkaitan dengan promosi judi online berpotensi memperoleh bobot yang lebih tinggi dibandingkan kata-kata umum yang banyak muncul pada berbagai jenis pesan [18]

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t) \quad (2)$$

Guna menjaga efisiensi komputasi, kedua metode ekstraksi fitur dikonfigurasi dengan membatasi jumlah kosakata menjadi 5.000 fitur paling representatif melalui parameter *max_features=5000*.

3. Pembagian Data
   Pembagian data dirancang secara sistematis dengan menggabungkan metode hold-out dan Stratified 10-Fold Cross Validation dengan pemisahan fungsi yang tegas. Pertama, seluruh dataset bersih (6.348 observasi) dibagi menggunakan metode stratified hold-out dengan proporsi 80% data latih (5.078 data) dan 20% data uji (1.270 data). Subset data latih sebesar 80% digunakan untuk proses pelatihan dan pemilihan model (model selection) melalui skema Stratified 10-Fold Cross Validation guna menguji stabilitas internal model dan menghitung standar deviasi performa antar-fold. Sementara itu, subset data uji sebesar 20% yang belum pernah dilihat sama sekali oleh model (unseen data) dialokasikan secara eksklusif untuk evaluasi akhir (final evaluation) guna mengukur kemampuan generalisasi model pada kondisi riil. Parameter stratify diterapkan secara konsisten pada seluruh tahapan pemisahan untuk memastikan proporsi kelas JUDOL dan NORMAL tetap seimbang. Pemilihan komposisi tersebut sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa penggunaan lebih dari 70% data untuk pelatihan cenderung menghasilkan performa klasifikasi yang lebih baik [19]

2.5 *Modelling*

Setelah fase representasi fitur numerik selesai dikonstruksi, alur penelitian berlanjut pada tahap implementasi algoritma klasifikasi. Eksperimen ini menerapkan dua algoritma pembelajaran mesin yang memiliki karakteristik kuat dalam menangani klasifikasi teks berdimensi tinggi secara paralel, yaitu SVM dan RF:

1. Support Vector Machine (SVM)
   Algoritma SVM dipilih karena keunggulannya yang konsisten dalam mencari hyperplane pemisah linier maupun non-linier yang optimal dengan margin maksimal di antara dua kelas data teks. Guna memastikan replikasi eksperimen, konfigurasi hyperparameter SVM ditetapkan secara spesifik menggunakan kernel Linear, parameter penalti keteraturan C = 1.0, parameter penyesuaian bobot kelas class_weight="balanced" untuk memberikan bobot proporsional terhadap variansi sampel, kriteria konvergensi toleransi 1e-4, batas iterasi maksimum max_iter = 10000, serta random_state = 42.

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \quad (3)$$

$$\hat{y} = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \quad (4)$$

$$w_j = \frac{N}{K \times N_j} \quad (5)$$

2. Random Forest (RF)
   Sebagai metode ensemble learning berbasis pohon keputusan (decision trees), RF diterapkan untuk mengeksploitasi interaksi fitur non-linier dan meningkatkan stabilitas prediksi melalui agregasi bootstrap. Konfigurasi hyperparameter Random Forest yang digunakan meliputi jumlah pohon keputusan n_estimators = 100, kriteria pemisahan Gini impurity (criterion="gini"), kedalaman pohon maksimal tanpa batasan (max_depth = None) hingga setiap daun murni, jumlah sampel minimum untuk membelah simpul internal min_samples_split = 2, jumlah sampel minimum pada daun min_samples_leaf = 1, pemanfaatan seluruh inti prosesor (n_jobs = -1), serta penguncian seed acak random_state = 42 untuk menjamin reproduktifitas hasil pengujian.

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1469
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), (x), \ldots, h_T(x)\} \quad (6)$$

2.6 Evaluasi dan Penyimpanan Model

Tahap akhir dari metodologi ini adalah melakukan pengujian matematis untuk mengukur validitas dan keandalan prediksi dari kombinasi fitur dan model yang telah dibangun. Proses evaluasi menggunakan metode Stratified K-Fold Cross Validation dengan konfigurasi 10 fold. Pemilihan jumlah *fold* tersebut bertujuan untuk menghasilkan estimasi performa yang lebih robust melalui pemanfaatan seluruh data secara bergantian sebagai data pelatihan dan data validasi, sekaligus menjaga keseimbangan yang baik antara bias dan variansi hasil evaluasi [20]. Selain itu, penerapan mekanisme *stratified* memastikan bahwa proporsi kelas "JUDOL" dan "NORMAL" pada setiap *fold* tetap merepresentasikan distribusi data asli, sehingga hasil pengujian menjadi lebih konsisten dan representatif [21]. Metrik evaluasi standar yang digunakan untuk mengukur kinerja komparatif meliputi empat parameter utama, yaitu *Accuracy* (mengukur ketepatan prediksi global), *Precision* (mengukur ketepatan penandaan kelas positif judi *online*), *Recall* (mengukur sensitivitas model dalam menjaring seluruh pesan judi *online* tanpa ada yang terlewat), serta *F1-Score* (nilai harmoni/keseimbangan antara *Precision* dan *Recall*) [22] . Selain itu, instrumen visual berupa *Confusion Matrix* digunakan untuk memetakan frekuensi kesalahan prediksi, baik dalam bentuk *False Positive* maupun *False Negative*.

