

IMPLEMENTASI ALGORITMA GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI SPAM *EMAIL*

Yugo Fahreno¹, Deni Arifianto², Miftahur Rahman³

Program Studi Teknik Informatika, Fakultas Teknik
Universitas Muhammadiyah Jember

Email: yugo.fahreno26@gmail.com¹, deni.arifianto@unmuhjember.ac.id²,

miftahurrahman@unmuhjember.ac.id³

ABSTRAK

Spam email menjadi salah satu permasalahan penting dalam teknologi informasi karena dapat mengganggu produktivitas, mencuri informasi sensitif, dan merusak reputasi sistem email. Penelitian ini mengimplementasikan algoritma *Gaussian Naive Bayes* (GNB) untuk mengklasifikasikan email *spam* dan *non-spam (ham)*. Algoritma GNB dipilih karena mampu menangani data kontinu serta memiliki proses pelatihan dan prediksi yang sederhana. Dataset yang digunakan terdiri dari 5.136 email yang terbagi ke dalam dua kategori, yaitu 4.496 email *ham* dan 640 email *spam*. Untuk mengatasi ketidakseimbangan data, diterapkan metode *Random Oversampling* (ROS) dengan menduplikasi data pada kelas minoritas sehingga jumlah dataset meningkat menjadi 8.992 data. Hasil penelitian menunjukkan bahwa model memperoleh tingkat akurasi sebesar 94,30%. Pada kelas *ham*, model menghasilkan *precision* sebesar 100%, *recall* sebesar 87,97%, dan *F1-score* sebesar 93,60%. Sementara itu, pada kelas *spam* diperoleh *precision* sebesar 90,24%, *recall* sebesar 100%, dan *F1-score* sebesar 94,87%. Hasil tersebut menunjukkan bahwa algoritma *Gaussian Naive Bayes* efektif dalam mendeteksi email *spam* dengan tingkat akurasi yang tinggi. Meskipun demikian, masih terdapat kesalahan klasifikasi pada email *ham* sehingga diperlukan pengembangan lebih lanjut untuk meningkatkan performa model, khususnya dalam mengurangi kesalahan klasifikasi pada email *non-spam*.

Kata Kunci: Klasifikasi *Spam* Email, Gaussian Naive Bayes, Machine Learning, Klasifikasi Teks, Deteksi *Spam*, Algoritma Klasifikasi

IMPLEMENTASI ALGORITMA GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI SPAM EMAIL

Yugo Fahreno¹, Deni Arifianto², Miftahur Rahman³

*Informatics Engineering Study Program, Faculty of Engineering,
Muhammadiyah University of Jember*

*Email: yugo.fahreno26@gmail.com¹, deni.arifianto@unmuhjember.ac.id²,
miftahurrahman@unmuhjember.ac.id³*

ABSTRACT

Email spam is a significant problem in information technology because it can disrupt productivity, steal sensitive information, and damage the reputation of email systems. This study implemented the Gaussian Naive Bayes (GNB) algorithm to classify spam and non-spam (ham) emails. The GNB algorithm was chosen because it is capable of handling continuous data and has a simple training and prediction process. The dataset used consists of 5,136 emails divided into two categories: 4,496 ham emails and 640 spam emails. To address data imbalance, the Random Oversampling (ROS) method was applied by duplicating data in the minority class so that the number of datasets increased to 8,992. The results showed that the model achieved an accuracy rate of 94.30%. In the ham class, the model produced a precision of 100%, a recall of 87.97%, and an F1-score of 93.60%. Meanwhile, in the spam class, the precision was 90.24%, a recall of 100%, and an F1-score of 94.87%. These results demonstrate that the Gaussian Naive Bayes algorithm is effective in detecting spam emails with a high degree of accuracy. However, there are still some misclassifications of ham emails, so further development is needed to improve model performance, particularly in reducing misclassifications of non-spam emails.

Keywords: Email Spam Classification, Gaussian Naive Bayes, Machine Learning, Text Classification, Spam Detection, Classification Algorithm