

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK DETEKSI SPAM KOMENTAR PADA PLATFORM TIKTOK

Mochamad Rifal Andriansyah^{1*}, Asep Budiman Kusdinar², Winda Apriandari³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Sukabumi; Jl. R. Syamsudin, S.H. No. 50, Kota Sukabumi, Jawa Barat 43113

Keywords: Deteksi Spam; TikTok; Naïve Bayes; Support Vector Machine; CRISP-DM; Text Mining.

Correspondent Email:
2230511036@ummi.ac.id

Abstrak. Perkembangan platform media sosial TikTok yang pesat menyebabkan meningkatnya interaksi pengguna dalam bentuk komentar, namun juga memunculkan permasalahan spam berupa promosi tidak relevan, tautan mencurigakan, konten penipuan, dan pesan bot. Penelitian ini bertujuan membangun model klasifikasi komentar spam dan non-spam pada TikTok menggunakan algoritma Naïve Bayes serta membandingkan performanya dengan Support Vector Machine (SVM). Metode penelitian menggunakan kerangka kerja CRISP-DM yang meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Dataset terdiri dari 1.300 komentar TikTok yang terbagi seimbang menjadi 650 komentar spam dan 650 komentar non-spam. Preprocessing teks meliputi case folding, cleansing, tokenizing, stopword removal, dan stemming, dilanjutkan pembobotan fitur TF-IDF ($\text{max_features}=2000$, $\text{ngram_range}=(1,2)$). Hasil pengujian menunjukkan Naïve Bayes mencapai akurasi, precision, recall, dan F1-score sebesar 98,46%, sedangkan SVM mencapai performa sempurna 100% pada seluruh metrik. Perbandingan menunjukkan SVM lebih unggul dibandingkan Naïve Bayes dalam mendeteksi komentar spam pada TikTok. Penelitian ini diharapkan menjadi referensi pengembangan sistem moderasi komentar otomatis berbasis machine learning pada media sosial.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. The rapid growth of the TikTok platform has increased user interaction through comments, but has also given rise to spam problems such as irrelevant promotions, suspicious links, fraudulent content, and bot messages. This study aims to build a classification model for spam and non-spam comments on TikTok using the Naïve Bayes algorithm and to compare its performance with Support Vector Machine (SVM). The research method refers to the CRISP-DM framework, covering Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The dataset consists of 1,300 TikTok comments evenly divided into 650 spam and 650 non-spam comments. Text preprocessing includes case folding, cleansing, tokenizing, stopword removal, and stemming, followed by TF-IDF feature weighting ($\text{max_features}=2000$, $\text{ngram_range}=(1,2)$). Results show that Naïve Bayes achieved 98.46% accuracy, precision, recall, and F1-score, while SVM achieved a perfect 100% on all metrics. The comparison shows that SVM outperforms Naïve Bayes in detecting spam comments on TikTok. This research is expected to serve as a reference for developing automated, machine-learning-based comment moderation systems on social media.

1. PENDAHULUAN

Pertumbuhan teknologi internet berpengaruh besar terhadap pemanfaatan media sosial sebagai alat komunikasi, berbagi informasi, promosi, kampanye, dan

penyampaian pendapat masyarakat. Salah satu platform media sosial yang cukup populer adalah TikTok, yaitu platform berbasis video pendek dengan jumlah pengguna yang sangat besar dan terus meningkat, termasuk di Indonesia [1]. Tingginya aktivitas pengguna

menyebabkan meningkatnya interaksi dalam bentuk komentar pada setiap konten yang diunggah. Namun, tingginya jumlah komentar juga menimbulkan permasalahan berupa munculnya komentar spam, yaitu komentar promosi berulang, tautan mencurigakan, komentar tidak relevan, serta pesan yang dikirim secara massal oleh akun tertentu.

