

# IMPLEMENTASI KLASIFIKASI KONTEN NEGATIF WEBSITE BERBASIS TEKS DENGAN NAIVE BAYES, K-NEAREST NEIGHBOR, DAN SUPPORT VECTOR MACHINE

Reihan Hanafi Wiyanatra<sup>1</sup>, Riska Amelia<sup>2</sup>

Program Studi Teknologi Rekayasa Komputer Jaringan, Fakultas Teknologi Industri,  
Universitas Bung Hatta<sup>1,2</sup>

Email : [reihandevml@gmail.com](mailto:reihandevml@gmail.com), [riskaamelia@bunghatta.ac.id](mailto:riskaamelia@bunghatta.ac.id)

---

## ABSTRAK

*Perkembangan teknologi informasi memudahkan akses informasi tetapi juga meningkatkan penyebaran konten negatif seperti pornografi, perjudian, dan phishing. Penelitian ini mengembangkan model klasifikasi konten negatif berbasis teks dengan algoritma Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN). Menggunakan metode kuantitatif eksperimental dan dataset dari GitHub, data diproses melalui web scraping dan pre-processing, serta fitur diekstraksi dengan TF-IDF. Model SVM menunjukkan akurasi tertinggi 82.24%, dengan sistem berhasil mengklasifikasikan konten negatif hingga 80%. Penelitian ini berkontribusi pada keamanan jaringan dengan model klasifikasi yang efektif.*

**Kata Kunci :** *Konten Negatif, Machine Learning, Web Scraping, Training, Confusion Matrix.*

## 1. PENDAHULUAN

Teknologi informasi berkembang pesat, terutama di Indonesia, yang membawa kemudahan dalam komunikasi dan akses informasi secara global. Meskipun teknologi ini menawarkan banyak manfaat, seperti mempermudah interaksi antar individu dan akses informasi di berbagai perangkat, ada dampak negatif yang signifikan. Salah satunya adalah penyebaran konten negatif, seperti pornografi, perjudian, dan phishing, yang dapat menimbulkan masalah serius seperti gangguan emosional, kehilangan kepercayaan diri, dan risiko menjadi korban kejahatan siber. Konten negatif ini juga dapat merusak kepercayaan masyarakat pada informasi daring dan menciptakan lingkungan yang tidak aman bagi pengguna.

Berdasarkan data dari Kementerian Komunikasi dan Informatika, Indonesia menerima 16.370 aduan terkait konten negatif per Maret 2022, dengan judi

menjadi pengaduan terbanyak. Penelitian ini bertujuan untuk mengatasi masalah ini dengan menggunakan beberapa metode machine learning, yaitu Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN). Dataset yang digunakan diambil dari repositori GitHub dan melibatkan proses web scraping dan pre-processing untuk menyiapkan data. Dengan mengklasifikasikan konten dalam kategori seperti gambling, pornografi, phishing, dan whitelist, penelitian ini berusaha mengidentifikasi dan mengelola konten negatif secara lebih efektif.

Penelitian ini bertujuan menciptakan model klasifikasi konten negatif berbasis teks yang efektif, mengingat pentingnya mengurangi dampak negatif dan risiko keamanan di internet. Penerapan machine learning terbukti efektif dalam penelitian sebelumnya, dengan akurasi tinggi untuk mengidentifikasi situs web berbahaya. Penelitian ini diharapkan dapat

berkontribusi pada keamanan dan kenyamanan pengguna internet.

## 2. METODOLOGI PENELITIAN

Penelitian ini menerapkan metode kuantitatif eksperimental dengan tiga tahap utama: perancangan model, alat, dan sistem, meliputi identifikasi masalah, studi literatur, dan penggunaan algoritma SVM, Naive Bayes, serta KNN. Sistem yang dirancang akan diuji dan dievaluasi menggunakan 2911 data dari repositori GitHub UT1 oleh olbat, yang berisi lebih dari 5 juta link blacklist dan whitelist.

## 3. HASIL DAN PEMBAHASAN

Proses training dan testing menggunakan dataset yang telah diekstraksi fitur dilakukan dengan model machine learning utama—SVM, Naive Bayes, dan K-Nearest Neighbors—dan hasilnya akan ditampilkan pada Tabel 1.1

Tabel 1.1 Hasil Persentase Akurasi Training Pada Setiap Model

Model	Akurasi Rata-Rata
GaussianNB	68.40%
MultinomialNB	70.59%
KNeighborsClassifier	30.02%
OneVsRestClassifier	81.07%
SVC	82.24%

Hasil training menunjukkan model Support Vector Machine (SVC) memiliki akurasi tertinggi 82.24%, diikuti oleh OneVsRestClassifier dengan 81.07%, dan Multinomial Naive Bayes dengan 70.59%. Gaussian Naive Bayes mencatat akurasi 68.40%, sementara K-Nearest Neighbors (KNN) memiliki akurasi terendah 30.02%. Performa model dapat bervariasi tergantung dataset dan pengaturan parameter.

## 4. KESIMPULAN

Berdasarkan analisis dan pengujian, sistem deteksi konten negatif berbasis teks dengan machine learning berhasil mengidentifikasi konten gambling, pornografi, phishing, dan whitelist secara

real-time. Model Support Vector Machine (SVC) mencatat akurasi tertinggi 82.24%, menjadikannya pilihan terbaik. Kesulitan terbesar terletak pada web scraping dan pelatihan model. Pengujian sistem menunjukkan keberhasilan 80% dalam klasifikasi konten negatif, menandakan efektivitas model dalam meningkatkan keamanan jaringan.

## 5. DAFTAR PUSTAKA

- [1] L. Yana Siregar, M. Irwan Padli Nasution Prodi Manajemen, and U. Negeri Islam Sumatera Utara, "Development of Information Technology on Increasing Business Online," *J. Ilm. Manaj. dan Bisnis*, vol. 2, no. 1, pp. 71–75, 2020, [Online]. Available: <https://doi.org/10.30606/hjimbhttp://journal.upp.ac.id/index.php/Hirarki>
- [2] R. R. Saputri and A. H. Yunial, "OKTAL: Jurnal Ilmu Komputer dan Science Implementasi *Machine learning* Pada Sistem Pendeteksi URL Bermuatan Konten Negatif Menggunakan Metode Algoritma *Naive Bayes* Dan *Support Vector Machine*," vol. 2, no. 11, pp. 3057–3062, 2023.
- [3] J. Laaksonen and E. Oja, "Classification with learning *K-Nearest Neighbors* s," in *Proceedings of International Conference on Neural Networks (ICNN'96)*, IEEE, pp. 1480–1483. doi: 10.1109/ICNN.1996.549118.
- [4] K. Kumar Mohbey and M. Acharya, "Python for *Machine learning*," *Basics Python Program. A Quick Guid. Beginners*, pp. 237–250, 2023, doi: 10.2174/9789815179637123010013
- [5] D. Sarkar, R. Bali, and T. Sharma, "The Python *Machine learning* Ecosystem," in *Practical Machine learning with Python*, Berkeley, CA: Apress, 2018, pp. 67–118. doi: 10.1007/978-1-4842-3207-1\_2.