Perbandingan Model Naive Bayes, SVM, dan KNN dalam Klasifikasi SMS Spam Menggunakan CountVectorizer

Nabila Ismiyati Mubarokah
Teknik Informatika, UIN Sunan Gunung Djati Bandung

Article Info

Article history:

Received month 05 July, 2025
Revised month 07 July, 2025
Accepted month 09, July, 2025

Keywords:

SMS Spam
Naïve Bayes
Support Vector Mechine
K-Nearest Neighbors
CountVectorizer

ABSTRACT (10 PT)

Pesan singkat (SMS) masih menjadi salah satu media komunikasi yang umum digunakan, namun kerentanannya terhadap penyalahgunaan untuk penyebaran spam menjadi tantangan serius. Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi teks—Multinomial Naive Bayes (NB), Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN)—dalam mendeteksi SMS spam menggunakan representasi fitur dari CountVectorizer. Dataset yang digunakan adalah SMS Spam Collection yang terdiri dari 5.572 pesan berlabel "spam" dan "ham".

Proses klasifikasi dilakukan dengan membagi data menjadi data latih dan data uji (80:20), dilanjutkan dengan pelatihan model dan evaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki akurasi tertinggi sebesar 99.19%, diikuti oleh SVM (98.65%) dan KNN (92.56%). Model NB juga mencatat precision sebesar 100% dan F1-score sebesar 0.97 pada kelas spam, yang menunjukkan performa tinggi dalam mengidentifikasi pesan spam dengan akurat dan efisien.

Dengan hasil tersebut, dapat disimpulkan bahwa Naive Bayes merupakan algoritma yang paling efektif dan efisien untuk klasifikasi SMS spam berbasis CountVectorizer. Penelitian ini diharapkan dapat menjadi acuan bagi pengembangan sistem deteksi spam berbasis machine learning yang ringan dan akurat, terutama pada platform komunikasi berbasis teks pendek.

Corresponding Author:

Aldy Rialdy Atmaja, MT.
Jurusan Teknik Informatika, UIN Sunan Gunung Djati Bandung
Email: aldyrialdy@uinsgd.ac.id

1. PENDAHULUAN

Pesan singkat (SMS) tetap menjadi salah satu media komunikasi utama di era digital, namun juga sering digunakan sebagai sarana penyebaran spam yang dapat merugikan pengguna melalui penipuan atau gangguan (spam) [1]. Oleh karena itu, deteksi dan klasifikasi SMS spam menjadi topik penelitian yang penting dan relevan dalam bidang Natural Language Processing (NLP) dan Machine Learning.

Berbagai pendekatan telah digunakan untuk tugas ini—khususnya algoritma Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN). Sharma et al. (2024) menunjukkan bahwa Multinomial Naive Bayes mampu mencapai akurasi hingga 98.65 % pada dataset SMS (UCI dan Bangla), mengungguli metode lain seperti KNN dan Random Forest [2]. Studi oleh Kumar et al. (2023) juga menemukan SVM unggul saat dibandingkan dengan Naive Bayes dan KNN pada berbagai dataset SMS, dengan SVM konsisten tampil terbaik [3]. Selain itu, Prasad & Christy (2025) menyimpulkan bahwa Naive Bayes multinomial mengungguli KNN dengan performa rata-rata 97.8 % dalam pendeteksian spam [4].

Lebih lanjut, Ahmadi et al. (2025) meneliti kombinasi algoritma dan metode ekstraksi fitur, dan melaporkan bahwa SVM serta Naive Bayes yang dipasangkan dengan TF-IDF melebihi bag-of-words dalam hal akurasi (~96–94 %), sementara KNN menunjukkan performa lebih rendah [5]. Disamping itu, Wang (2023)

dalam tinjauan komprehensif tentang filter spam menegaskan bahwa ketiga algoritma tersebut—Naive Bayes, KNN, dan SVM—masing-masing memiliki kekuatan dan kelemahan, tergantung pada skenario penerapan [6].

Metode transformasi teks menjadi representasi fitur numerik sangat penting dalam proses klasifikasi. Salah satu teknik populer adalah CountVectorizer, yang mengubah korpus teks ke dalam bentuk matriks frekuensi kata. Contohnya, Orenday (2023) berhasil mencapai akurasi ~98 % dengan F1-score ~0.93 dalam dataset SMS menggunakan Multinomial Naive Bayes dan CountVectorizer [7].

Dengan demikian, perbandingan antara Naive Bayes, SVM, dan KNN pada klasifikasi SMS spam menggunakan fitur dari CountVectorizer sangat relevan dan diperlukan. Tujuan artikel ini adalah mengevaluasi dan membandingkan ketiga metode tersebut untuk menentukan mana yang paling efektif dalam mendeteksi SMS spam berdasarkan data yang telah diolah dengan CountVectorizer..