Keberadaan spam dapat mengganggu kenyamanan pengguna, menurunkan kualitas interaksi, serta berpotensi menimbulkan risiko keamanan digital. Berbeda dengan platform lain seperti Twitter atau Instagram, komentar pada TikTok memiliki karakteristik unik, seperti penggunaan bahasa informal, singkatan, emoji, serta pola komentar yang singkat dan repetitif. Penelitian deteksi spam telah banyak dilakukan pada platform lain seperti Twitter, SMS, dan email [2], [13], [15], namun penelitian yang secara khusus membahas komentar TikTok masih terbatas, dan belum banyak penelitian yang melakukan perbandingan algoritma klasifikasi pada data komentar TikTok.

Data tekstual pada media sosial umumnya tidak standar, sehingga memerlukan tahapan preprocessing yang cermat, mencakup tokenisasi, pelarasan kosakata, penyaringan, dan stemming sebelum proses klasifikasi dilakukan [6]. Salah satu algoritma yang umum digunakan untuk klasifikasi pesan spam adalah Naïve Bayes, yaitu metode probabilistik yang efektif menangani data teks dan sering dijadikan metode dasar dalam penelitian text mining [3], [7], [14].

Untuk memperkuat hasil penelitian, performa Naïve Bayes dibandingkan dengan Support Vector Machine (SVM) sebagai algoritma pembanding. SVM dipilih karena bekerja dengan mencari hyperplane optimal yang memisahkan dua kelas data secara maksimal, sehingga menghasilkan batas keputusan yang lebih tegas dibandingkan metode probabilistik. SVM juga terbukti efektif pada data berdimensi tinggi seperti representasi TF-IDF, tidak mudah mengalami overfitting pada ruang fitur yang besar, serta mampu menangani data yang tidak sepenuhnya linier melalui fungsi kernel. Penelitian-penelitian sebelumnya menunjukkan bahwa SVM secara konsisten menghasilkan akurasi tinggi dalam

klasifikasi teks, termasuk deteksi spam pada media sosial [2], [9], [12].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk: (1) membangun model klasifikasi komentar spam dan non-spam pada platform TikTok menggunakan algoritma Naïve Bayes; (2) menerapkan tahapan preprocessing pada data komentar TikTok sebelum proses klasifikasi; dan (3) membandingkan performa algoritma Naïve Bayes dengan SVM dalam mendeteksi komentar spam pada platform TikTok. Kebaruan penelitian ini terletak pada penerapan sekaligus perbandingan algoritma Naïve Bayes dan SVM secara khusus pada data komentar TikTok berbahasa Indonesia, yang sejauh ini belum banyak dikaji, sehingga diharapkan dapat menjadi acuan pengembangan sistem moderasi komentar otomatis berbasis machine learning.

2. TINJAUAN PUSTAKA

2.1. *Kajian Keislaman*

Berdasarkan QS. Al-Hujurat ayat 6, umat Islam diperintahkan untuk melakukan tabayyun, yaitu memeriksa kebenaran suatu informasi sebelum mempercayai dan menyebarkannya. Prinsip ini sejalan dengan penelitian deteksi spam yang bertujuan mengidentifikasi komentar yang tidak valid, menyesatkan, atau tidak relevan, sehingga informasi yang diterima pengguna media sosial menjadi lebih terpercaya.

2.2. *Spam pada Media Sosial dan TikTok*

Spam pada media sosial merupakan pesan atau konten yang disebar secara massal dan tidak dikehendaki pengguna tanpa persetujuan terlebih dahulu, umumnya bertujuan mempromosikan produk, meluncurkan penipuan, menyebarkan informasi tidak benar, atau mengarahkan pengguna ke tautan mencurigakan. Pada TikTok, spam umumnya berupa komentar promosi berulang, tautan mencurigakan, komentar tidak relevan, serta pesan massal oleh akun tertentu [4]. Fenomena serupa juga banyak dikaji pada platform lain, misalnya komentar spam perjudian daring di YouTube [16] dan spam pada layanan pesan singkat (SMS) [13], [15].

