JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274

DETEKSI KOMENTAR SPAM PERJUDIAN ONLINE DI YOUTUBE MENGGUNAKAN K-NEAREST NEIGHBORS DAN NAÏVE BAYES

Idil Ode[1)], Christian Dwi Suhendra*[2)], Marlinda Sanglise[3)]
1. Teknik Informatika, Fakultas Teknik, Universitas Papua, Indonesia
2. Teknik Informatika, Fakultas Teknik, Universitas Papua, Indonesia
3. Teknik Informatika, Fakultas Teknik, Universitas Papua, Indonesia

Article Info
Kata Kunci: Judi Online; K-Nearest Neighbors; Komentar Spam; Machine Learning; Naïve Bayes

Keywords: K-Nearest Neighbors, Machine Learning, Naïve Bayes, Online Gambling, Spam Comments

Article history:
Received 18 April 2025
Revised 23 May 2025
Accepted 2 June 2025
Available online 1 March 2026

DOI :
https://doi.org/10.29100/jipi.v11i1.7682

* Corresponding author.
Christian Dwi Suhendra
E-mail address:
c.suhendra@unipa.ac.id

ABSTRAK
Penyebaran komentar spam perjudian daring di YouTube semakin meningkat dan mengganggu pengalaman pengguna. Komentar-komentar ini sering kali menggunakan pola bahasa tertentu, seperti simbol dan variasi kata, untuk menghindari deteksi otomatis. Meskipun YouTube telah menerapkan sistem filter spam, banyak komentar masih lolos deteksi, yang menunjukkan perlunya metode yang lebih akurat dan adaptif. Penelitian ini bertujuan untuk menganalisis pola bahasa dalam komentar spam perjudian daring dan mengembangkan model deteksi berbasis pembelajaran mesin menggunakan K-Nearest Neighbors (KNN) dan Naïve Bayes (NB). Dataset dikumpulkan menggunakan YouTube Data API v3, diproses melalui tahapan normalisasi, cleaning, analisis statistik, visualisasi, dan pelabelan sebelum diterapkan pada pemodelan pembelajaran mesin. Fitur teks diekstraksi menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Bag of Words (BoW), dengan pengoptimalan hiperparameter dilakukan menggunakan GridSearchCV untuk meningkatkan kinerja model. Evaluasi model didasarkan pada metrik accuracy dan F1-Score. Hasilnya menunjukkan bahwa model Naïve Bayes dengan TF-IDF mencapai kinerja tertinggi, dengan akurasi 99% setelah normalisasi dan 98% setelah cleaning, mengungguli Naïve Bayes dengan BoW dan KNN. Temuan ini berkontribusi pada pengembangan sistem deteksi spam yang lebih efektif, khususnya dalam mengidentifikasi pola bahasa tertentu yang digunakan dalam komentar spam perjudian di YouTube. Dengan pendekatan ini, diharapkan dapat dikembangkan sistem moderasi otomatis yang lebih efisien untuk menyaring komentar spam dan meningkatkan kualitas interaksi pengguna di platform YouTube.

ABSTRACT
The proliferation of online gambling spam comments on YouTube is increasing and disrupting the user experience. These comments often use certain language patterns, such as symbols and word variations, to avoid automatic detection. Although YouTube has implemented a spam filter system, many comments still escape detection, indicating the need for more accurate and adaptive methods. This study aims to analyze the language patterns in online gambling spam comments and develop a machine learning-based detection model using K-Nearest Neighbors (KNN) and Naïve Bayes (NB). The dataset was collected using YouTube Data API v3, processed through normalization, cleaning, statistical analysis, visualization, and labeling stages before being applied to machine learning modeling. Text features were extracted using Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW), with hyperparameter optimization performed using GridSearchCV to improve model performance. Model evaluation was based on accuracy and F1 score metrics. The results show that the Naïve Bayes model with TF-IDF achieves the highest performance, with an accuracy of 99% after normalization and 98% after cleaning, outperforming Naïve Bayes with BoW and KNN. These findings contribute to the development of a more effective spam detection system, especially in identifying specific language patterns used in gambling spam comments on YouTube. With this approach, it is expected that a more efficient automatic moderation system can be developed to filter spam comments and improve the quality of user interactions on the YouTube platform.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274

I. PENDAHULUAN

YOUTUBE sebagai salah satu platform media sosial terbesar menghadapi lonjakan komentar spam, terutama yang berhubungan dengan promosi perjudian online. Komentar spam dapat berupa teks biasa, kombinasi teks dan emoji, atau bahkan hanya terdiri dari simbol emoji. Penggunaan teknik penyamaran semacam ini tidak hanya mengganggu komunikasi, tetapi juga meningkatkan ambiguitas dalam diskusi online [1]. Keberadaan komentar semacam ini tidak hanya mengganggu pengalaman pengguna, tetapi juga berisiko menargetkan kelompok rentan, seperti anak-anak dan remaja, yang dapat terjerumus dalam praktik perjudian ilegal [2].

Lebih jauh, komentar spam judi online yang tidak terdeteksi dapat menimbulkan dampak sosial yang signifikan. Menurut Organisasi Kesehatan Dunia (WHO), sekitar 1,2% populasi orang dewasa global mengalami gangguan perjudian, dan bahaya dari perjudian yang dilakukan oleh orang lain juga tersebar luas [3]. Perjudian telah terbukti mengancam kesehatan masyarakat dengan meningkatkan risiko gangguan mental dan bunuh diri, serta mendorong kemiskinan melalui pengalihan pengeluaran rumah tangga dari kebutuhan pokok. Selain itu, dampaknya meluas ke aspek sosial seperti kerusakan hubungan, kekerasan dalam keluarga, pengabaian anak, serta kesulitan finansial. Praktik perjudian juga berpotensi terkait dengan kejahatan yang menghasilkan pendapatan seperti pencurian dan penipuan serta menjadi sarana pencucian uang dan sumber korupsi dalam institusi sipil maupun aktivitas politik perusahaan [3].

Meskipun YouTube telah menyediakan fitur laporkan untuk menangani komentar spam [4], fitur ini masih belum cukup efektif dalam mendeteksi komentar spam judi online secara menyeluruh. Sebagai contoh, Google memblokir lebih dari 100 ribu situs judi berisi spam setiap minggu, tetapi upaya ini masih berkelanjutan dan membutuhkan kerja sama berbagai pihak untuk menciptakan lingkungan online yang aman [5]. Oleh karena itu, diperlukan pendekatan baru yang lebih akurat dan efisien dalam mendeteksi komentar spam judi online.

Pendekatan berbasis machine learning telah digunakan untuk mendeteksi berbagai jenis spam, termasuk email [6], SMS [7], dan komentar spam di media sosial lainnya [8], [9]. Namun, pendekatan ini seringkali tidak memadai untuk menangani komentar spam judi online yang memiliki karakteristik unik. Pada spam email misalnya, pesan-pesan yang dikirimkan biasanya berupa salinan pesan yang sama yang disebarkan secara massal untuk memaksa penerima melihat konten yang tidak diinginkan, sehingga memakan waktu dan bandwidth yang besar bagi pengguna [6]. Sementara itu, komentar spam judi online cenderung jauh lebih tidak terstruktur dan kompleks. Penelitian sebelumnya menyebutkan bahwa konten seperti ini menggunakan berbagai teknik penyamaran, seperti simbol, emoji, kesalahan ketik yang disengaja, singkatan, dan pencampuran bahasa, untuk menghindari deteksi otomatis [10]. Filtering otomatis komentar spam di YouTube juga menjadi tantangan besar bahkan bagi metode klasifikasi yang telah mapan, karena pesan-pesan tersebut sangat pendek dan sering kali dipenuhi dengan bahasa gaul, simbol, dan singkatan [11]. Komentar spam bisa terdiri dari teks biasa dalam format ASCII/Unicode Latin, atau kombinasi teks dan emoji yang disusun secara berurutan [1]. Meskipun YouTube telah menyediakan alat bawaan untuk mengendalikan spam, alat ini masih belum cukup efektif untuk melawan konten komentar berbahaya dan spam [11].

