ABSTRAK

Banyak pengguna yang telah menjadi korban penipuan internet. Dengan adanya bentuk kejahatan seperti email *spam, spoofing,* dan *phishing* menjadikan keaslian dari sebuah informasi didalam *email* menjadi diragukan. Oleh karena itu, penelitian ini menggunakan *email spam* sebagai objek penelitian. Penelitian ini bertujuan untuk mengimplementasikan dan menghitung akurasi dari Algoritma *Support Vector Machine* (SVM) menggunakan kernel *polynomial* untuk mengklasifikasikan *email spam* dengan label *fraud* dan *phishing*. Adapun manfaat dari penelitian ini dapat mempermudah dalam mengklasifikasikan *email spam* dengan lebih akurat. Proses yang dilakukan dalam penelitian ini dimulai dari mengambil data email yang terdapat pada situs *kaggle.com*, proses *pre-processing* teks, pembobotan data atau ekstraksi fitur dengan TF-IDF normalisasi, klasifikasi menggunakan SVM, dan menghitung performa model dengan *confusion matrix*. Berdasarkan pengujian dengan *confusion matrix* menggunakan 80 data *training* dan 80 data *testing* dibagi dengan kondisi skenario 1 sampai dengan 4, didapatkan hasil pada penelitian menunjukkan kernel *polynomial* menggunakan parameter C=1, gamma=scale, degree=2, dengan skenario 1 (60%:40%) menghasilkan akurasi sebesar 78.12 %. Skenario 2 (70%:30%) menghasilkan akurasi sebesar 91.67 %. Skenario 3 (80%:20%) menghasilkan akurasi sebesar 93.75%. Skenario 4 (90%:10%) menghasilkan akurasi sebesar 100%. Hasil ini menunjukan bahwa SVM dapat mengklasifikasikan email spam dengan baik. Akan tetapi, kualitas dan kuantitas dataset serta menentukan parameter sangat berpengaruh pada akurasi data sehingga dapat menyebabkan *underfitting* atau *overfitting*. Data email dengan jumlah besar dapat berpengaruh dalam proses pengujian yang lama.

Kata kunci : Klasifikasi, *Email, Email Spam, Fraud, Phishing, Pre-Processing,* TF-IDF, *Support Vector Machine, Polynomial*.

ABSTRACT

Many people have become victims of internet fraud. With the existence of forms of crime such as spam email, spoofing, and phishing, the authenticity of the information in the email becomes doubtful. Therefore, this research uses spam email as the object of research. This research aims to implement and calculate the accuracy of the Support Vector Machine (SVM) algorithm using a polynomial kernel to classify spam emails with fraud and phishing labels. The benefits of this research can make it easier to classify spam emails more accurately. The process carried out in this study starts from taking email data contained on the kaggle.com site, text pre-processing, data weighting or feature extraction with TF-IDF normalization, classification using SVM, and calculating model performance with a confusion matrix. Based on testing with confusion matrix using 80 data training and 80 data testing divided by scenario conditions 1 to 4, the results showed that the polynomial kernel using parameters C = 1, gamma = scale, degree = 2, with scenario 1 (60%: 40%) resulted in an accuracy of 78.12%. Scenario 2 (70%:30%) produces an accuracy of 91.67%. Scenario 3 (80%:20%) resulted in an accuracy of 93.75%. Scenario 4 (90%:10%) resulted in 100% accuracy. These results show that SVM can classify spam emails well. However, the quality and quantity of the dataset and determining the parameters significantly affect the accuracy of the data so it can cause underfitting or overfitting. Large amounts of email data can affect the long testing process.

Keywords: Klasifikasi, *Email, Email Spam, Fraud, Phishing, Pre-Processing,* TF-IDF, *Support Vector Machine, Polynomial*.