2.3. *Text Mining dan Preprocessing*

Text mining merupakan metode untuk menggali informasi dari data teks tidak terstruktur dengan mengintegrasikan teknik machine learning, data mining, natural language processing, dan pencarian informasi [7]. Tahap preprocessing seperti case folding, tokenisasi, stopwords removal, dan stemming berperan penting dalam meningkatkan kualitas data serta memengaruhi hasil klasifikasi [6]. Pembobotan fitur dilakukan menggunakan metode TF-IDF, yang menggabungkan frekuensi kemunculan kata dalam dokumen dan tingkat kelangkaannya pada seluruh dokumen untuk mengubah teks menjadi representasi numerik [8].

2.4. *Naïve Bayes dan Support Vector Machine*

Naïve Bayes merupakan metode klasifikasi probabilistik berbasis Teorema Bayes yang mengasumsikan independensi antar atribut. Meskipun asumsi tersebut sederhana, algoritma ini terbukti memberikan kinerja klasifikasi yang baik pada data teks [3], [14]. Support Vector Machine (SVM) merupakan algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan data ke dalam kelas tertentu, dan banyak digunakan dalam klasifikasi teks karena mampu menangani data berdimensi tinggi secara efektif [9], [12].

2.5. *Confusion Matrix dan CRISP-DM*

Confusion matrix digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi terhadap data aktual, menghasilkan nilai True Positive, True Negative, False Positive, dan False Negative yang menjadi dasar perhitungan accuracy, precision, recall, dan F1-score [10]. CRISP-DM merupakan metodologi data mining yang terdiri atas tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment, dan telah terbukti membantu menghasilkan model yang sesuai dengan kebutuhan analisis data pada berbagai penelitian klasifikasi [11].

2.6. *Penelitian Terdahulu dan Research Gap*

Penelitian deteksi spam dan klasifikasi teks sebagian besar telah dilakukan pada platform email, Twitter, SMS, YouTube, dan artikel berbahasa Indonesia menggunakan algoritma Naïve Bayes, K-Nearest Neighbor, maupun SVM [2], [3], [12], [13], [15], [16]. Bahkan pada jurnal JITET sendiri, algoritma Naïve Bayes dan SVM telah diterapkan untuk analisis sentimen pada media sosial X [17]. Namun demikian, penelitian yang secara khusus membahas deteksi spam pada komentar TikTok, khususnya yang membandingkan performa Naïve Bayes dan SVM, masih terbatas. Tabel 1 merangkum posisi penelitian ini terhadap beberapa penelitian terdahulu.

Tabel 1. Perbandingan dengan Penelitian Terdahulu

Peneliti	Fokus Penelitian	Perbedaan
[7]	Text mining & sentimen Twitter dengan Naïve Bayes	Objek Twitter, bukan TikTok
[2]	Deteksi spam Twitter dengan Naïve Bayes	Tidak membandingkan SVM
[3]	Naïve Bayes vs K-NN pada artikel Indonesia	Pembanding K-NN, bukan SVM
[15]	Klasifikasi SMS spam Indonesia dengan SVM	Tidak membandingkan NB; objek SMS
[17]	Naïve Bayes & SVM untuk sentimen di platform X	Fokus sentimen, bukan deteksi spam

Berdasarkan tabel tersebut, penelitian ini mengisi celah dengan menggunakan data komentar TikTok berbahasa Indonesia serta membandingkan performa Naïve Bayes dan SVM untuk mendapatkan model klasifikasi spam yang optimal.

3. METODE PENELITIAN

Penelitian ini menggunakan metodologi data mining yang mengacu pada kerangka kerja CRISP-DM guna mengidentifikasi komentar spam pada platform TikTok. Tahapan CRISP-DM ditunjukkan pada Gambar 1.



Gambar 3.1. Tahapan Metode CRISP-DM
(Sumber: Dwi Ramo Sanjaya)

Gambar 1. Tahapan Metode CRISP-DM

3.1. *Business Understanding*

Tahap ini mengidentifikasi permasalahan penelitian, yaitu banyaknya komentar spam pada TikTok yang mengandung materi promosi, tautan berisiko, dan pesan berulang. Tujuan utama penelitian adalah mengembangkan model klasifikasi yang mampu mengkategorikan komentar TikTok menjadi spam dan non-spam menggunakan algoritma Naïve Bayes, dengan SVM sebagai pembanding.

