Analisis Perbandingan Algoritma Klasifikasi Naive Bayes dan Support Vector Machine Dalam Efektifitas Pendeteksi Email Spam = Comparative Analysis of The Classification Algorithm Naive Bayes and Support Vector Machine in The Effectiveness of Spam Email Detection

Fikri Afif Musyaffa, author

Abstrak

Spam email merupakan salah satu masalah yang sangat sering dialami dalam komunikasi digital. Penelitian ini bertujuan untuk membandingkan efektifitas dua algoritma klasifikasi Naïve Bayes dan Support Vector Machine (SVM) dalam mendeteksi email spam. Tahapan penelitian dimulai dari pengumpulan data, pemrosesan teks seperti penghapusan angka, tanda baca, dan huruf kapital, penghapusan kata-kata umum, stemming, dan transformasi teks menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Dataset dibagi menjadi dua bagian yaitu data latih dan data uji dengan perbandingan 80% data latih dan 20% data uji. Hyperparameter yang digunakan pada metode Naive Bayes adalah nilai alpha, sedangkan pada SVM adalah nilai C, gamma dan kernel Radial Basis Function (RBF). Evaluasi menggunakan parameter metrik akurasi, presisi, recall, dan F1 score. Hasil penelitian menunjukkan metode SVM dengan hyperparameter tuning dan teks preprocessing mendapatkan nilai akurasi 98,74% sedangkan metode naïve bayes hanya 98,35%. Sehingga dapat disimpulkan bahwa metode Support Vector Machine lebih efektif dibandingkan metode Naïve Bayes dalam mendeteksi email spam.

.....Spam email is one of the most frequently encountered issues in digital communication. This study aims to compare the effectiveness of two classification algorithms, Naïve Bayes and Support Vector Machine (SVM), in detecting spam emails. The research stages begin with data collection, followed by text processing such as removing numbers, punctuation, and capital letters, removing common words, stemming, and text transformation using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The dataset is divided into two parts: training data and testing data, with a ratio of 80% training data and 20% testing data. The hyperparameter used for the Naïve Bayes method is the alpha value, while for SVM, the hyperparameters are the values of C, gamma, and the Radial Basis Function (RBF) kernel. Evaluation is conducted using accuracy, precision, recall, and F1 score metrics. The results show that the SVM method, with hyperparameter tuning and text processing, achieved an accuracy of 98.74%, whereas the Naïve Bayes method only achieved 98.35%. Therefore, it can be concluded that the Support Vector Machine method is more effective than the Naïve Bayes method in detecting spam emails.