Setelah analisis komparatif performa selesai dilakukan dan kombinasi model terbaik telah ditentukan, alur eksperimen ditutup dengan proses purifikasi sistem. Model dengan performa tertinggi diekspor dan disimpan ke dalam penyimpanan lokal dalam format biner .pkl memanfaatkan pustaka joblib. Bersamaan dengan itu, seluruh dokumentasi nilai metrik evaluasi hasil pengujian disimpan dalam berkas metadata terpisah berformat .json. Penyimpanan terstruktur ini bertujuan agar model final siap dipanggil kembali secara instan (*deployment*) tanpa perlu melalui proses pelatihan ulang, untuk selanjutnya diintegrasikan ke dalam arsitektur sistem moderasi otomatis berbasis ekstensi peramban (*Chrome Extension*).

2.7 Deployment Chrome Extension

Alur kerja pengembangan ekstensi peramban pada penelitian ini dilakukan melalui beberapa tahapan sistematis. Tahap pertama adalah pendefinisian konfigurasi utama melalui pembuatan file manifest.json, yang memuat metadata, izin akses (permissions), serta penentuan komponen ekstensi. Tahap kedua berfokus pada pengembangan antarmuka pengguna (User Interface) berupa popup, menggunakan kombinasi popup.html, popup.css, dan popup.js untuk menangani interaksi pengguna. Tahap selanjutnya adalah implementasi fungsionalitas inti yang terdiri dari content script (content.js dan content.css) untuk berinteraksi dan memanipulasi elemen pada halaman web aktif (DOM), serta background script (background.js) untuk menangani proses di latar belakang dan mengelola komunikasi antar komponen. Tahap terakhir adalah pengujian fungsionalitas secara lokal menggunakan mode pengembang (developer mode) pada peramban untuk memastikan seluruh komponen dan integrasi sistem berjalan dengan baik.

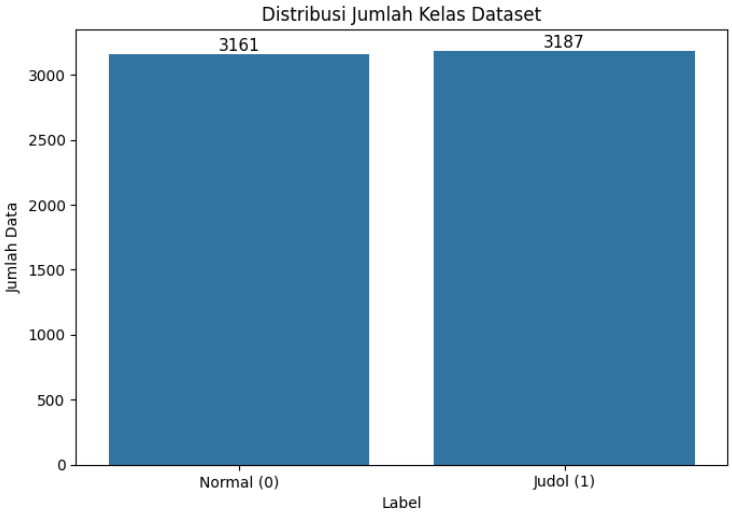
3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi Data

Tahap pertama dalam analisis hasil eksperimen diawali dengan pemetaan karakteristik struktural dari data yang berhasil dihimpun. Dataset tersebut memuat total volume sebanyak 6.350 baris observasi. Sebelum diintegrasikan ke dalam *pipeline* pelatihan, dilakukan audit kualitas data (*data auditing*) untuk mendeteksi keberadaan entri yang tidak valid. Hasil pemeriksaan struktural menunjukkan tidak ditemukannya baris data duplikat secara keseluruhan, namun terdeteksi adanya 2 (dua) baris data kosong (*null values*) pada kolom pesan teks utama. Penanganan secara instan dilakukan dengan mengeliminasi entri kosong tersebut, sehingga menghasilkan total baris data bersih sebanyak 6.348 observasi yang siap dialirkan ke tahapan berikutnya. Sifat akurasi dari model pembelajaran mesin sangat dipengaruhi oleh keseimbangan sebaran data antar-kelas. Berdasarkan hasil ekstraksi label biner, dataset ini memiliki tingkat keseimbangan yang relatif seimbang. Kategori pesan yang mengandung promosi perjudian *online* (Label 1) berjumlah 3.187 data, sedangkan pesan komunikasi interaksi normal (Label 0) berjumlah 3.161 data.

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1470
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 2. Jumlah Data Tiap Kelas

Keseimbangan sebaran dokumen secara biner ini memberikan keuntungan metodologis yang besar. Algoritma klasifikasi dapat mempelajari bobot fitur dari kedua belah kelas secara adil, sehingga meminimalkan risiko bias prediksi terhadap kelas mayoritas sejak awal fase pelatihan model.

3.2 Preprocessing Teks

Tahapan *preprocessing text* memegang peranan krusial dalam mereduksi varians linguistik dan membersihkan data tekstual dari *noise* konseptual. Standardisasi dokumen dievaluasi secara bertahap melalui tiga proses pembersihan yang saling memengaruhi, dengan hasil sebagai berikut:

1. *Case Folding*
   Proses ini berhasil menyeragamkan seluruh variasi kapitalisasi teks menjadi bentuk huruf kecil (*lowercase*) standar. Langkah ini secara signifikan mengompresi keanekaragaman indeks kosakata unik dalam korpus teks tanpa menghilangkan makna kontekstualnya.