Selain tantangan dalam struktur bahasa, komentar di YouTube memiliki karakteristik lain yang membedakannya dari platform seperti Twitter dan Facebook [12]. Salah satu keunikan utamanya adalah sifat multibahasa komentar, yang mencakup lebih dari 50 bahasa berbeda yang digunakan secara aktif di platform ini [13]. Keanekaragaman bahasa dan ekspresi budaya tersebut [13] mengharuskan penggunaan strategi deteksi yang mampu menangkap variasi pola komunikasi, termasuk dalam mengidentifikasi komentar spam. Selain itu, komentar di YouTube biasanya memiliki keterkaitan langsung dengan konten video, sehingga konteks komentar menjadi sangat beragam dan sulit diprediksi [14]. Faktor-faktor ini menyebabkan deteksi komentar spam di YouTube memerlukan pendekatan yang lebih adaptif dan kontekstual dibandingkan media sosial lainnya.

Bahkan ditemukan bahwa kata-kata tertentu seperti “link”, “situs”, “prediksi”, “jackpot”, “maxwin”, dan “togel” sering digunakan sebagai fitur tekstual utama dalam promosi judi online di media sosial, khususnya Twitter [15]. Keragaman pola bahasa dan upaya penyamaran yang dilakukan oleh pelaku membuat deteksi komentar semacam ini jauh lebih menantang dibandingkan spam konvensional, karena memerlukan pendekatan yang lebih adaptif terhadap konteks dan bentuk ekspresi yang tidak standar.

Penelitian oleh [16] berfokus pada analisis sentimen publik terhadap perjudian menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM), dengan hasil akurasi tertinggi sebesar 98% untuk SVM dan 93% untuk

Naïve Bayes. Namun, penelitian ini tidak secara spesifik membahas deteksi komentar spam judi online. Penelitian lain oleh [17] lebih menitikberatkan pada pemetaan opini masyarakat terkait fenomena perjudian online berdasarkan data media sosial, tanpa berfokus pada deteksi spam atau teknik penyamaran dalam komentar. Sementara itu, penelitian [18] menerapkan teknik Natural Language Processing (NLP) untuk memahami sentimen masyarakat terhadap perjudian online di Twitter, tetapi tidak secara khusus menangani deteksi komentar spam yang menggunakan simbol atau emoji sebagai bentuk penyamaran.

Penelitian ini menerapkan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan Bag of Words (BoW) untuk ekstraksi fitur, karena keduanya merupakan teknik yang umum digunakan dalam merepresentasikan data teks ke dalam bentuk numerik, dengan kelebihan dan kekurangan masing-masing dalam proses analisis teks [19].

Meskipun embedding modern seperti BERT, Word2Vec, dan Doc2Vec memiliki kemampuan dalam merepresentasikan fungsi semantik dan kontekstual dari bahasa, beberapa studi menunjukkan bahwa metode statistik seperti TF-IDF dan BoW tetap menjadi pilihan yang efektif untuk deteksi spam. Studi oleh [20] menunjukkan bahwa metode statistik seperti TF-IDF, BoW, dan N-grams memberikan performa terbaik ketika dipasangkan dengan klasifikasi seperti Multinomial Naïve Bayes (MNB), Random Forest (RF), dan Extra Trees (ET). Word2Vec tidak menunjukkan hasil yang kompetitif dalam dataset terbatas, sementara Doc2Vec terlihat menjanjikan namun membutuhkan data yang lebih besar untuk meningkatkan performa. Selain itu, [21] menemukan bahwa TF-IDF secara konsisten mengungguli BoW dalam mendeteksi spam SMS, terutama ketika digunakan bersama dengan Naïve Bayes dan SVM, menunjukkan skor presisi dan recall yang tinggi. Di sisi lain, studi [22] menunjukkan bahwa BERT efektif dalam menghadapi serangan seperti Mad-lib karena kemampuannya menangkap makna kontekstual dan mengenali istilah di luar kosakata standar. Namun, secara komputasi, BERT jauh lebih berat dibanding BoW, dan kinerjanya hanya unggul secara signifikan pada spam yang telah dimanipulasi secara adversarial. Bahkan, untuk kasus spam yang tidak dimanipulasi, BoW mampu memberikan hasil yang sebanding. Studi tersebut juga mengusulkan kombinasi model encoding sebagai pendekatan untuk mendeteksi perubahan perilaku spam, namun tidak menyarankan BERT sebagai solusi tunggal. Oleh karena itu, pemilihan TF-IDF dan BoW dalam penelitian ini didasarkan pada bukti empiris dari ketiga studi tersebut yang menunjukkan efektivitas dan efisiensi metode statistik, terutama dalam konteks teks pendek dan tidak terstruktur seperti komentar spam di YouTube.

Selain itu, penelitian ini menggunakan dua algoritma klasifikasi, yaitu K-Nearest Neighbors (KNN) dan Naïve Bayes (NB), untuk membandingkan kinerja keduanya [23] dalam mendeteksi komentar spam judi online. KNN dipilih karena kemampuannya dalam mengelompokkan teks berdasarkan kesamaan fitur dengan data [24] yang telah diberi label, sedangkan Naïve Bayes dipilih karena pendekatannya yang berbasis probabilitas [25], sehingga lebih efisien dalam klasifikasi teks.

Dataset yang digunakan diperoleh melalui proses scraping komentar YouTube, kemudian diproses melalui beberapa tahapan utama, yaitu normalisasi data, cleaning data, analisis statistik dan visualisasi data, pelabelan data, serta pemodelan machine learning dan evaluasi model. Setiap tahapan ini dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik sebelum diterapkan dalam model deteksi spam.

Tujuan penelitian ini adalah untuk mengembangkan model deteksi komentar spam judi online yang lebih akurat dan efisien, serta mengevaluasi perbandingan kinerja KNN dan Naïve Bayes dalam mengklasifikasikan komentar spam menggunakan metode TF-IDF dan BoW, yang telah terbukti efektif dalam deteksi teks di berbagai aplikasi, seperti deteksi ujaran kebencian pada tweet [26].

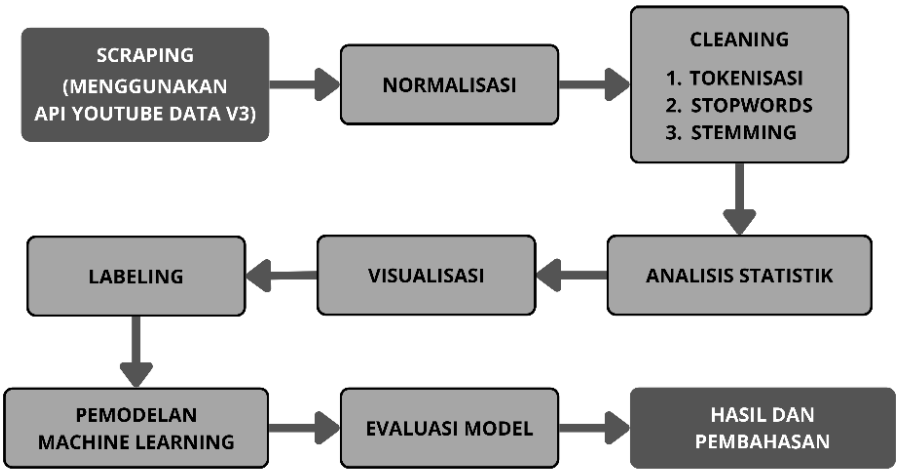
Kebaruan dari penelitian ini terletak pada analisis pola bahasa unik dalam komentar spam judi online serta pengujian efektivitas machine learning dalam mendeteksi komentar-komentar tersebut. Penelitian ini belum banyak dibahas dalam literatur yang ada, terutama yang berfokus pada platform seperti YouTube.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem deteksi otomatis yang lebih efektif untuk mendeteksi komentar spam judi online di YouTube, serta membantu menciptakan lingkungan online yang lebih aman bagi pengguna.

II. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan utama yang dapat dilihat pada Gambar 1.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274



Gambar. 1. Diagram Alur Metode Penelitian

A. Scraping Data

Proses scraping dilakukan menggunakan pemrograman Python [27] melalui YouTube API v3 untuk mengumpulkan komentar dari video dalam kategori gaming, edukasi, dan hiburan dengan total 25.093 komentar.

B. Normalisasi Data

Proses normalisasi bertujuan untuk menyelaraskan format komentar agar dapat dianalisis dengan lebih baik. Langkah-langkah yang dilakukan meliputi konversi karakter Unicode atau simbol spesial ke bentuk standar, penghapusan tanda baca yang tidak diperlukan, serta penggabungan huruf yang terpisah akibat spasi berlebih. Selain itu, komentar yang hanya berisi simbol, angka tanpa konteks, timestamp, tautan, dan komentar yang kosong atau tidak relevan untuk memastikan data yang bersih dan terstruktur.

C. Cleaning Data

Proses cleaning bertujuan untuk menghilangkan komentar yang tidak relevan atau mengandung noise agar analisis dapat dilakukan secara lebih akurat.

