MIDI 2025 Conference

Enhancing Cybersecurity Risk Assessment through Machine Learning: A Study on Random Forest and XGBoost

Abstract

The modern world web-based system usage is increased, cybersecurity attacks are becoming more frequent, largely due to outdated software components, improper installations, and ignored ventures. Programs like Nikto, OWASP ZAP, and OpenVAS are frequently used for scanning, although they mostly rely on static rules and outdated threat signatures. These results are then in false positives, necessitating a great deal of standard interventions. Better methods for determining the severity of cyberattacks are obviously needed, as they are becoming more sophisticated. This research uses a dataset of 3,000 real-world incidents to present a machine learning-based method for predicting the criticality of cyber-attacks. Random Forest and XGBoost are the two machine learning algorithms that are used in the report for the evaluation process. Different variables like attack type, economic impact, and different protective measures are taken into consideration with 20% of labelled dataset as for identifying serious threats, critical incidents proved to be quite challenging. Among the evaluation of models, Random Forest Classification outperformed with an accuracy score of 71.4% and 0.509 ROC-AUC Score. XG Boost Followed behind, with slightly lower accuracy at 65.3% and 0.500 ROC-AUC Score. However, when it came to detecting rare but crucial cases, the results revealed certain limitations. Precision of 21.5% and recall value dropped to 16.1%, reflecting many critical cases were still to be overlooked. However, the study also highlighted that type of attack and industry affected are standing as most significant contributors.

This research establishes a foundation for incorporating AI models for assessments of security risk indicating single approach might not be sufficient. Stronger detection capabilities may be possible with a combination of approaches, particularly for low-frequency, high-impact threats that are frequently overlooked by conventional techniques.