Tabel 1. Hasil Casefolding

Sebelum	Sesudah
Saya Nonton Apa Ini Tribun Tv Kenapa Acara Wayang Tribun Night Carnival Tak Kunjung Dimulai	saya nonton apa ini tribun tv kenapa acara wayang tribun night carnival tak kunjung dimulai
JAM 4 TAPI PEMBUKAAN	jam 4 tapi pembukaan
HBD JOGJA	hbd jogja
selamat ulang tahu Jogja aku semoga Jogja semakin berubah semoga Jogja tetap istimewa Yogyakarta	selamat ulang tahu jogja aku semoga jogja semakin berubah semoga jogja tetap istimewa yogyakarta

2. Reduksi *Noise* dan *Leet Speak*
   Karakter-karakter non-alfabetik yang tidak membawa nilai informasi prediktif, seperti angka, tanda baca, emoji, dan simbol dekoratif, berhasil dieliminasi menggunakan pendekatan *Regular Expression*. Di samping itu, fungsi ini secara cerdas menyelesaikan kendala manipulasi karakter berbasis kemiripan visual (*LeetSpeak*) dengan mengonversi angka atau simbol kembali ke huruf aslinya (seperti "0" menjadi "o", "4" menjadi "a", dan "@" menjadi "a").

Tabel 2. Hasil Cleaning Text

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1471
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Sebelum	Sesudah
udah lewat 22 menit ni	udah lewat menit ni
L1nk G4c0r M4xw1n h4r1 1n1	link gacor maxwin hari ini

3. Penghapusan Spasi Berlebih

Proses penyaringan tahap akhir ini berfokus pada perbaikan anomali jarak antarkata yang sengaja dimodifikasi oleh pelaku spam. Pola penulisan yang renggang direkonstruksi secara otomatis menjadi jarak spasi tunggal yang baku. Selain mengoptimalkan kualitas sintaksis data, langkah ini berhasil memulihkan susunan kata yang sengaja dipecah untuk mengelabui filter penyaringan, sehingga meningkatkan kesiapan teks untuk dibaca oleh mesin pembobot fitur.

Tabel 3. Hasil Pemulihan Struktur Kata

Sebelum	Sesudah
udah lewat menit ni	udah lewat menit ni
S.L.O.T G.A.C.O.R P.A.S.T.I J.P	slot gacor pasti jp
D-E-P-O-S-I-T-O murah meriah menang jutaan	deposito murah meriah menang jutaan

3.3 Evaluasi Hasil Tahap Visualisasi Data

Visualisasi WordCloud dilakukan untuk mengetahui kata-kata yang paling sering muncul pada dataset divisualikan pada Gambar 2 berikut.

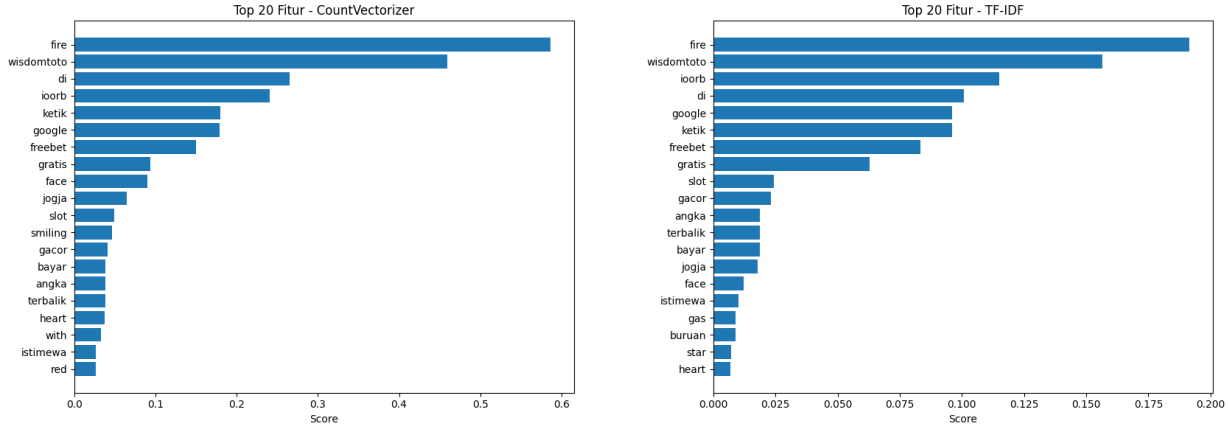


Gambar 3. Wordcloud Pada Dataset

Berdasarkan hasil WordCloud pada gambar 3, terlihat bahwa kata seperti “wisdomtoto”, “ketik”, “google”, “freebet”, “slot”, dan “gaSSSScor” memiliki ukuran yang lebih besar dibandingkan kata lainnya. Hal ini menunjukkan bahwa kata-kata tersebut sering digunakan dalam promosi judi online pada live chat YouTube. Kata “wisdomtoto” muncul sebagai kata paling dominan karena banyak chat mengandung promosi menuju situs tertentu. Selain itu, pola kalimat seperti “ketik wisdomtoto di google” juga sering ditemukan pada data spam judi online. Kata “freebet”, “slot”, dan “gacor” merupakan istilah yang umum digunakan dalam promosi perjudian online untuk menarik perhatian pengguna. Visualisasi WordCloud membantu dalam memahami pola linguistik pada dataset sehingga dapat diketahui kata-kata yang menjadi indikator utama dalam klasifikasi chat judi online. Hasil visualisasi ini juga mendukung proses ekstraksi fitur menggunakan Count Vectorizer karena kata-kata dengan frekuensi tinggi akan memiliki kontribusi besar terhadap proses pembelajaran model. Hasil menunjukkan bahwa promosi judi online memiliki pola penggunaan kata tertentu yang berulang dan dapat dimanfaatkan dalam proses klasifikasi. Selain WordCloud, dilakukan visualisasi bobot fitur Count Vectorizer untuk melihat kata-kata yang memiliki kontribusi terbesar terhadap model.

