

ABSTRAK

Nicholas Kaory (03082200015)

Analisis Perbandingan Metode Naïve Bayes dan Metode Support Vector Machine Terhadap Klasifikasi Kepada Spam Email.

(vii+58 halaman; 14 gambar; 7 tabel; 7 lampiran)

Email merupakan alat komunikasi yang digunakan oleh masyarakat untuk memberikan informasi secara cepat melalui digital pada zaman ini. Walaupun email digunakan untuk memberikan informasi kepada penerima, spam juga terdapat mengirim ke pengguna untuk mencuri informasi atau memberi virus untuk penerima. Oleh karena itu, diperlukan deteksi untuk menyaring email dan spam. Penelitian ini bertujuan untuk membuat aplikasi *Orange3* dengan menggunakan metode SVM dan *Naïve Bayes* dan membandingkan dua metode tersebut untuk mencari metode yang lebih akurat untuk mengklasifikasi email. *Dataset* terdapat dalam *website kaggle* mengandung 5157 data. Proses perhitungan adalah *preprocess data*, transformasi ke TF-IDF, klasifikasi dengan menggunakan metode SVM dan *Naïve Bayes*. Hasil evaluasi dari kedua metode tersebut adalah *Naïve Bayes* memiliki AUC sebesar 0,949, akurasi setinggi 0,969, *F-1 score* dengan 0,970, persisi sebesar 0,971, *recall* dengan nilai 0,971 dan MCC sebesar 0,854 sedangkan hasil pengujian untuk SVM memiliki AUC sebesar 0,950, akurasi setinggi 0,948, *F-1 score* dengan 0,950, persisi sebesar 0,953, *recall* dengan nilai 0,948 dan MCC sebesar 0,760. Hasil menunjukkan bahwa metode *Naïve Bayes* merupakan metode yang lebih akurat untuk mengklasifikasi spam email.

Kata kunci : email, klasifikasi, *Naïve Bayes*, spam, SVM

Referensi : 20 (2019 – 2024)

ABSTRACT

Nicholas Kaory (03082200015)

Comparative Analysis of the Naïve Bayes Method and the Support Vector Machine Method on Classifying Spam Email.

(vii+58 pages; 14 figures; 7 tables; 7 appendixes)

Email is a communication tool used by people to quickly deliver information digitally in this era. Although email is used to convey information to recipients, spam emails are also sent to users to steal information or deliver viruses to recipients. Therefore, detection is needed to filter emails and spam. This study aims to create an application in Orange3 using the SVM and Naïve Bayes methods and to compare these two methods to find the more accurate one for classifying emails. The dataset, available on the Kaggle website, contains 5,157 data points. The calculation process involves data preprocessing, transformation to TF-IDF, and classification using SVM and Naïve Bayes methods. The evaluation results of both methods show that Naïve Bayes has an AUC of 0.949, an accuracy of 0.969, an F-1 score of 0.970, a precision of 0.971, a recall of 0.971, and an MCC of 0.854, whereas the SVM testing results show an AUC of 0.950, an accuracy of 0.948, an F-1 score of 0.950, a precision of 0.953, a recall of 0.948, and an MCC of 0.760. The results indicate that the Naïve Bayes method is more accurate for classifying spam emails.

Keywords : *email, classify, Naïve Bayes, spam, SVM*

Reference : 20 (2019 – 2024)