2. METODE

Metode dalam penelitian ini mencakup enam tahap utama: pengumpulan data, praproses, representasi fitur, pelatihan model, evaluasi performa, dan visualisasi hasil. Penelitian ini bertujuan membandingkan performa tiga algoritma klasifikasi teks—Multinomial Naive Bayes (NB), Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN)—untuk mendeteksi spam dalam pesan singkat (SMS) menggunakan representasi CountVectorizer.

2.1 Pengumpulan Data

Dataset yang digunakan adalah *SMS Spam Collection Dataset* yang terdiri dari 5.574 pesan SMS dalam bahasa Inggris, dengan dua kategori: “ham” (pesan normal) dan “spam” (pesan sampah). Dataset ini diambil dari repositori GitHub milik Markham dan bersumber dari UCI Machine Learning Repository [5]. Dataset tersebut telah banyak digunakan dalam penelitian sebelumnya sebagai benchmark untuk klasifikasi teks [1].

2.2 Praproses Data

Praproses dilakukan dengan mengonversi label teks ke bentuk numerik. Label “ham” dikonversi menjadi 0 dan “spam” menjadi 1. Selanjutnya, data dibagi menjadi data latih (80%) dan data uji (20%) dengan fungsi train_test_split dari pustaka scikit-learn. Praproses ini memungkinkan data digunakan dalam pipeline pembelajaran mesin berbasis supervised learning [8].

Berbeda dengan beberapa studi yang menerapkan tahapan praproses lanjutan seperti stopword removal, stemming, atau lemmatization, dalam implementasi ini tidak dilakukan manipulasi terhadap isi teks secara eksplisit. Hal ini sejalan dengan beberapa studi yang menunjukkan bahwa untuk algoritma berbasis frekuensi kata seperti Naive Bayes, CountVectorizer cukup mampu menangkap pola distribusi kata tanpa perlu manipulasi linguistik tambahan [2].

2.3 Ekstraksi Fitur dengan CountVectorizer

Teks diubah menjadi representasi numerik menggunakan CountVectorizer, yang menghasilkan matriks dokumen-kata berdasarkan frekuensi kata (bag-of-words). Pendekatan ini dianggap sederhana namun efektif dalam tugas klasifikasi spam SMS, karena mampu menangkap representasi dasar dari distribusi kata dalam teks [4]. Dalam beberapa studi, penggunaan CountVectorizer dipasangkan dengan model seperti Naive Bayes atau SVM menunjukkan performa yang kompetitif dengan metode ekstraksi fitur yang lebih kompleks seperti TF-IDF [9].

2.4 Pelatihan dan Prediksi Model

Tiga model pembelajaran mesin dilatih menggunakan scikit-learn, masing-masing dengan parameter default:

a. Multinomial Naive Bayes (NB):
   Cocok untuk data diskret seperti frekuensi kata. Dalam berbagai studi, model ini menunjukkan akurasi tinggi (>95%) dalam klasifikasi spam [4], [5].
b. Support Vector Machine (SVM):
   Digunakan dengan kernel default (SVC()), yang dalam scikit-learn berarti kernel RBF. SVM dikenal mampu mengatasi data berdimensi tinggi seperti teks [1], [9].
c. K-Nearest Neighbors (KNN):
   Digunakan dengan jumlah tetangga k=5, model ini melakukan klasifikasi berdasarkan mayoritas kelas dari lima tetangga terdekat berdasarkan jarak Euclidean. KNN sering digunakan sebagai pembanding baseline dalam tugas klasifikasi teks [10].

2.5 Evaluasi Performa

Evaluasi dilakukan dengan membandingkan akurasi dari masing-masing model menggunakan metrik accuracy score dari scikit-learn. Selain itu, dilakukan evaluasi lanjutan menggunakan classification report untuk menghitung nilai precision, recall, dan F1-score dari model Naive Bayes, yang menjadi model utama pertama yang diuji [8]. Evaluasi dilakukan hanya satu kali pada data uji, tanpa teknik validasi silang atau pengulangan eksperimen.

2.6 Visualisasi

Akurasi masing-masing model divisualisasikan dalam bentuk grafik batang menggunakan pustaka matplotlib. Visualisasi ini memberikan gambaran komparatif performa antar model terhadap data uji secara intuitif dan mudah dipahami [2].

3. HASIL PENELITIAN

Penelitian ini dilakukan dengan mengimplementasikan tiga algoritma klasifikasi teks—Multinomial Naive Bayes (NB), Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN)—pada dataset SMS Spam Collection. Proses klasifikasi didasarkan pada fitur yang dihasilkan oleh metode CountVectorizer, yang merepresentasikan data teks ke dalam bentuk matriks frekuensi kata.