1) Case Folding dan Tokenisasi: Mengubah teks menjadi huruf kecil dan memecahnya menjadi kata-kata.
2) Penghapusan Stopwords: Menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis menggunakan daftar stopwords dari NLTK serta daftar tambahan kata-kata umum dalam komentar.
3) Stemming: Mengubah kata menjadi bentuk dasar menggunakan Sastrawi.
4) Pembersihan Data Tidak Relevan: Menghapus komentar yang kosong setelah proses sebelumnya.

D. Analisis Statistik dan Visualisasi Data

Analisis statistik dilakukan untuk memahami karakteristik komentar setelah melalui proses normalisasi dan cleaning. Langkah pertama adalah menghitung panjang setiap komentar berdasarkan jumlah karakter dan kata, termasuk menghitung rata-rata, serta mengidentifikasi komentar dengan panjang terpendek dan terpanjang.

Setelah itu, dilakukan visualisasi data untuk melihat pola kata yang sering muncul. Grafik batang digunakan untuk menampilkan 20 kata yang paling sering muncul, sementara word cloud dibuat untuk memberikan gambaran visual mengenai distribusi kata dalam komentar. Teknik ini membantu memahami tren kata yang dominan dalam dataset.

E. Labeling Data

Setelah mendapatkan pola kata yang sering muncul, langkah berikutnya adalah labeling data untuk mengklasifikasikan komentar menjadi spam atau non-spam. Proses ini dilakukan secara semi-otomatis dengan mendeteksi kata kunci yang sering muncul dalam komentar spam judi online.

Labeling dimulai dengan identifikasi kata kunci seperti “slot”, “gacor”, “jackpot”, dan angka khas promosi judi seperti “777”, “4d”, dan lainnya. Komentar yang mengandung kata atau pola ini langsung diberi label spam. Untuk menghindari kesalahan (false positives), beberapa frasa seperti “slot waktu”, “bonus pelajaran”, atau istilah dalam game seperti “hero”, “buff”, dan “lvl” dikecualikan dari deteksi spam. Selain itu, angka ID panjang juga tidak dianggap sebagai spam.

Setelah labeling otomatis selesai, hasilnya diperiksa secara manual untuk memastikan akurasi. Koreksi dilakukan jika ada komentar yang salah terdeteksi. Dengan metode ini, proses labeling menjadi lebih cepat tanpa

mengorbankan keakuratan.

F. *Pemodelan Machine Learning dan Evaluasi Model*

Dataset dibagi menjadi 80% data latih untuk melatih model dalam mengenali pola data dan 20% data uji untuk mengevaluasi performa model dalam mengklasifikasikan data baru [28]. Sebelum dilakukan klasifikasi, data diekstraksi fiturnya menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan Bag of Words (BoW) agar teks dapat direpresentasikan dalam bentuk numerik. Setelah itu, dua algoritma machine learning diterapkan, yaitu K-Nearest Neighbors (KNN) dan Naïve Bayes. Pemilihan KNN didasarkan pada fleksibilitasnya dalam menangani berbagai distribusi data, sementara Naïve Bayes dipilih karena efisiensinya dalam klasifikasi teks berbasis probabilitas.

Untuk mengoptimalkan performa model, dilakukan tuning hyperparameter dengan GridSearchCV [29] guna mencari kombinasi parameter terbaik yang dapat meningkatkan kinerja model. Pada model KNN, jumlah tetangga diuji dengan nilai 3, 5, 7, dan 9. Sedangkan pada Naïve Bayes, nilai alpha diuji dengan 0,1, 0,5, dan 1,0. Model dievaluasi menggunakan metrik utama, yaitu accuracy, precision, recall, dan F1-score, untuk mengukur kinerja klasifikasi berdasarkan perbandingan hasil prediksi dengan data actual [30]. Dari keempat metrik tersebut, F1-score dipilih sebagai metrik utama karena mampu menangkap keseimbangan antara precision dan recall dalam mendeteksi komentar spam. Selain itu, untuk memahami distribusi prediksi model terhadap data aktual, digunakan Confusion Matrix yang menunjukkan kategori prediksi, yaitu True Positive, True Negative, False Positive, dan False Negative [31].

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari setiap tahapan penelitian yang telah dilakukan, mulai dari pengumpulan data hingga penerapan metode yang digunakan. Setiap tahapan dijelaskan berdasarkan temuan yang diperoleh selama proses analisis.

A. *Scraping Data*

Scraping dilakukan menggunakan YouTube API v3, menghasilkan 25.093 komentar dari tiga playlist video dalam kategori gaming, edukasi, dan hiburan. Komentar yang terkumpul terdiri dari 8.187 komentar dari playlist pertama, 8.676 komentar dari playlist kedua, dan 8.230 komentar dari playlist ketiga.

B. *Normalisasi Data*

Proses normalisasi menghasilkan data komentar yang lebih terstruktur. Dari total 25.093 komentar, sebanyak 992 komentar dihapus karena hanya berisi simbol, angka tanpa konteks, timestamp, dan tautan. Setelah normalisasi, tersisa 24.101 komentar.

Tabel 1 menunjukkan beberapa contoh komentar sebelum dan sesudah normalisasi.

TABEL I
NORMALISASI DATA

Sebelum Normalisasi	Sesudah Normalisasi
Maxwin tanpa ribet melalui **DELICUAN** 🎲 💎	Maxwin tanpa ribet melalui DELICUAN
SPECIAL MEMBER BARU BOSKU.. DIJAMIN GACOR SODA69 BOSKU.. SEMOGA BERMANFAAT BOSKU.. SEMOGA WORK IT DI AKUN BOSKU SEMUA : LINK-https://t.ly/S69-FB	SPECIAL MEMBER BARU BOSKU DIJAMIN GACOR SODA69 BOSKU SEMOGA BERMANFAAT BOSKU SEMOGA WORK IT DI AKUN BOSKU SEMUA LINK
🔴 🔴 sudah terbukti **BIR123** 💜 💜	sudah terbukti BIR123
Mirip kok, TAPI SAYA TAHU ITU COMPUTER :) 100k🖤	Mirip kok TAPI SAYA TAHU ITU COMPUTER 100k
7:54 KAYA SKYLAR GEBUG ARAN 😂 😂 😂 😂	KAYA SKYLAR GEBUG ARAN
🅿️ **U L A U**777 selalu punya fitur keren	PULAU777 selalu punya fitur keren

C. *Cleaning Data*

Proses cleaning data dilakukan melalui tahapan tokenisasi, penghapusan stopwords, dan stemming. Setelah proses cleaning, sebanyak 356 komentar dihapus karena dianggap tidak relevan, sehingga jumlah komentar yang tersisa untuk analisis adalah 23.745 komentar. Hasil dari proses ini dapat dilihat pada Tabel 2.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274

TABEL II
HASIL CLEANING DATA

Sebelum Cleaning	Tokenisasi	Stopwords	Stemming
Maxwin tanpa ribet melalui DELICUAN	[‘maxwin’, ‘tanpa’, ‘ribet’, ‘melalui’, ‘delicuan’]	maxwin ribet delicuan	maxwin ribet delicuan
SPECIAL MEMBER BARU BOSKU DIJAMIN GACOR SODA69 BOSKU SEMOGA BERMANFAAT BOSKU SEMOGA WORK IT DI AKUN BOSKU SEMUA LINK	[‘special’, ‘member’, ‘baru’, ‘bosku’, ‘dijamin’, ‘gacor’, ‘soda69’, ‘bosku’, ‘semoga’, ‘bermanfaat’, ‘bosku’, ‘semoga’, ‘work’, ‘it’, ‘di’, ‘akun’, ‘bosku’, ‘semua’, ‘link’]	special member bosku dijamin gacor soda69 bosku semoga bermanfaat bosku semoga work it akun bosku link	special member bos jamin gacor soda69 bos moga manfaat bos moga work it akun bos link
sudah terbukti BIR123	[‘sudah’, ‘terbukti’, ‘bir123’]	terbukti bir123	bukti bir123
Mirip kok TAPI SAYA TAHU ITU COMPUTER 100k	[‘mirip’, ‘kok’, ‘tapi’, ‘saya’, ‘tahu’, ‘itu’, ‘computer’, ‘100k’]	computer 100k	computer 100k
KAYA SKYLAR GEBUG ARAN	[‘kaya’, ‘skylar’, ‘gebug’, ‘aran’]	kaya skylar gebug aran	kaya skylar gebug aran
PULAU777 selalu punya fitur keren	[‘pulau777’, ‘selalu’, ‘punya’, ‘fitur’, ‘keren’]	pulau777 fitur keren	pulau777 fitur keren

D. Analisis Statistik dan Visualisasi Data

Setelah melalui proses normalisasi dan cleaning, dilakukan analisis statistik untuk memahami karakteristik komentar. Salah satu aspek yang dianalisis adalah panjang komentar berdasarkan jumlah karakter dan kata yang ditunjukkan pada Tabel 3.

