

## **Correlation Impact by Random Forest towards Prediction of Phishing Website**

### **Abstract**

Phishing is an online identity theft that lure unsuspecting victims into phishing website that asks for their personal information such as financial credentials or online access. The most obvious phishing attacks use spam emails to lure victims into visiting its phishing website. The less obvious attack is when a phisher spoofes legitimate sites to bring the victims to the phishing website by presenting a visually similar websites. To address this issue, this paper focuses on the effectiveness of the Random Forest algorithm in predicting a website whether a phishing or legitimate website. The prediction model is developed using the classification methodology and the results revealed that the prediction accuracy of Random Forest is only 66.66%. This is due to potential high correlation among the features when training the dataset since each Random Tree in the forest should protect each other from individual errors. In the future, this project is hoped to investigate features selection of the same dataset because the phishing website prediction model requires features that at least some predictive power.