Setelah dilakukan pelatihan dan pengujian, hasil evaluasi menunjukkan bahwa masing-masing model memiliki performa yang berbeda dalam mengklasifikasikan SMS sebagai spam atau bukan.

3.1 Deskripsi Dataset

Dataset terdiri dari 5.572 pesan SMS, yang masing-masing telah diberi label sebagai spam atau ham (non-spam). Setelah dibagi dengan rasio 80:20, diperoleh sebanyak 4.457 data latih dan 1.115 data uji. Dataset ini merupakan benchmark standar dalam tugas klasifikasi spam

3.2 Akurasi Model

Berikut adalah hasil akurasi dari ketiga model klasifikasi yang diuji:

Model	Akurasi
Naive Bayes (MultinomialNB)	99.19%
Support Vector Machine	98.65%
K-Nearest Neighbors (KNN)	92.56%

Tabel 1.1 Hasil Akurasi Model

Hasil tersebut menunjukkan bahwa Naive Bayes merupakan model dengan akurasi tertinggi dalam mendeteksi SMS spam, disusul oleh SVM dan terakhir KNN. Keunggulan Naive Bayes konsisten dengan berbagai studi terdahulu yang menyatakan bahwa algoritma ini sangat cocok untuk data teks pendek dengan fitur frekuensi kata.

3.3 Laporan Klasifikasi Naive Bayes

Evaluasi lebih mendalam dilakukan pada model Naive Bayes untuk mengukur metrik precision, recall, dan F1-score. Hasil evaluasinya disajikan dalam Tabel 2:

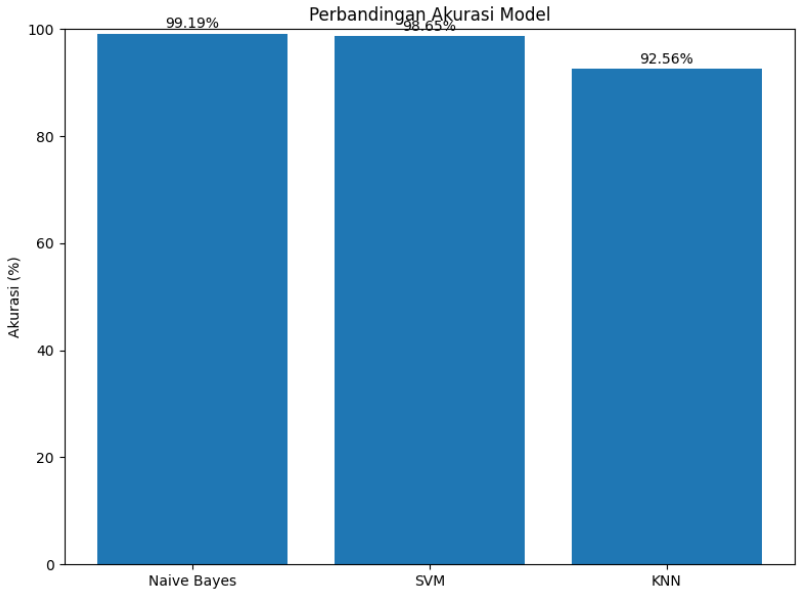
Kelas	Precision	Recall	F1-Score	Support
Ham (0)	0.99	1.00	1.00	966
Spam (1)	1.00	0.94	0.97	149
Macro Avg	1.00	0.97	0.98	1115
Weighted Avg	0.99	0.99	0.99	1115

Tabel 1.2 Evaluasi Naive Bayes

a. Recall pada kelas spam adalah 94%, menunjukkan bahwa sebagian besar pesan spam berhasil dikenali oleh model.
b. Precision pada kelas spam sebesar 100% berarti tidak ada kesalahan klasifikasi spam palsu (false positives).
c. F1-score keseluruhan menunjukkan performa yang sangat seimbang.

3.4 Visualisasi Akurasi

Hasil akurasi dari ketiga model divisualisasikan menggunakan grafik batang berikut:



Gambar 1.1 Visualisasi Akurasi

Visualisasi memperkuat temuan bahwa Naive Bayes memberikan performa terbaik, diikuti SVM dan KNN. Performa KNN yang lebih rendah dapat disebabkan oleh sensitivitasnya terhadap dimensi tinggi dan ketidakefisienannya dalam menangani data teks besar tanpa reduksi fitur.

4. CONCLUSION

Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi teks—Multinomial Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors—dalam mendeteksi spam pada pesan singkat (SMS) menggunakan representasi fitur berbasis CountVectorizer.