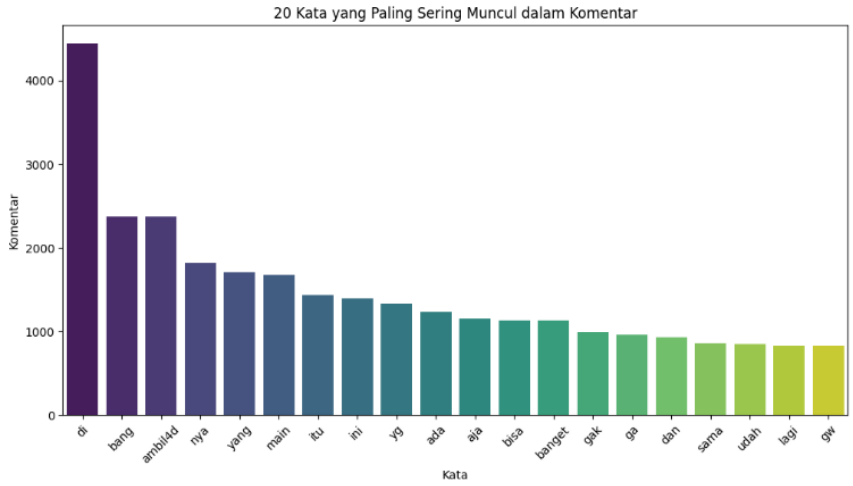
TABEL III
HASIL ANALISIS STATISTIK

Metode	Rata-Rata Panjang (Karakter)	Rata-Rata Panjang (Kata)	Panjang Terpendek (Karakter)	Panjang Terpanjang (Karakter)	Panjang Terpendek (Kata)	Panjang Terpanjang (Kata)
Normalisasi Data	43,12	7,52	2	1.570	1	299
Cleaning Data	27,78	4,72	1	1.570	1	299

Setelah proses cleaning, rata-rata panjang komentar mengalami penurunan, karena kata-kata umum yang tidak memiliki makna signifikan telah dihapus melalui proses stopwords removal dan stemming.

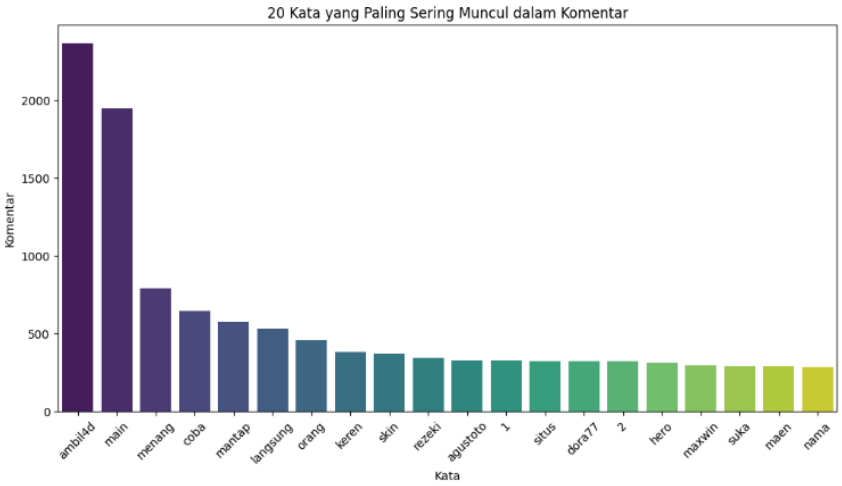
1) Frekuensi Kata yang Paling Sering Muncul

Analisis berikutnya adalah melihat kata-kata yang paling sering muncul dalam komentar. Grafik batang yang ditampilkan pada Gambar 2 menunjukkan 20 kata yang paling sering muncul setelah normalisasi, sedangkan Gambar 3 menampilkan kata-kata yang paling sering muncul setelah proses cleaning.



Gambar. 2. Frekuensi 20 Kata Paling Sering Muncul Setelah Normalisasi

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274



Gambar. 3. Frekuensi 20 Kata paling Sering Muncul Setelah Cleaning

Setelah proses cleaning, beberapa kata umum berhasil dihapus, sehingga kata-kata yang tersisa lebih relevan dalam analisis.

2) Wordcloud untuk Visualisasi Distribusi Kata

Wordcloud digunakan untuk menggambarkan distribusi kata dalam komentar setelah normalisasi dan setelah cleaning. Visualisasi ini dapat dilihat pada Gambar 4 untuk wordcloud setelah normalisasi dan Gambar 5 untuk wordcloud setelah cleaning, yang menunjukkan perubahan pola kata setelah proses pembersihan data.



Gambar. 4. Wordcloud Kata yang Sering Muncul Setelah Normalisasi



Gambar. 5. Wordcloud Kata yang Sering Muncul Setelah Cleaning

Dari visualisasi wordcloud pada Gambar 4, masih banyak kata-kata umum yang kurang relevan. Sedangkan setelah proses cleaning pada Gambar 5, kata-kata yang lebih spesifik dan bermakna mulai lebih dominan,

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274

menunjukkan bahwa proses cleaning berhasil menyaring komentar menjadi lebih informatif.

Selain itu, dari hasil analisis frekuensi kata, terlihat bahwa setelah cleaning, kata-kata terkait promosi seperti "ambil4d", "main", dan "menang" masih sering muncul, mengindikasikan pola komentar yang masih didominasi oleh promosi perjudian online. Dengan adanya analisis ini, dapat disimpulkan bahwa proses cleaning efektif dalam meningkatkan kualitas data.

E. Labeling Data

Proses labeling data dilakukan untuk mengklasifikasikan komentar sebagai spam atau non-spam secara semi-otomatis. Proses ini memanfaatkan kata kunci yang sering muncul pada tahap analisis statistik dan visualisasi data dalam komentar spam judi online serta menerapkan filter untuk menghindari false positives.

1) Labeling Setelah Normalisasi

Pada tahap ini, komentar masih dalam bentuk awal setelah normalisasi, tanpa penghapusan stopwords atau stemming yang dapat dilihat pada Tabel 4.

TABEL IV
HASIL LABELING DATA SETELAH NORMALISASI

Komentar Setelah Normalisasi	Label
Maxwin tanpa ribet melalui DELICUAN	spam
SPECIAL MEMBER BARU BOSKU DIJAMIN GACOR SODA69 BOSKU SEMOGA BERMANFAAT BOSKU SEMOGA WORK IT DI AKUN BOSKU SEMUA LINK	spam
sudah terbukti BIR123	spam
Mirip kok TAPI SAYA TAHU ITU COMPUTER 100k	non-spam
KAYA SKYLAR GEBUG ARAN	non-spam
PULAU777 selalu punya fitur keren	spam

Komentar spam terdeteksi berdasarkan kata kunci seperti "maxwin", "gacor", dan "777".

2) Labeling Setelah Cleaning

Setelah stopwords dihapus dan stemming dilakukan, proses labeling diulang untuk melihat perubahan klasifikasi, sebagaimana ditampilkan dalam Tabel 5.

TABEL V
HASIL LABELING DATA SETELAH CLEANING

Komentar Setelah Cleaning	Label
maxwin ribet delicuan	spam
special member bos jamin gacor soda69 bos moga manfaat bos moga work it akun bos link	spam
bukti bir123	spam
computer 100k	non-spam
kaya skylar gebug aran	non-spam
pulau777 fitur keren	spam

Hasil menunjukkan komentar spam tetap dapat dikenali, sementara komentar non-spam tetap tidak berubah. Cleaning membantu membuat teks lebih ringkas tanpa mengurangi akurasi deteksi.

F. Pemodelan Machine Learning dan Evaluasi Model

Pada tahap pemodelan, dilakukan eksperimen untuk mencari kombinasi parameter terbaik dari algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes (NB) menggunakan GridSearchCV. Pada KNN, jumlah tetangga yang diuji adalah 3, 5, 7, dan 9, sedangkan pada Naïve Bayes, nilai alpha diuji dengan 0,1, 0,5, dan 1,0.