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1472
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 4. Perbandingan Ekstraksi Fitur

Gambar di atas menunjukkan perbandingan 20 fitur teratas yang dihasilkan oleh dua metode ekstraksi teks, yaitu Count Vectorizer dan TF-IDF. Metode Count Vectorizer menekankan frekuensi kemunculan kata dalam korpus, sehingga kata seperti fire, wisdomtoto, dan di memiliki skor tertinggi. Sementara itu, TF-IDF menilai kata berdasarkan seberapa informatif kata tersebut dalam konteks dokumen, bukan hanya frekuensinya, sehingga kata seperti slot, gacor, dan angka muncul lebih menonjol dibandingkan Count Vectorizer. Perbedaan ini menunjukkan bahwa Count Vectorizer efektif untuk mengidentifikasi kata yang paling sering muncul, sedangkan TF-IDF lebih mampu menangkap kata yang memberikan kontribusi informasi unik pada dokumen. Kombinasi kedua metode dapat memberikan gambaran lengkap mengenai kata kunci yang dominan dan relevan dalam analisis teks.

3.4 Modelling dan Evaluasi

Disini akan dibahasi hasil dari model yang dibuat baik pada saat pelatihan maupun pengujian. Hasil pelatihan dapat dilihat pada Tabel 4 berikut.

Tabel 4. Perbandingan Hasil Pelatihan K-Fold

Fitur Ekstraktor	Model	Acc Mean ± SD	F1 Mean ± SD
Count Vectorizer	SVM	0,9992 ± 0,0000	0,9992 ± 0,0000
Count Vectorizer	Random Forest	0,9992 ± 0,0000	0,9992 ± 0,0000
TF-IDF	SVM	0,9976 ± 0,0008	0,9976 ± 0,0008
TF-IDF	Random Forest	0,9992 ± 0,0000	0,9992 ± 0,0000

Pada Tabel 4 ditunjukkan hasil evaluasi Stratified 10-Fold Cross Validation pada data training (5.078 data). Seluruh kombinasi fitur ekstraktor dan model klasifikasi menghasilkan performa yang sangat tinggi dan konsisten. Kombinasi Count Vectorizer dengan SVM memperoleh nilai akurasi rata-rata sebesar 0,9992 dengan standar deviasi (SD) sebesar ±0,0000, serta F1-score rata-rata sebesar 0,9992 ±0,0000. Nilai yang identik juga diperoleh pada kombinasi Count Vectorizer dengan Random Forest serta TF-IDF dengan Random Forest. Sementara itu, kombinasi TF-IDF dengan SVM menghasilkan performa sedikit lebih rendah, yaitu akurasi rata-rata 0,9976 ±0,0008 dan F1-score rata-rata 0,9976 ±0,0008. Nilai standar deviasi yang sangat kecil pada seluruh skenario menunjukkan bahwa stabilitas model terjaga secara prima di seluruh lipatan pengujian. Capaian performa yang mendekati 100% ini perlu dianalisis secara kritis terhadap karakteristik dataset. Audit data lebih lanjut menunjukkan bahwa dari 3.187 baris pesan berlabel JUDOL, hanya terdapat 36 variasi template pesan unik yang dikirimkan secara masif dan berulang oleh akun bot spam sepanjang siaran langsung. Karakteristik pesan spam bot tersebut memiliki leksikon promosi yang sangat seragam dan kontras terhadap obrolan penonton normal. Tingginya keterpisahan fitur (separability) antara kedua kelas serta tingginya repetisi pola pesan spam inilah yang menjadi faktor utama tingginya akurasi klasifikasi pada dataset ini. Hasil pengujian pada data uji independen (test set) dapat dilihat pada Tabel 5.