Berdasarkan hasil eksperimen terhadap dataset SMSSpamCollection yang terdiri dari 5.572 pesan, diperoleh bahwa Multinomial Naive Bayes (NB) memiliki kinerja terbaik dengan akurasi sebesar 99.19%, diikuti oleh SVM sebesar 98.65%, dan KNN sebesar 92.56%. Selain itu, NB juga mencatat nilai precision 100% dan F1-score 0.97 pada kelas spam, menunjukkan kemampuannya dalam mengenali spam dengan sangat akurat dan minim kesalahan.

Keunggulan NB disebabkan oleh kemampuannya menangani data berdimensi tinggi dan bersifat diskrit seperti representasi frekuensi kata, serta efisiensinya dari segi komputasi. Di sisi lain, KNN menunjukkan performa yang paling rendah dalam eksperimen ini, yang kemungkinan besar disebabkan oleh sensitivitasnya terhadap fitur berdimensi tinggi serta ketergantungannya terhadap distribusi data pada fase prediksi.

Dengan demikian, dapat disimpulkan bahwa Naive Bayes merupakan algoritma yang paling sesuai dan efisien untuk tugas klasifikasi spam SMS berbasis CountVectorizer dalam skenario dataset teks pendek dan seimbang. Hasil ini sejalan dengan sejumlah studi sebelumnya yang menyatakan bahwa NB sangat efektif untuk klasifikasi dokumen berbasis frekuensi kata.

Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi representasi fitur lain seperti TF-IDF, serta mengevaluasi performa model dengan teknik validasi silang (cross-validation) atau pengulangan eksperimen untuk meningkatkan reliabilitas hasil. Selain itu, pendekatan berbasis model pembelajaran mendalam atau transformer juga dapat dipertimbangkan untuk membandingkan efektivitasnya terhadap model klasik.

DAFTAR PUSTAKA

[1] D. A. Oyeyemi and A. K. Ojo, “SMS Spam Detection and Classification to Combat Abuse in Telephone Networks Using Natural Language Processing,” *Journal of Advances in Mathematics and Computer Science*, vol. 38, no. 10, pp. 144–156, Jun. 2024, doi: 10.9734/JAMCS/2023/v38i101832.

[2] S. Kumar and D. Sharma, “A Comparative Study of Machine Learning Classifiers for Different Language Spam SMS Detection: Performance Evaluation and Analysis,” *Advances in Artificial Intelligence Research*, vol. 4, no. 2, pp. 69–77, Dec. 2024, doi: 10.54569/AAIR.1549781.

[3] S. Kumar and S. Gupta, “Legitimate and spam SMS classification employing novel Ensemble feature selection algorithm,” *Multimed Tools Appl*, vol. 83, no. 7, pp. 19897–19927, Feb. 2024, doi: 10.1007/S11042-023-16327-4/METRICS.

[4] J. K. Prasad and S. Christy, “SMS Spam Detection Using Multinational Naive Bayes Algorithm Compared with K-Nearest Neighbor Algorithm,” *AIP Conf Proc*, vol. 3270, no. 1, Apr. 2025, doi: 10.1063/5.0262686/3343800.

[5] M. Ahmadi, M. Khajavi, A. Varmaghani, A. Ala, K. Danesh, and D. Javaheri, “Leveraging Large Language Models for Cybersecurity: Enhancing SMS Spam Detection with Robust and Context-Aware Text Classification,” Feb. 2025, Accessed: Jul. 09, 2025. [Online]. Available: https://arxiv.org/pdf/2502.11014

[6] Y. Wang, “Research on Spam Filters using: SVM, Naïve Bayes, and KNN,” pp. 574–580, Nov. 2023, doi: 10.2991/978-94-6463-300-9_59.

[7] jcorenday, “SMS Spam Classifier using CountVectorizer,” 2023, *https://github.com/jcorenday/sms-spam-classification*.

[8] D. Dharrao, P. Gaikwad, S. V. Gawai, A. M. Bongale, K. Patel, and A. Singh, “Classifying SMS as Spam or Ham: Leveraging NLP and Machine Learning Techniques,” *International Journal of Safety and Security Engineering*, vol. 14, no. 1, pp. 289–296, Feb. 2024, doi: 10.18280/IJSSE.140128.

[9] Y. Li, R. Zhang, W. Rong, and X. Mi, “SpamDam: Towards Privacy-Preserving and Adversary-Resistant SMS Spam Detection,” Apr. 2024, Accessed: Jul. 09, 2025. [Online]. Available: https://arxiv.org/pdf/2404.09481

[10] N. Ghatasheh, I. Altaharwa, and K. Aldebei, “Modified Genetic Algorithm for Feature Selection and Hyper Parameter Optimization: Case of XGBoost in Spam Prediction,” *IEEE Access*, vol. 10, pp. 84365–84383, 2022, doi: 10.1109/ACCESS.2022.3196905.