Setelah mendapatkan parameter terbaik, model kemudian dievaluasi menggunakan dua metode ekstraksi fitur, yaitu Term Frequency-Inverse Document Frequency (TF-IDF) dan Bag of Words (BoW). Performa model diukur menggunakan metrik akurasi, dengan F1-score sebagai metrik utama pada hasil akurasi parameter karena mampu menangkap keseimbangan antara precision dan recall dalam mendeteksi komentar spam.

1) Hasil Akurasi Parameter Setelah Normalisasi

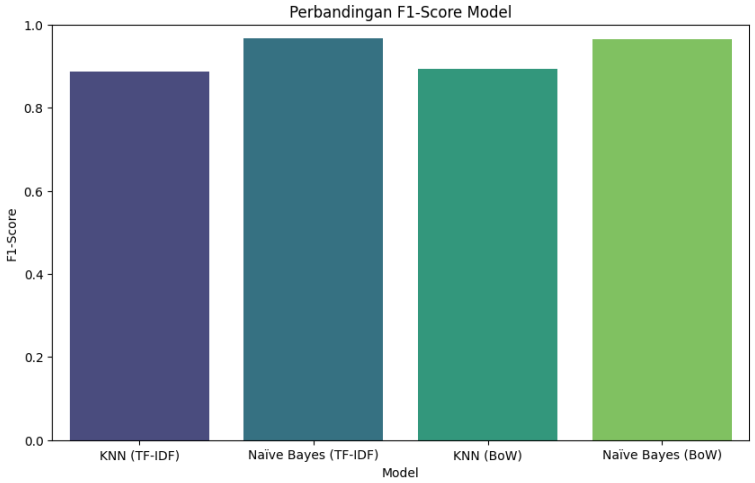
Berdasarkan Tabel 6, model Naïve Bayes menunjukkan performa terbaik dibandingkan KNN.

Tabel VI
HASIL AKURASI PARAMETER SETELAH NORMALISASI

Model	Parameter	F1-Score
KNN	3	0,882894
	5	0,851339
	7	0,834506
	9	0,818445
Naïve Bayes	0,1	0,969047
	0,5	0,968127
	1,0	0,967872

Model Naïve Bayes dengan alpha 0,1 memiliki F1-score tertinggi sebesar 96,9%, sedikit lebih unggul dibandingkan Naïve Bayes dengan alpha lainnya. Sementara itu, KNN menunjukkan performa terbaik pada k=3 dengan F1-score sebesar 88,2%, tetapi masih lebih rendah dibandingkan Naïve Bayes.

Sebagai perbandingan, performa F1-score dari semua model setelah normalisasi divisualisasikan dalam grafik Gambar 6.



Gambar. 6. Perbandingan F1-Score Model Setelah Normalisasi

2) Evaluasi Model Setelah Normalisasi

Setelah mendapatkan parameter terbaik, model dievaluasi lebih lanjut menggunakan metode ekstraksi fitur TF-IDF dan BoW. Hasil evaluasi akurasi untuk masing-masing model dapat dilihat pada Tabel 7, di mana "1" mewakili spam dan "0" mewakili non-spam.

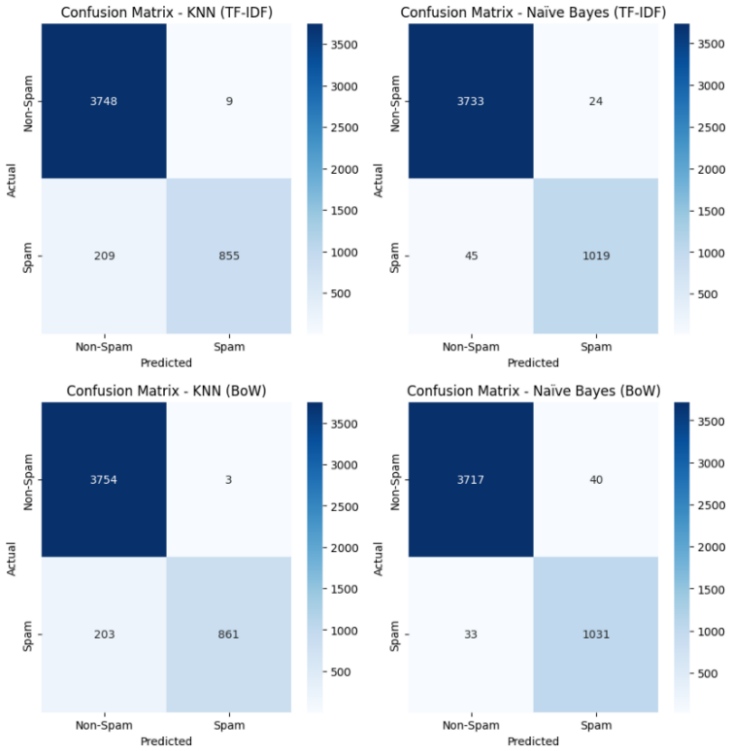
TABEL VII
PERBANDINGAN SEMUA MODEL SETELAH NORMALISASI

Model	Precision (0)	Recall (0)	F1-Score (0)	Precision (1)	Recall (1)	F1-Score (1)	Accuracy
KNN (TF-IDF)	0,95	1,00	0,97	0,99	0,80	0,89	0,95
Naïve Bayes (TF-IDF)	0,99	0,99	0,99	0,98	0,96	0,97	0,99
KNN (BoW)	0,95	1,00	0,97	1,00	0,81	0,89	0,96
Naïve Bayes (BoW)	0,99	0,99	0,99	0,96	0,97	0,97	0,98

Dari hasil evaluasi, terlihat bahwa Naïve Bayes dengan TF-IDF tetap unggul dengan akurasi 99%, menunjukkan performa yang sangat baik dalam mendeteksi komentar spam. Keunggulan metode TF-IDF dalam eksperimen ini tidak hanya terlihat dari nilai akurasi, tetapi juga dapat dijelaskan dari segi prinsip kerjanya. TF-IDF mampu menekan bobot kata-kata umum yang sering muncul baik di komentar [26][32] spam maupun non-spam, seperti "bosku" atau "link", sehingga kata-kata tersebut tidak mendominasi dalam pembentukan fitur. Dengan demikian, model lebih fokus pada kata-kata yang lebih spesifik terhadap kelas tertentu, meningkatkan kemampuan dalam membedakan komentar spam dan non-spam. Hal ini membuat model berbasis TF-IDF lebih tahan terhadap noise dibandingkan BoW, yang hanya mengandalkan frekuensi tanpa mempertimbangkan distribusi antar dokumen [33].

Untuk memahami bagaimana model melakukan klasifikasi setelah normalisasi, digunakan Confusion Matrix, sebagaimana ditunjukkan dalam Gambar 7, yang memperlihatkan distribusi prediksi model terhadap data aktual.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274



Gambar. 7. Confusion Matrix Semua Model Setelah Normalisasi

Selain itu, dilakukan pula analisis terhadap sejumlah komentar yang salah diklasifikasikan, baik sebagai false negative maupun false positive. Misalnya, model KNN berbasis TF-IDF masih gagal mengenali komentar spam seperti "Baru coba main di AERO88 eh rezekinya langsung gacir mantap banget", "idihh aksi amat sangat nih web AMBIL4D ngerasain maxwin juga", dan "Jangan hanya bermimpi tentang kemenangan raihlah bersama MAUT88". Ketiganya diklasifikasikan sebagai non-spam, padahal jelas mengandung unsur promosi situs judi online. Kemungkinan besar, kesalahan ini terjadi karena komentar tersebut menggunakan kata-kata yang terdengar positif atau umum seperti "mantap", "gacir", dan "kemenangan", serta variasi penulisan nama situs judi yang tidak tercakup dalam daftar kata kunci atau fitur yang dikenali model.

Sebaliknya, terdapat pula komentar non-spam yang diklasifikasikan sebagai spam, seperti "sholat di M4SJ1D gacor banget pahalanya" dan "ortu cewek pasti bangga lihat anaknya sudah bisa merokok". Kedua komentar ini memiliki konteks yang sebenarnya tidak berkaitan dengan promosi judi, namun penggunaan istilah seperti "gacor" atau penyampaian yang ambigu membuat model salah mengenalinya sebagai spam. Ini menunjukkan bahwa meskipun model cukup efektif secara umum, masih ada tantangan dalam mengenali konteks sebenarnya dari kata atau frasa tertentu yang multitafsir.

3) Hasil Akurasi Parameter Setelah Cleaning

Berdasarkan hasil akurasi setelah cleaning pada Tabel 8, model Naïve Bayes tetap menunjukkan performa terbaik dibandingkan KNN.