3.2. *Data Understanding*

Data understanding dilakukan dengan mengumpulkan komentar dari beberapa video TikTok yang diperoleh dari dataset publik di platform Kaggle. Dataset terdiri dari 1.300 komentar yang telah dilabeli, terbagi seimbang menjadi 650 komentar spam dan 650 komentar non-spam.

3.3. *Data Preparation*

Data komentar diproses melalui tahapan text preprocessing yang meliputi: (1) case folding untuk menyeragamkan huruf menjadi huruf kecil; (2) cleansing untuk menghapus URL, simbol, angka, dan emoji; (3) tokenizing untuk memecah komentar menjadi token kata; (4) stopword removal untuk menghapus kata tidak bermakna termasuk kata informal seperti "nih", "sih", "deh"; dan (5) stemming menggunakan pustaka PySastrawi. Setelah preprocessing, data teks diubah menjadi representasi numerik menggunakan TF-IDF dengan parameter $\text{max_features}=2000$ dan $\text{ngram_range}=(1,2)$, menghasilkan matriks fitur berukuran $1.300 \text{ dokumen} \times 2.000 \text{ fitur}$.

3.4. *Modeling*

Dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan `train_test_split` dan `stratify=y` agar distribusi kelas seimbang, menghasilkan 1.040 data latih dan 260 data uji. Model dibangun menggunakan Multinomial Naïve Bayes ($\alpha=1.0$) sebagai metode utama, dan SVM kernel linear ($C=1.0$) sebagai metode pembanding.

3.5. *Evaluation*

Evaluasi dilakukan menggunakan confusion matrix untuk memperoleh nilai accuracy, precision, recall, dan F1-score, sebagaimana dirumuskan pada Persamaan (1) sampai (4) [10].

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Precision} = TP/(TP+FP) \quad (2)$$

$$\text{Recall} = TP/(TP+FN) \quad (3)$$

$$F1 = 2 \times (\text{Prec} \times \text{Rec}) / (\text{Prec} + \text{Rec}) \quad (4)$$

Evaluasi tambahan dilakukan melalui Stratified K-Fold Cross Validation yang dibungkus dalam satu pipeline scikit-learn bersama TF-IDF vectorizer, sehingga TF-IDF hanya di-fit pada data latih pada setiap fold dan tidak terjadi kebocoran informasi (data leakage) dari data uji.

3.6. *Deployment*

Tahap deployment berupa penyajian model klasifikasi dalam bentuk antarmuka interaktif berbasis ipywidgets pada lingkungan Google Colaboratory. Perlu ditegaskan bahwa sistem ini masih berupa prototipe yang belum terintegrasi dengan API resmi TikTok, sehingga pengujian bersifat simulasi input tunggal.

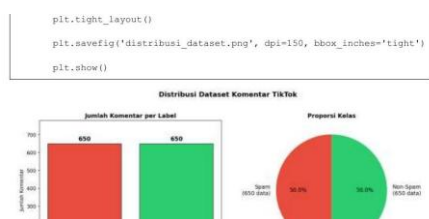
3.7. *Perangkat Penelitian*

Penelitian menggunakan komputer dengan sistem operasi Windows 11, prosesor AMD Ryzen 3 7320U, RAM 8 GB, dan SSD 474 GB. Perangkat lunak meliputi Google Colaboratory, Python, Microsoft Excel, serta pustaka scikit-learn dan PySastrawi.

