Machine Learning based Predictive Modelling of Cybersecurity Threats Utilising Behavioural Data

ABSTRACT

With the rapid advancement of technology in Malaysia, the number of cybercrimes is also increasing. To stop the increase in cybercrimes, everyone, including normal citizens, needs to know how secure they are while using digital appliances. A system is developed to predict the risk of users based on their behaviour when they are online using real-life behavioural data obtained from a private university’s 207 undergraduates. Five supervised machine learning methods are being tested which are: Regression Logistics, K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM), and Naïve Bayesian Classifier with the aid of a tool, RapidMiner. The algorithms are used to construct, test, and validate three categories of cybercrime threat (Malware, Social Engineering, and Password Attack) predictive models. It was found that KNN model produces the highest accuracy and lowest classification error for all three categories of cybercrime threat. This system is believed to be crucial in alerting users with details of whether the consumer behaviour risk is high or low and what further actions can be taken to increase awareness. This system aims to prevent the rise in cybercrimes by providing a prediction of their risk levels in cybersecurity to encourage them to be more proactive in cybersecurity.