TABEL VIII
HASIL AKURASI PARAMETER SETELAH CLEANING

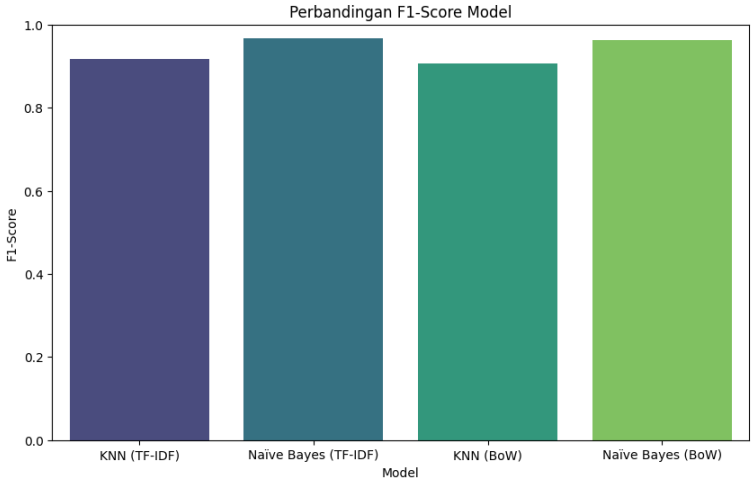
Model	Parameter	F1-Score
KNN	3	0,899202
	5	0,869404
	7	0,842971
	9	0,822246
Naïve Bayes	0,1	0,964449
	0,5	0,962457
	1,0	0,962554

Model Naïve Bayes dengan alpha 0,1 memiliki F1-score tertinggi sebesar 96,4%, sedikit lebih rendah dibandingkan dengan hasil setelah normalisasi sebesar 96,9%. Penurunan ini terjadi karena pembersihan data melalui penghilangan stopwords, stemming dengan pustaka Sastrawi dan penghapusan komentar tidak relevan

dapat menghilangkan konteks kata yang penting dalam distribusi probabilistik, yang sangat berpengaruh pada kinerja model Naïve Bayes. Meskipun data menjadi lebih bersih, penghapusan kata-kata umum dan perubahan kata ke bentuk dasarnya bisa menyebabkan hilangnya informasi yang sebelumnya berguna untuk klasifikasi.

Sementara itu, model KNN dengan nilai k=3 justru menunjukkan peningkatan F1-score dari 88,3% menjadi 89,9% setelah cleaning. Hal ini karena KNN lebih bergantung pada jarak antar representasi teks. Dengan teks yang sudah dibersihkan dari kata-kata tidak penting dan distandarkan melalui stemming, representasi vektor menjadi lebih konsisten, sehingga model KNN dapat melakukan pencocokan antar data lebih akurat dan efektif.

Sebagai perbandingan, performa F1-score dari semua model setelah cleaning dapat dilihat pada Gambar 8.



Gambar. 8. Perbandingan F1-Score Model Setelah Cleaning

4) Evaluasi Model Setelah Cleaning

Model kembali dievaluasi menggunakan metode ekstraksi fitur TF-IDF dan BoW setelah mendapatkan parameter terbaik. Hasil evaluasi akurasi untuk masing-masing model dapat dilihat pada Tabel 9, di mana "1" mewakili spam dan "0" mewakili non-spam.

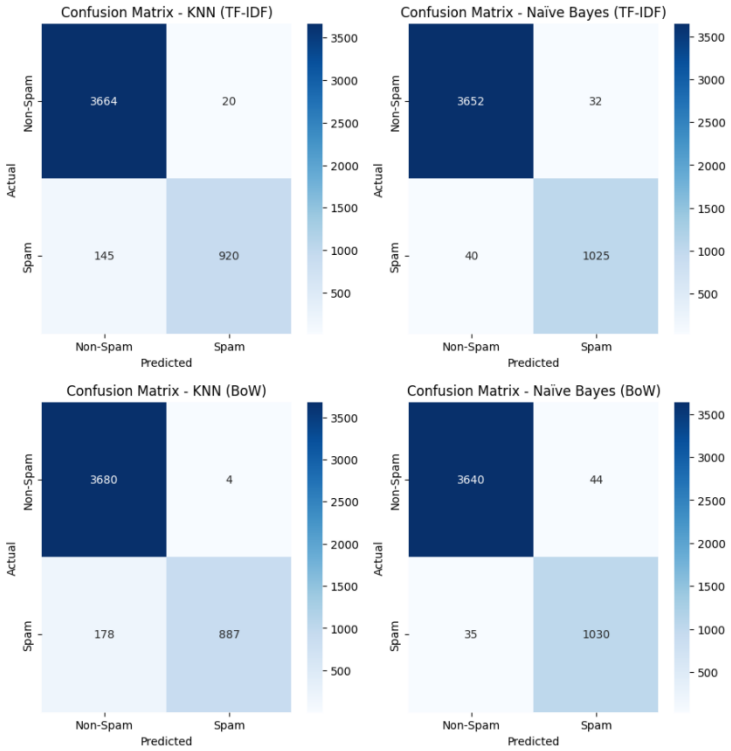
TABEL IX
PERBANDINGAN SEMUA MODEL SETELAH CLEANING

Model	Precision (0)	Recall (0)	F1-Score (0)	Precision (1)	Recall (1)	F1-Score (1)	Accuracy
KNN (TF-IDF)	0,96	0,99	0,98	0,98	0,86	0,92	0,97
Naïve Bayes (TF-IDF)	0,99	0,99	0,99	0,97	0,96	0,97	0,98
KNN (BoW)	0,95	1,00	0,98	1,00	0,83	0,91	0,96
Naïve Bayes (BoW)	0,99	0,99	0,99	0,96	0,97	0,96	0,98

Dari hasil evaluasi, terlihat bahwa Naïve Bayes dengan TF-IDF tetap unggul dengan akurasi sebesar 98%. Meskipun Naïve Bayes dengan BoW juga mencapai akurasi yang sama, model dengan TF-IDF memiliki keseimbangan precision dan recall yang lebih optimal, sebagaimana ditunjukkan oleh F1-score yang lebih stabil. Hal ini menunjukkan bahwa metode TF-IDF lebih efektif dalam menangkap pola bahasa komentar spam dibandingkan BoW, sehingga meningkatkan akurasi deteksi dengan lebih sedikit kesalahan klasifikasi. Keunggulan ini dapat dijelaskan lebih dalam melalui karakteristik metode TF-IDF, di mana kata-kata yang muncul terlalu sering di semua komentar akan diberikan bobot rendah melalui mekanisme IDF (Inverse Document Frequency) [34]. Dengan demikian, kata-kata umum yang tidak membantu dalam klasifikasi akan diabaikan, dan fokus model diarahkan pada kata-kata yang lebih relevan untuk membedakan spam dan non-spam.

Untuk menganalisis kinerja model setelah proses cleaning, digunakan Confusion Matrix sebagaimana ditampilkan dalam Gambar 9, yang menunjukkan distribusi hasil prediksi model terhadap data aktual.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.261-274



Gambar. 9. Confusion Matrix Semua Model Setelah Cleaning

Confusion Matrix pada Gambar 9 menunjukkan peningkatan akurasi model setelah proses cleaning, namun masih terdapat sejumlah komentar yang diklasifikasikan secara keliru. Model Naïve Bayes dengan TF-IDF, misalnya, salah mengklasifikasikan komentar seperti "main aero88 secara otomatis hati bahagia rezeki atur jamin" dan "untung maen s1tus ligalexus lek gue lek s1tus b0nus enak maen ligalexus" sebagai non-spam. Kedua komentar tersebut menggunakan teknik pengaburan seperti penulisan alternatif (misalnya "s1tus", "b0nus") yang mungkin luput dari deteksi algoritma. Selain itu, pola bahasa yang menyerupai kalimat motivasi juga berkontribusi terhadap kesalahan klasifikasi ini.

Sementara itu, ada juga komentar yang sebenarnya tidak mengandung unsur spam tetapi dikenali sebagai spam, seperti "terima nyata ngapain ngambek" dan "wardennya kasih jumpspell healernya kasih clonespell". Kedua komentar ini mungkin diklasifikasikan salah karena struktur kalimatnya tidak umum atau mengandung kata yang dianggap tidak relevan, padahal dari konteksnya tidak ada unsur promosi atau ajakan tertentu. Kasus-kasus ini menunjukkan pentingnya pemahaman kontekstual dalam pemodelan deteksi spam, khususnya pada domain komentar YouTube yang sangat beragam gaya bahasanya.