4. HASIL DAN PEMBAHASAN

4.1. Data Understanding

Dataset yang digunakan terdiri dari 1.300 komentar TikTok yang telah melalui proses pembersihan data, terdiri dari kolom komentar dan kolom label berupa spam atau non-spam. Distribusi dataset menunjukkan keseimbangan sempurna antara kedua kelas, masing-masing 650 komentar, sebagaimana ditunjukkan pada Gambar 2. Data komentar spam mencakup komentar promosi dan iklan, penipuan berupa iming-iming hadiah atau pinjaman tanpa jaminan, ajakan mengikuti akun (sub4sub, follback), komentar bot, serta komentar kasar dan tidak relevan. Adapun komentar non-spam terdiri dari komentar apresiasi, pertanyaan terkait konten, diskusi, dan komentar netral yang relevan dengan topik video.



Gambar 2. Distribusi Komentar Spam dan Non-Spam

4.2. Data Preparation

Tahap data preparation mentransformasikan data mentah menjadi format bersih dan terstruktur melalui tahapan case folding, cleansing, tokenizing, stopwords removal, dan stemming. Contoh hasil perbandingan data sebelum dan sesudah preprocessing disajikan pada Tabel 2.

Tabel 2. Contoh Hasil Preprocessing Data

No	Komentar Asli	Hasil Preprocessing	Label
1	Videonya keren banget kak, sangat bermanfaat!	video keren manfaat	Non-Spam
2	Cek BIO aku ya KAK Ada DANA kaget 10JT hri ini!!	cek bio dana kaget hari	Spam
3	Klik link di profil menang hadiah jutaan rupiah	klik link profil menang hadiah juta rupiah	Spam
4	Sub4sub yuk kita tumbuh bareng	subsub tumbuh bareng	Spam

5	Makasih udah berbagi ilmu! Sangat membantu	kasih bagi ilmu bantu	Non-Spam
---	--	-----------------------	----------

Setelah preprocessing, data teks diubah menjadi representasi numerik menggunakan TF-IDF dengan parameter `max_features=2000` dan `ngram_range=(1,2)`, menghasilkan matriks fitur berukuran 1.300×2.000 yang digunakan sebagai input pada tahap pemodelan.

4.3. Modeling

Dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan `stratify=y` agar distribusi kelas tetap seimbang, sebagaimana ditunjukkan pada Tabel 3. Model dibangun menggunakan Multinomial Naïve Bayes ($\alpha=1.0$) dan SVM kernel linear ($C=1.0$), yang dilatih pada 1.040 data latih dan dievaluasi pada 260 data uji.

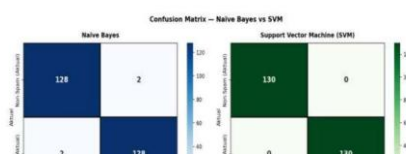
Tabel 3. Distribusi Data Latih dan Data Uji

Jenis Data	Jumlah	%
Data Latih	1.040	80%
Data Uji	260	20%
Total	1.300	100%

4.4. Evaluation

Hasil evaluasi model Naïve Bayes pada 260 data uji menghasilkan nilai True Positive sebesar 128, True Negative 128, False Positive 2, dan False Negative 2, sehingga diperoleh accuracy, precision, recall, dan F1-score masing-masing sebesar 98,46%. Model salah mengklasifikasikan 2 komentar non-spam sebagai spam (false positive) dan 2 komentar spam yang lolos sebagai non-spam (false negative).

Model SVM menghasilkan True Positive 130, True Negative 130, tanpa False Positive maupun False Negative sama sekali, sehingga mencapai accuracy, precision, recall, dan F1-score sempurna sebesar 100%. Confusion matrix kedua model ditunjukkan pada Gambar 3.

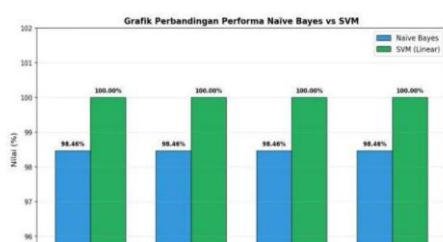


Gambar 3. Confusion Matrix Naïve Bayes dan SVM

Perbandingan performa kedua algoritma pada seluruh metrik evaluasi disajikan pada Tabel 4 dan Gambar 4.