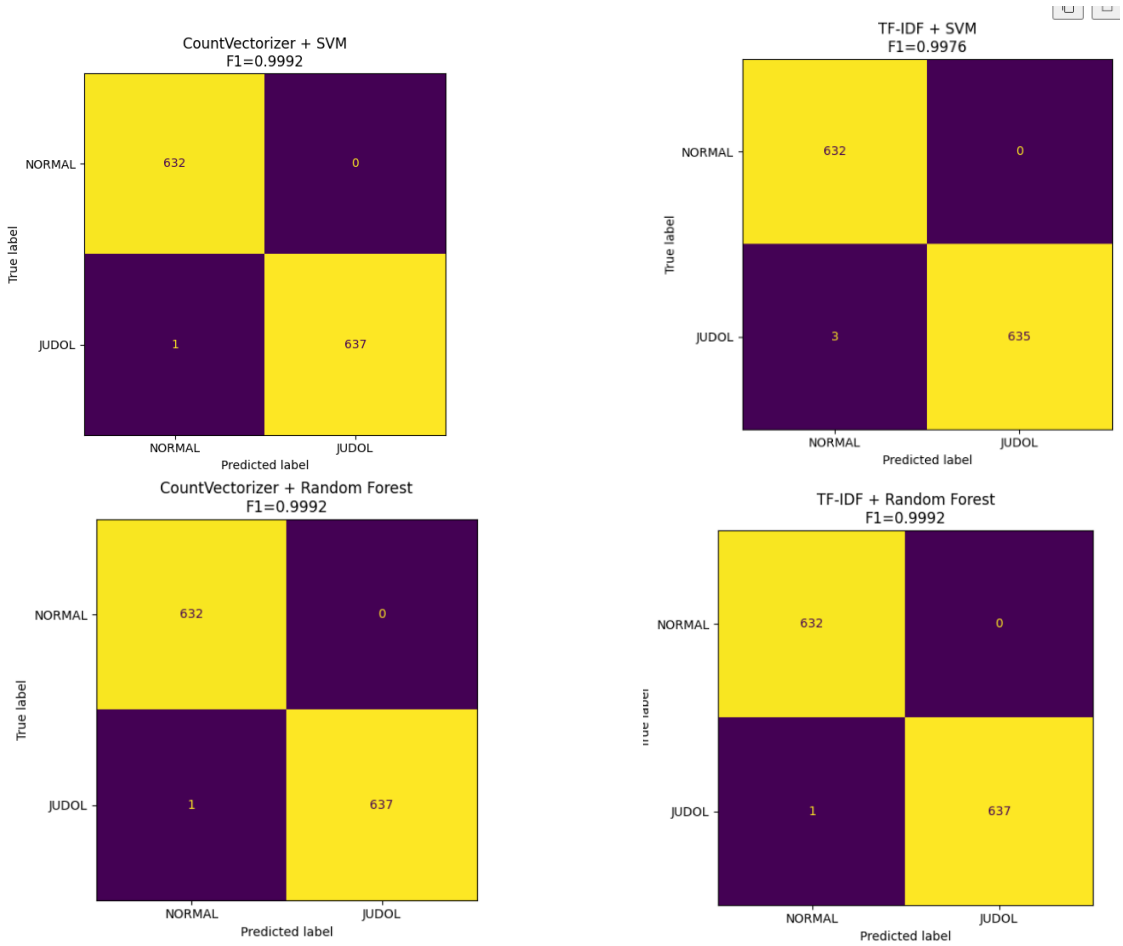
Tabel 5. Hasil Testing Model

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1473
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Fitur Ekstraktor	Model	Accuracy	Precision	Recall	F1 Score
Count Vectorizer	SVM	0,9992	1,0000	0,9984	0,9992
Count Vectorizer	RF	0,9992	1,0000	0,9984	0,9992
TF-IDF	SVM	0,9976	1,0000	0,9953	0,9976
TF-IDF	RF	0,9992	1,0000	0,9984	0,9992

Hasil testing model menunjukkan bahwa seluruh kombinasi fitur ekstraktor dan algoritma klasifikasi menghasilkan performa yang sangat tinggi. Kombinasi Count Vectorizer dengan SVM, Count Vectorizer dengan RF, dan TF-IDF dengan RF memperoleh nilai akurasi sebesar 0,9992, precision sebesar 1,0000, recall sebesar 0,9984, dan F1-score sebesar 0,9992. Hasil ini menunjukkan bahwa ketiga kombinasi tersebut mampu mengklasifikasikan data dengan tingkat kesalahan yang sangat rendah. Sementara itu, kombinasi TF-IDF dengan SVM menghasilkan akurasi sebesar 0,9976, precision sebesar 1,0000, recall sebesar 0,9953, dan F1-score sebesar 0,9976. Meskipun nilainya sedikit lebih rendah dibandingkan kombinasi lainnya, performa tersebut tetap menunjukkan kemampuan klasifikasi yang sangat baik. Secara keseluruhan, Random Forest menunjukkan performa yang stabil pada kedua metode ekstraksi fitur, sedangkan SVM memperoleh hasil terbaik ketika menggunakan Count Vectorizer, yang diperoleh dari nilai confussion matrix pada Gambar 5.



Gambar 5. Confusion Matrix

Confusion matrix menunjukkan bahwa kombinasi Count Vectorizer dengan SVM menghasilkan kinerja klasifikasi yang sangat baik. Model berhasil mengklasifikasikan seluruh 632 data NORMAL dengan benar sebagai NORMAL. Pada kelas JUDOL, model berhasil mengklasifikasikan 637 data dengan benar sebagai JUDOL, sedangkan 1 data JUDOL salah diklasifikasikan sebagai NORMAL. Nilai F1-score sebesar 0,9992 menunjukkan bahwa kombinasi ini memiliki tingkat ketepatan dan keseimbangan klasifikasi yang sangat tinggi.