Berdasarkan Confusion Matrix setelah normalisasi dan cleaning, model Naïve Bayes dengan TF-IDF menunjukkan kemampuan yang lebih baik dalam menjaga keseimbangan antara precision, recall dan F1-score, yang menghasilkan kesalahan klasifikasi lebih rendah dibandingkan model lainnya. Model ini lebih efektif dalam mendeteksi komentar spam, dengan akurasi yang lebih tinggi dalam mengklasifikasikan baik komentar non-spam maupun spam. Sebaliknya, KNN memiliki kinerja yang kurang optimal pada spam, karena model ini terlalu cenderung memprediksi komentar spam sebagai non-spam, yang mengarah pada tingginya angka false negatives.

Tingginya false negatives pada model K-Nearest Neighbors (KNN) disebabkan oleh beberapa faktor, antara lain ketergantungan pada kedekatan fitur yang dapat menyesatkan model dalam membedakan komentar spam dari komentar non-spam, kelemahan dalam menangkap nuansa bahasa, serta pengaruh dimensionalitas tinggi yang menyebabkan jarak antar vektor menjadi kurang bermakna. Selain itu, ketidakseimbangan jumlah komentar spam dan non-spam di dalam dataset juga dapat memengaruhi kinerja model.

Implikasi dari tingginya false negatives ini adalah risiko kegagalan dalam mendeteksi spam, yang mengarah pada tetap terjadinya komentar spam di platform, berpotensi merusak kredibilitas kanal dan berisiko hukum atau finansial. Sebagai perbandingan, model Naive Bayes lebih efektif dalam mendeteksi spam tanpa menghasilkan false negatives, karena pendekatan probabilistiknya yang lebih mampu menangkap pola kata-kata spam yang tersembunyi.

Secara keseluruhan, hasil eksperimen menunjukkan bahwa pemilihan metode ekstraksi fitur dan algoritma klasifikasi berpengaruh signifikan terhadap performa model. Naïve Bayes dengan TF-IDF terbukti menjadi metode

terbaik dalam mendeteksi komentar spam judi online di YouTube, dengan performa yang lebih stabil dan deteksi spam yang lebih akurat dibandingkan model lainnya.

Meskipun penelitian ini berfokus pada algoritma klasik seperti Naïve Bayes dan K-Nearest Neighbors, penting untuk memposisikan hasil ini dalam konteks perkembangan terkini di bidang deteksi spam, khususnya dengan munculnya pendekatan berbasis transformer seperti BERT (Bidirectional Encoder Representations from Transformers). Model berbasis transformer telah menunjukkan performa unggul dalam banyak tugas pemrosesan bahasa alami [35], termasuk klasifikasi teks [36] dan deteksi spam [22][37], berkat kemampuannya memahami konteks kata secara mendalam dan menangani ambiguitas bahasa [38]. Beberapa studi sebelumnya menunjukkan bahwa BERT dan turunannya mampu mencapai F1-score yang sangat tinggi dalam klasifikasi spam, bahkan dalam domain informal seperti media social [39] dan komentar daring [40][41].

Namun demikian, model-model tersebut umumnya membutuhkan sumber daya komputasi yang besar dan waktu pelatihan yang lebih lama dibandingkan metode klasik. Dalam konteks ini, hasil eksperimen menunjukkan bahwa Naïve Bayes yang dikombinasikan dengan TF-IDF mampu menghasilkan performa sangat kompetitif dengan akurasi 99% setelah normalisasi dan 98% setelah cleaning, meskipun dengan kompleksitas dan kebutuhan komputasi yang jauh lebih rendah. Ini menandakan bahwa untuk aplikasi berskala kecil hingga menengah atau sistem yang membutuhkan efisiensi, pendekatan klasik seperti Naïve Bayes masih sangat relevan dan efektif. Kontribusi penelitian ini terletak pada pembuktian bahwa dengan preprocessing yang tepat dan pemilihan fitur yang sesuai, model konvensional tetap mampu memberikan kinerja tinggi dalam mendeteksi komentar spam, tanpa memerlukan infrastruktur canggih seperti yang dibutuhkan oleh model transformer.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa Naïve Bayes dengan TF-IDF adalah model terbaik untuk mendeteksi komentar spam judi online di YouTube, dengan akurasi 99% setelah normalisasi dan 98% setelah cleaning. Model ini lebih unggul dibandingkan Naïve Bayes dengan BoW, karena memiliki keseimbangan precision, recall, dan F1-score yang lebih baik. Sementara itu, KNN kurang efektif dalam tugas ini. Hasil juga menegaskan bahwa TF-IDF lebih baik daripada BoW dalam menangkap pola bahasa spam. Sebagai saran, penelitian selanjutnya dapat mengeksplorasi algoritma lain seperti SVM atau Deep Learning, menerapkan Word Embeddings untuk pemahaman konteks yang lebih baik, serta memperluas dataset agar model lebih general. Implementasi dalam bot moderasi otomatis juga dapat menjadi langkah berikutnya untuk mendeteksi spam secara real-time dan menjaga kualitas diskusi di YouTube.

UCAPAN TERIMA KASIH

Terima Kasih yang tulus diberikan kepada Universitas Papua, Dosen Pembimbing, Orang tua dan teman-teman serta semua pihak yang telah mendukung dan membimbing pada penelitian ini.

DAFTAR PUSTAKA

[1] A. R. Chrismanto, E. Winarko, dan Y. Suyanto, "EiAP-BC: A Novel Emoji Aware Inter-Attention Pair Model for Contextual Spam Comment Detection Based on Posting Text," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 23, no. 12, hal. 1–30, Des 2024, doi: 10.1145/3696663.

[2] Y. Pirnando, Y. Kusnadi, dan P. Angkupi, "Mekanisme Dan Regulasi Penegakan Hukum Terhadap Streamer Game yang Menyampaikan Informasi Tentang Judi Online Oknum yang mempromosikan platform judi online ke media online karena turut mendukung terjadinya perjudian online , Penelitian ini mengeksplorasi," *Muhammadiyah Law Rev.*, vol. 8, no. 1, hal. 1–8, 2024.

[3] World Health Organization, "Gambling," World Health Organization. Diakses: 24 April 2025. [Daring]. Tersedia pada: https://www.who.int/news-room/fact-sheets/detail/gambling

[4] I. Network, "Spam Komentar Judi Online di YouTube Semakin Meresahkan, Begini Cara Mengatasinya," ION Network. Diakses: 14 Maret 2025. [Daring]. Tersedia pada: https://blog.ionnetwork.co.id/spam-komentar-judi-online-di-youtube-semakin-meresahkan-begini-cara-mengatasinya/

[5] L. Kristianti, "Cara Google tangani komentar judi online yang berseliweran di YouTube," ANTARA. Diakses: 14 Maret 2025. [Daring]. Tersedia pada: https://www.antaranews.com/berita/4657029/cara-google-tangani-komentar-judi-online-yang-berseliweran-di-youtube/

[6] A. N. Salim, A. Adryani, dan T. Sutabri, "Deteksi Email Spam dan Non-Spam Berdasarkan Isi Konten Menggunakan Metode K-Nearest Neighbor dan Support Vector Machine," *Syntax Idea*, vol. 6, no. 2, hal. 991–1001, Mar 2024, doi: 10.46799/syntax-idea.v6i2.3052.

[7] A. Wara Putera dan Y. Dewi Lestari, "Klasifikasi SMS Spam Menggunakan Algoritma K-Nearest Neighbor," *Jikstra*, vol. 5, no. 01, hal. 43–55, 2023.

[8] P. V Dhote, R. Ramteke, A. R. Patel, P. Dewalker, dan P. A. Chaube, "Instagram Spam Detection (ISD)," vol. 8, no. 5, hal. 573–583, 2024, [Daring]. Tersedia pada: www.ijtsrd.com/papers/ijtsrd69419.pdf

[9] L. Das, L. Ahuja, dan A. Pandey, "Analysis of Twitter Spam Detection Using Machine Learning Approach," *Proc. 3rd Int. Conf. Intell. Eng. Manag. ICIEM 2022*, no. April 2022, hal. 764–769, 2022, doi: 10.1109/ICIEM54221.2022.9853100.

[10] A. R. Chrismanto, A. K. Sari, dan Y. Suyanto, "Enhancing Spam Comment Detection on Social Media With Emoji Feature and Post-Comment Pairs Approach Using Ensemble Methods of Machine Learning," *IEEE Access*, vol. 11, hal. 80246–80265, 2023, doi: 10.1109/ACCESS.2023.3299853.