Tabel 4. Perbandingan Performa Naïve Bayes dan SVM

Metrik	NB	SVM
Accuracy	98,46%	100%
Precision	98,46%	100%
Recall	98,46%	100%
F1-Score	98,46%	100%



Gambar 4. Grafik Perbandingan Performa Naïve Bayes vs SVM

Berdasarkan Gambar 4, algoritma SVM mengungguli Naïve Bayes pada seluruh metrik evaluasi, mencapai performa sempurna (100%), sedangkan Naïve Bayes memperoleh nilai 98,46% pada seluruh metrik. Hasil ini konsisten dengan temuan penelitian terdahulu yang menunjukkan keunggulan SVM pada klasifikasi teks berdimensi tinggi menggunakan representasi TF-IDF [9], [12], meskipun pada penelitian deteksi spam Twitter, Naïve Bayes juga terbukti mampu memberikan performa yang cukup baik [2]. Performa sempurna 100% pada data uji tersebut juga perlu dimaknai secara hati-hati dan tidak dapat diartikan sebagai model yang sempurna secara umum. Evaluasi tambahan melalui Stratified K-Fold Cross Validation dilakukan untuk memperoleh estimasi performa yang lebih stabil dan dapat direproduksi tanpa risiko kebocoran data.

4.5. Deployment

Model terbaik, yaitu SVM, diimplementasikan pada antarmuka interaktif berbasis ipywidgets di lingkungan Google Colaboratory yang memungkinkan pengguna memasukkan komentar baru untuk diklasifikasikan secara langsung oleh model Naïve Bayes dan SVM secara berdampingan. Pengujian terhadap komentar "klik link di bio"

berhasil diklasifikasikan sebagai spam oleh kedua model, meskipun kalimatnya singkat dan tidak memuat kata kunci spam yang eksplisit. Sebaliknya, komentar "pengen banget bisa bikin konten bagus ini" secara konsisten diklasifikasikan sebagai non-spam oleh kedua model.

Hasil pengujian tersebut menunjukkan bahwa sistem mampu mengenali pola komentar promosi maupun komentar apresiasi yang wajar, sekalipun menggunakan bahasa informal dan tidak baku. Perlu ditegaskan bahwa sistem yang dibangun masih berupa prototipe yang berjalan pada lingkungan notebook dan belum terintegrasi dengan API resmi TikTok, sehingga klaim kesiapan produksi (production-ready) dan performa pada kondisi real-time memerlukan penelitian lanjutan yang menggunakan data komentar TikTok secara langsung (live streaming comments).

5. KESIMPULAN

1. Tahapan preprocessing teks (case folding, cleansing, tokenizing, stopword removal, stemming) berhasil diterapkan pada dataset komentar TikTok. Pembobotan fitur TF-IDF ($\text{max_features}=2000$, $\text{ngram_range}=(1,2)$) menghasilkan matriks berukuran 1.300×2.000 sebagai representasi numerik dalam pemodelan.
2. Algoritma Naïve Bayes berhasil mendeteksi komentar spam dengan performa sangat baik, memperoleh akurasi, precision, recall, dan F1-score sebesar 98,46% pada 260 data uji, dengan 4 kesalahan klasifikasi (2 false positive dan 2 false negative).
3. Perbandingan menunjukkan algoritma SVM dengan kernel linear lebih unggul dibandingkan Naïve Bayes, mencapai akurasi, precision, recall, dan F1-score sempurna sebesar 100% tanpa kesalahan klasifikasi, sehingga SVM terbukti lebih efektif memisahkan kelas spam dan non-spam pada dataset komentar TikTok.
4. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan dikumpulkan langsung dari TikTok melalui web scraping, mempertimbangkan algoritma deep learning seperti LSTM atau

BERT sebagai pembanding, mengembangkan model menjadi sistem moderasi komentar otomatis yang terintegrasi dengan API resmi TikTok, serta meningkatkan preprocessing dengan normalisasi kata tidak baku (slang) dan penanganan emoji yang lebih komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika, Universitas Muhammadiyah Sukabumi, serta kepada dosen pembimbing atas bimbingan yang diberikan selama proses penelitian dan penyusunan naskah ini. Ucapan terima kasih juga disampaikan kepada keluarga dan seluruh pihak yang telah memberikan dukungan hingga penelitian ini dapat diselesaikan.