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1474
This Journal is licensed under a Creative Commons Attribution 4.0 International License

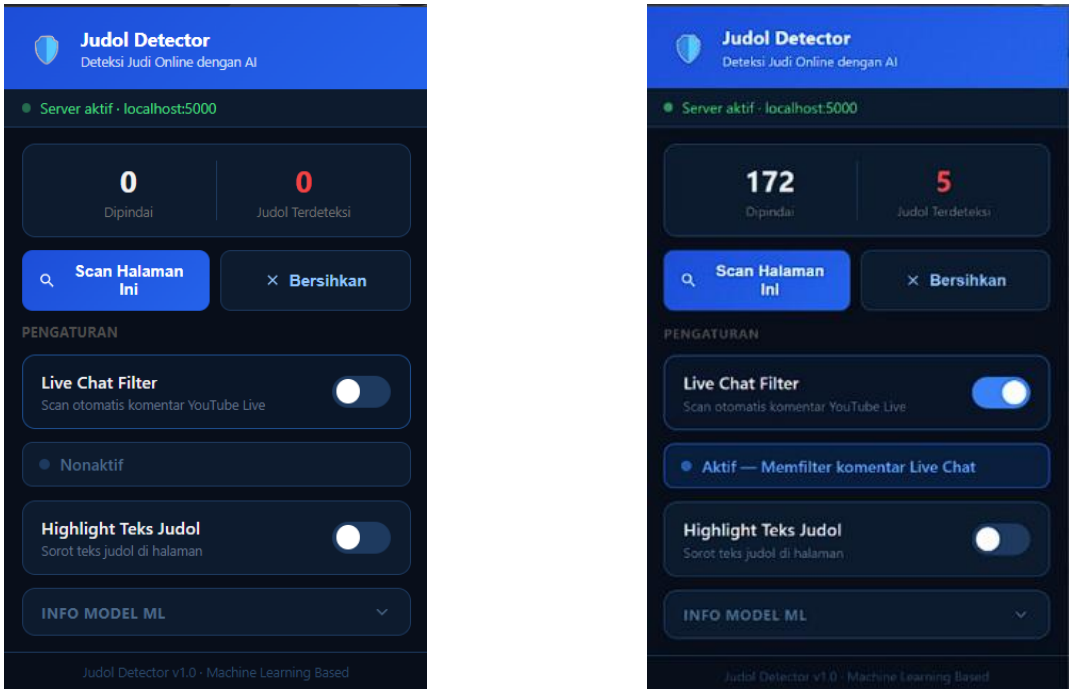
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Pada kombinasi TF-IDF dengan SVM, model juga mampu mengklasifikasikan seluruh 632 data NORMAL dengan benar sebagai NORMAL. Namun, pada kelas JUDOL, terdapat 3 data yang salah diklasifikasikan sebagai NORMAL, sedangkan 635 data berhasil diklasifikasikan dengan benar sebagai JUDOL. Nilai F1-score sebesar 0,9976 menunjukkan bahwa performa model tetap sangat baik, meskipun sedikit lebih rendah dibandingkan kombinasi Count Vectorizer dengan SVM. Secara keseluruhan, hasil ini menunjukkan bahwa SVM memberikan performa lebih optimal ketika menggunakan Count Vectorizer dibandingkan TF-IDF pada data yang digunakan.

Untuk perbandingan yang adil, evaluasi confusion matrix juga dilakukan secara mendalam pada algoritma Random Forest. Pada kombinasi Count Vectorizer dengan Random Forest serta TF-IDF dengan Random Forest, model berhasil mengklasifikasikan seluruh 632 data NORMAL dengan benar tanpa ada False Positive (FP = 0). Pada kelas JUDOL, sebanyak 637 data berhasil diklasifikasikan dengan tepat sebagai JUDOL (True Positive = 637) dan hanya 1 data JUDOL yang terlewat atau salah diklasifikasikan sebagai NORMAL (False Negative = 1). Rincian metrik per kelas untuk Random Forest menunjukkan Precision 1,0000 dan Recall 0,9984 pada kelas JUDOL, serta Precision 0,9984 dan Recall 1,0000 pada kelas NORMAL dengan F1-score makro sebesar 0,9992. Capaian ini membuktikan bahwa arsitektur ensemble pohon pada Random Forest memiliki daya tahan dan akurasi klasifikasi yang setara dengan pemisah margin optimal SVM dalam menangani leksikon spam live chat.

3.5 Implementasi Extension

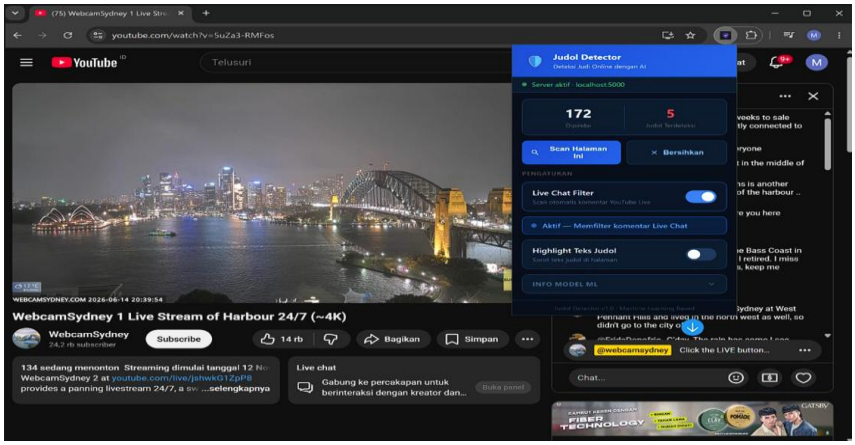
Setelah proses pemodelan selesai, model terbaik dipilih dan disimpan berdasarkan kombinasi fitur ekstraktor dan algoritma klasifikasi yang memberikan performa tertinggi. Model tersebut disimpan dalam format .pkl. Pendekatan ini memungkinkan model untuk langsung digunakan dalam aplikasi eksternal, seperti ekstensi Chrome, tanpa perlu melakukan pelatihan ulang. Dengan menyertakan vectorizer atau model embedding yang sesuai, ekstensi dapat melakukan preprocessing teks dan prediksi secara real-time dengan hasil klasifikasi yang konsisten. Metode penyimpanan ini memastikan replikasi performa model, mempermudah integrasi ke dalam aplikasi web, serta mendukung pengembangan fitur tambahan seperti deteksi otomatis konten NORMAL dan JUDOL pada data teks pengguna.