[11] A. Sinhal dan M. Maheshwari, "YouTube: Spam Comments Filtration using Hybrid Ensemble Machine Learning Models," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 12, no. 10, hal. 169–182, 2022, doi: 10.46338/ijetae1022_18.

[12] G. Airlangga, "Spam Detection on YouTube Comments Using Advanced Machine Learning Models: A Comparative Study," *Brill. Res. Artif.*

Intell., vol. 4, no. 2, hal. 500–508, Okt 2024, doi: 10.47709/BRILLIANCE.V4I2.4670.
[13] H. Sankar Dutta, "YT-30M: A multi-lingual multi-category dataset of YouTube comments," *arXiv*, hal. arXiv:2412.03465, Des 2024, doi: 10.48550/ARXIV.2412.03465.
[14] S. Teng, K. W. Khong, S. P. Sharif, dan A. Ahmed, "Youtube video comments on healthy eating: Descriptive and predictive analysis," *JMIR Public Heal. Surveill.*, vol. 6, no. 4, 2020, doi: 10.2196/19618.
[15] R. B. Perdana, Ardin, I. Budi, A. B. Santoso, A. Ramadiah, dan P. K. Putra, "Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 8, hal. 942–949, 2024, doi: 10.14569/IJACSA.2024.0150893.
[16] A. Maulana dan A. Yuliana, "ANALISIS SENTIMEN OPINI PUBLIK TERKAIT JUDI ONLINE PADA PENGGUNA APLIKASI X MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MECHINE," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, hal. 3706–3714, Okt 2024, doi: 10.23960/jitet.v12i3S1.5187.
[17] M. F. A. Shidiq dan D. Alita, "ANALISIS SENTIMEN MASYARAKAT TERHADAP KASUS JUDI ONLINE MENGGUNAKAN DATA DARI MEDIA SOSIAL X PENDEKATAN NAIVE BAYES DAN SVM," *Simika*, vol. 8, no. 1, hal. 24–35, 2025.
[18] O. N. Julianti, N. Suarna, dan W. Prihartono, "PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANALISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, hal. 2936–2941, Mei 2024, doi: 10.36040/jati.v8i3.9613.
[19] K. Tri Putra, M. Amin Hariyadi, dan C. Crysdian, "Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm," *J. Cahaya MAndalika*, hal. 1449–1463, 2023.
[20] P. Joseph dan S. Y. Yerima, "A comparative study of word embedding techniques for SMS spam detection," *Proc. - 2022 14th IEEE Int. Conf. Comput. Intell. Commun. Networks, CICN 2022*, hal. 149–155, 2022, doi: 10.1109/CICN56167.2022.10008245.
[21] M. Ahmadi, M. Khajavi, A. Varmaghani, A. Ala, K. Danesh, dan D. Javaheri, "Leveraging Large Language Models for Cybersecurity: Enhancing SMS Spam Detection with Robust and Context-Aware Text Classification," *Arxiv*, Feb 2025, Diakses: 26 April 2025. [Daring]. Tersedia pada: https://arxiv.org/pdf/2502.11014
[22] S. Rojas-Galeano, "Using BERT Encoding to Tackle the Mad-lib Attack in SMS Spam Detection," 2021, [Daring]. Tersedia pada: http://arxiv.org/abs/2107.06400
[23] M. D. Taufiqurahman, S. Anraeni, dan H. Darwis, "Analisis Sentimen Komentar Konten Kreator Gaming Menggunakan Metode Naive Bayes dan KNN," *LINIER Lit. Inform. dan Komput.*, vol. 1, no. 4, hal. 317–327, 2024.
[24] A. Wijaya, M. Rivaldo, dan M. R. Pribadi, "Analisis Sentimen Terhadap Aplikasi Mitra Darat Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor," *AICOMS Appl. Inf. Technol. Comput. Sci.*, vol. 3, no. 1, hal. 9–14, 2024.
[25] M. Iksan Maulana, M. Martanto, dan U. Hayati, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Topik Berita Pada Situs Detik.Com," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, hal. 3733–3742, 2024, doi: 10.36040/jati.v8i3.9779.
[26] S. Akuma, T. Lubem, dan I. T. Adom, "Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets," *Int. J. Inf. Technol.*, vol. 14, no. 7, hal. 3629–3635, 2022, doi: 10.1007/s41870-022-01096-4.
[27] D. N. Larasakti, A. Aziz, dan D. Aditya, "Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 5, hal. 132–142, 2023.
[28] A. Purnama dan D. Hamidin, "Metode Algoritma Logistic Regression dalam Klasifikasi Email Spam," *SHIFT*, vol. 5, no. 1, hal. 39–47, 2025.
[29] L. N. AINA, V. R. S. NASTITI, dan C. S. K. ADITYA, "Implementasi Extra Trees Classifier dengan Optimasi Grid Search CV pada Prediksi Tingkat Adaptasi," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 9, no. 1, hal. 78–88, 2024, [Daring]. Tersedia pada: https://ejurnal.itenas.ac.id/index.php/mindjournal/article/view/10899
[30] G. Hakim, T. N. Fatyanosa, dan A. W. Widodo, "Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT," *JPTIIK J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 10, hal. 1–10, 2024.
[31] M. Z. Ramadhan dan R. Mubarak, "ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN BAHAN BAKAR MINYAK PADA MEDIA SOSIAL YOUTUBE DENGAN METODE K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, hal. 984–991, 2025.
[32] S. Dai *et al.*, "AI-based NLP section discusses the application and effect of bag-of-words models and TF-IDF in NLP tasks," *J. Artif. Intell. Gen. Sci. ISSN3006-4023*, vol. 5, no. 1, hal. 13–21, Jun 2024, doi: 10.60087/JAIGS.V5I1.149.
[33] A. F. Nuraini, R. D. Pertiwi, M. Z. Subarkah, dan K. Ferawati, "Analisis Sentimen Pelaksanaan Vaksinasi Covid-19 secara Massal pada Media Sosial Twitter," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, hal. 1385–1394, Nov 2022, doi: 10.34123/SEMNASOFFSTAT.V2022I1.1564.
[34] A. Addiga, S. Bagui, A. Addiga, dan S. Bagui, "Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency," *J. Comput. Commun.*, vol. 10, no. 8, hal. 117–128, Agu 2022, doi: 10.4236/JCC.2022.108008.
[35] A. Filali, E. A. Abdellaoui Alaoui, dan M. Merras, "Enhancing spam detection with GANs and BERT embeddings: A novel approach to imbalanced datasets," *Procedia Comput. Sci.*, vol. 236, hal. 420–427, 2024, doi: 10.1016/j.procs.2024.05.049.
[36] A. R. Hanum *et al.*, "Analisis Kinerja Algoritma Klasifikasi Teks Bert dalam Mendeteksi Berita Hoaks," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, hal. 537–546, Jul 2024, doi: 10.25126/JTIIK.938093.
[37] Y. Guo, Z. Mustafaoglu, dan D. Koundal, "Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms," *J. Comput. Cogn. Eng.*, vol. 2, no. 1, hal. 5–9, Apr 2022, doi: 10.47852/bonviewJCCE2202192.
[38] K. A. DeLong, S. Trott, dan M. Kutas, "Offline dominance and zeugmatic similarity normings of variably ambiguous words assessed against a neural language model (BERT)," *Behav. Res. Methods*, vol. 55, no. 4, hal. 1537–1557, 2023, doi: 10.3758/s13428-022-01869-6.
[39] B. M. Jones dan M. Omar, "Detection of Twitter Spam with Language Models: A Case Study on How to Use BERT to Protect Children from Spam on Twitter," *Proc. - 2023 Congr. Comput. Sci. Comput. Eng. Appl. Comput. CSCE 2023*, hal. 511–516, 2023, doi: 10.1109/CSCE60160.2023.00089.
[40] T. Sahmoud dan M. Mikki, "Spam Detection Using BERT," *Arxiv*, hal. 2–7, Jun 2022, doi: 10.48550/arXiv.2206.02443.
[41] M. I. H. Emon, K. N. Iqbal, M. H. K. Mehedi, M. J. A. Mahbub, dan A. A. Rasel, "Detection of Bangla Hate Comments and Cyberbullying in Social Media Using NLP and Transformer Models," in *Communications in Computer and Information Science*, vol. 1613 CCIS, Springer, Cham, 2022, hal. 86–96. doi: 10.1007/978-3-031-12638-3_8.