DAFTAR PUSTAKA

- [1] We Are Social and Hootsuite, "Digital 2024: Global Overview Report," 2024. [Online]. Available: <https://wearesocial.com/>
- [2] A. Wahyuningtyas, I. S. Sitanggang, and H. Khotimah, "Deteksi Spam pada Twitter Menggunakan Algoritme Naïve Bayes," *Jurnal Ilmu Komputer dan Agri-Informatika*, vol. 7, no. 1, pp. 31-40, 2020.
- [3] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, pp. 427-434, 2018.
- [4] M. Edwin, A. Chairul, and R. Dewi, "Memahami Spam Terhadap Digitalisasi Masyarakat di Media Sosial," *Jurnal Pengabdian Masyarakat*, vol. 2, no. 1, 2023.
- [5] S. A. P. Perdana, T. B. Aji, and R. Ferdiana, "Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 3, 2021.
- [6] B. Hakim, "Analisa Sentimen Data Text Preprocessing pada Data Mining dengan Menggunakan Machine Learning," *Journal of Business and Audit Information Systems*, vol. 4, no. 2, pp. 16-22, 2021.
- [7] R. Tengku, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia di Twitter Menggunakan K-Fold Cross Validation dan Naïve Bayes Classifier," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178-185, 2022.
- [8] R. Fitriansyah, "Pemanfaatan Vector Space Model Algoritma Nazief Andriani Pembobotan TF-IDF pada Prototipe Klasifikasi Teks Bahasa Indonesia," *Jurnal Teknologi Informasi (JUTECH)*, vol. 2, no. 1, pp. 37-47, 2021.
- [9] S. Hidayat, H. Napitupulu, and N. Gusriani, "Penerapan Model Support Vector Machine pada Kasus Klasifikasi Teks Berdasarkan Tujuan SDGs Ke Tiga, Empat, dan Enam," *SISINFO: Jurnal Sistem Informasi dan Informatika*, vol. 6, no. 2, pp. 115-124, 2024.
- [10] M. K. Suryadewiansyah and T. E. E. Tju, "Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert," *Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)*, vol. 8, no. 2, pp. 81-88, 2022.
- [11] Y. A. Singgalen, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, pp. 237-248, 2023.
- [12] M. A. Sofyan, N. Rahaningsih, and R. D. Dana, "Deteksi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, 2024.
- [13] R. Lutfiyani and N. Retnowati, "Implementasi Pendeteksian Spam Email Menggunakan Metode Text Mining dengan Algoritma Naïve Bayes dan Decision Tree J48," *Jurnal Komputer dan Informatika*, vol. 9, no. 2, pp. 244-252, 2021.
- [14] N. Q. Fitriyah, H. Oktavianto, and H., "Deteksi Spam Pada Email Berbasis Fitur Konten Menggunakan Naïve Bayes," *JUSTINDO (Jurnal Sistem & Teknologi Informasi Indonesia)*, vol. 5, no. 1, pp. 1-7, 2020.
- [15] M. H. S. Ajat, "Klasifikasi SMS Spam dengan Komparasi Metode SVM dan Naïve Bayes," *Method: Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 1, pp. 31-34, 2023.
- [16] L. Ode et al., "Deteksi Komentar Spam Perjudian Online di YouTube Menggunakan K-Nearest Neighbors dan Naïve Bayes," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 2026.
- [17] M. G. Andriawan and T. Ernawati, "Penggunaan Algoritma Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen Konflik Palestina dan Israel pada

Platform X," Jurnal Informatika dan Teknik
Elektro Terapan, vol. 12, no. 3, 2024, doi:
10.23960/jitet.v12i3.4943.