Gambar 6. Sebelum dan Sesudah Scan

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1475
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9944
Hal 1465-1477
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 7. Case Penggunaan Extension

Gambar 6 mengilustrasikan antarmuka ekstensi browser "Judol Detector v1.0" sebagai media deployment model machine learning. Karena berkas model (.pkl) berbasis Python tidak dapat dieksekusi langsung oleh Chrome Extension yang berjalan dengan JavaScript, ekstensi ini menggunakan arsitektur client-server. Model terbaik (Count Vectorizer dengan Random Forest/SVM) dimuat pada backend REST API lokal yang berjalan di localhost. Ekstensi berperan sebagai klien yang mengambil teks pesan dari DOM live chat YouTube, mengirimkannya ke endpoint API melalui permintaan HTTP, dan menerima label prediksi ("JUDOL" atau "NORMAL") sebagai respons.

Pendekatan operasional utama ekstensi bertumpu pada fitur Auto-scan halaman. Fitur ini dirancang untuk mengatasi tingginya volume dan kecepatan aliran pesan dengan memantau dan memindai setiap pesan baru secara otomatis dan real-time, tanpa intervensi manual pengguna. Setiap pesan yang masuk dikirim ke backend untuk diklasifikasikan, dan apabila terdeteksi sebagai promosi judi online, ekstensi menyorot teks pesan tersebut menggunakan fitur Highlight teks judol agar mudah diidentifikasi.

Pengujian kuantitatif dilakukan pada lingkungan komputasi dengan prosesor Intel Xeon 2,80 GHz (1 vCPU), RAM 3 GB, sistem operasi Ubuntu 24.04, Python 3.12.3, dan scikit-learn 1.8.0, menggunakan 1.270 pesan uji dengan 10 pengulangan setelah 50 prediksi pemanasan (total 12.700 prediksi). Pada lapisan inferensi murni (preprocessing, transformasi vectorizer, dan prediksi), latensi rata-rata adalah 0,46 ± 0,13 ms per pesan (median 0,43 ms, p95 0,66 ms, p99 0,89 ms), dengan throughput 2.164 pesan per detik. Pada lapisan end-to-end yang mencakup komunikasi HTTP ke API lokal (Flask) menggunakan klien Python, pengujian terhadap 3.810 permintaan menghasilkan latensi rata-rata 3,19 ± 0,75 ms (p95 4,62 ms, p99 5,35 ms) dengan throughput sekitar 300–320 pesan per detik dan tingkat keberhasilan deteksi 99,92%. Sistem kemudian diuji dengan laju kedatangan 50, 100, dan 200 pesan per detik selama 30 detik. Latensi rata-rata masing-masing adalah 4,8 ms, 4,5 ms, dan 4,3 ms (p99 9,3 ms, 8,6 ms, dan 14,0 ms), dan seluruh pesan terlayani sesuai laju kedatangan tanpa antrean yang menumpuk. Pengujian ini mensimulasikan alur komunikasi ekstensi dengan API lokal sehingga belum mencakup waktu pemrosesan DOM pada sisi browser.

4. KESIMPULAN

Penelitian ini berhasil membangun sistem klasifikasi chat judi online pada live chat YouTube dengan 6.348 data bersih (3.187 JUDOL dan 3.161 NORMAL) menggunakan Count Vectorizer dan TF-IDF yang dikombinasikan dengan SVM dan Random Forest. Count Vectorizer dengan SVM dipilih sebagai model terbaik berdasarkan rata-rata F1-score Stratified 10-Fold Cross Validation tertinggi (0,9990 ± 0,0010), dan memperoleh accuracy 0,9992, precision 1,0000, recall 0,9984, serta F1-score 0,9992 pada data uji independen dengan satu pesan JUDOL terlewat. Perbedaan antar-kombinasi berada dalam kisaran simpangan baku sehingga tidak dapat dianggap signifikan. Model dideploy pada Chrome Extension yang terhubung ke backend REST API lokal dan, pada lingkungan 1 vCPU, mencapai latensi inferensi rata-rata 0,46 ms per pesan (2.164 pesan per detik), latensi end-to-end rata-rata 3,19 ms, serta tingkat keberhasilan deteksi 99,92%. Keterbatasan penelitian adalah dataset berasal dari satu sumber siaran dengan hanya 36 template pesan JUDOL unik dan 59,1% data uji identik dengan data latih, sehingga performa mendekati sempurna belum membuktikan generalisasi terhadap pola spam baru (accuracy 0,9926 pada pemisahan berbasis grup teks). Selain itu, pengujian kinerja belum mencakup pemrosesan DOM di browser, sistem bergantung pada backend lokal, dan klasifikasi masih biner untuk bahasa Indonesia. Penelitian selanjutnya dapat menggunakan dataset multi-sumber dan multi-bahasa serta model deep learning seperti LSTM atau BERT..

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1476
This Journal is licensed under a Creative Commons Attribution 4.0 International License

REFERENCES

[1] M. N. Sari and J. Veri, "PERAN MEDIA SOSIAL DALAM KOMUNIKASI PUBLIK DI ERA DIGITAL: KAJIAN LITERATUR SISTEMATIS BERBASIS PRISMA," *Jurnal Daya Saing*, vol. 11, no. 3, pp. 669–677, Oct. 2025.

[2] H. Lin, "`Let's purchase coloured live chat messages': the impact of user engagement withs Super Chat on YouTube," *Inf. Commun. Soc.*, vol. 28, no. 9, pp. 1608–1626, Jul. 2025.

[3] A. Fahrudin *et al.*, "Online gambling addiction: Problems and solutions for policymakers and stakeholders in Indonesia," *J. Infrastruct. Pol. Dev.*, vol. 8, no. 11, p. 9077, Oct. 2024.

[4] P. A. N. Martha, "Legal review of online gambling crimes in Indonesia and its impact on the younger generation," *Journal of Social Research*, vol. 4, no. 8, pp. 2213–2218, Aug. 2025.

[5] E. Yang, D. D. Lewis, and O. Frieder, "TAR on social media: A framework for online content moderation," Aug. 2021.

[6] J. S. Andersen, O. Zukunft, and W. Maalej, "REM: Efficient semi-automated real-time moderation of online forums," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021.

[7] A. C. Zahro *et al.*, "Fairer Public Complaint Classification on LaporGub: Integrating XLM-RoBERTa with Focal Loss for Imbalance Data," *sinkron*, vol. 9, no. 4, pp. 1850–1862, Oct. 2025, doi: 10.33395/sinkron.v9i4.15260.

[8] K. Szabó Nagy, J. Kapusta, and M. Munk, "Feature extraction from unstructured texts as a combination of the morphological and the syntactic analysis and its usage in fake news classification tasks," *Neural Comput. Appl.*, Sep. 2023.

[9] K. Dumbre, B. Shiragapur, Y. More, and V. Kale, "Enhanced frequency-based word embedding approach for text analysis: Comparative study with existing method," in *2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA)*, IEEE, Oct. 2023.

[10] S. Senbel, "Fast and memory-efficient TFIDF calculation for text analysis of large datasets," in *Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices*, in Lecture Notes in Computer Science. , Cham: Springer International Publishing, 2021, pp. 557–563.

[11] D. Kaibassova, B. Mukhametzhanova, D. Tokseit, A. Kubegenova, and M. Kozhanov, "Document analysis via combined vectorization and machine learning approaches," *ijirss*, vol. 8, no. 4, pp. 2195–2204, Jul. 2025.

[12] Z. Labd, S. Bahassine, K. Housni, F. Z. A. Hamou Aadi, and K. Benabbes, "Text classification supervised algorithms with term frequency inverse document frequency and global vectors for word representation: a comparative study," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 14, no. 1, p. 589, Feb. 2024.

[13] S. Bose and S. Bose, "Random Forests: The wisdom of crowds in action," *Journal of Emerging Trends in Computer Science and Applications*, vol. 1, no. 1, pp. 67–91, Nov. 2025.

[14] M. Maindola *et al.*, "Utilizing random forests for high-accuracy classification in medical diagnostics," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, Sep. 2024, pp. 1679–1685.

[15] Yaemico, "Dataset Judol Youtube Live Chat Ultah Jogja Ke-268," Kaggle. Accessed: Jun. 01, 2026. [Online]. Available: https://www.kaggle.com/datasets/yaemico/deteksi-judi-online?select=youtube_chat_jogja_clean.csv

[16] M. Ariyanto, F. Alzami, R. Rakhmat Sani, I. Gamayanto, M. Naufal, and S. Winarno, "Addressing Extreme Class Imbalance in Multilingual Complaint Classification Using XLM-RoBERTa," 2026. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[17] P. Cichosz, "Bag of words and embedding text representation methods for medical article classification," *Int. J. Appl. Math. Comput. Sci.*, vol. 33, no. 4, 2023.

[18] H. S. Lubis, M. K. M. Nasution, and A. Amalia, "Performance of term frequency - inverse document frequency and K-means in government service identification," in *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, IEEE, Jul. 2024, pp. 772–777.

[19] H. Bichri, A. Chergui, and M. Hain, "Investigating the impact of train / test split ratio on the performance of pre-trained models with custom datasets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, 2024.

[20] I. K. Nti, Department of Computer Science, U. of E. Informatics, G. Natural Resources Sunyani, O. Nyarko-Boateng, and J. Aning, "Performance of machine learning algorithms with different K values in K-fold CrossValidation," *Int. J. Inf. Technol. Comput. Sci.*, vol. 13, no. 6, pp. 61–71, Dec. 2021.

[21] R. Vinston Raja, J. Jayashankari, S. Sheela, S. Jancy Sickory Daisy, G. G. Gokilam, and M. Robinson Joel, "Metrics and techniques for evaluating machine learning models and optimization algorithms," in *Advances in Computational Intelligence and Robotics*, IGI Global Scientific Publishing, 2025, pp. 193–222.

[22] Rafi Jonathan Siger, Muhammad Naufal, and Farrikh Alzami, "Analisis Performa Class Weight Dan Focal Loss Pada Model Indobert Untuk Klasifikasi Teks Depresi Berbahasa Indonesia," *JURIKOM (Jurnal Riset Komputer)*, vol. 13, no. 2, pp. 558–568, Apr. 2026, doi: 10.30865/jurikom.v13i2.9620.

Copyright © 2026 The **Muhammad Taufiq Ramadhan**, Page 1477
This Journal is licensed under a Creative Commons Attribution 4.0 International License