**Individual In-depth Report**

**Member name: Justin Young**

**Evaluated by: Yeshwanth Reddy Chennur**

**Date: 09/24/2023**

**Tasks Assigned:**

* Literature review of “Machine Learning-based web security intrusion detection system”.

**Summary:**

* This paper aims to compare and identify machine learning based intrusion detection systems against specific attacks (SQL injection and XSS cross site scripting) without sacrificing web server quality.
* The paper begins by providing some research on the chosen attacks, analyzing their principles, classifications and statistical data on frequency of occurrence and keywords.
* A comparative analysis of seven algorithms is conducted on their performance in detection of the SQL based attack
  + AVM, Adaboost, Decision Tree, Random Forest, Logistic Stiff Regression, KNN and Bayesian
* The results of this experiment indicate that all seven algorithms had a perfect rate, but the Adaboost model had the highest accuracy rate of the seven with a recall value of 0.962
* The next experiment is a text disambiguation process on the samples using a defined text classification on the XSS data.
  + Model parameters are optimized for this experiment by reducing the capacity of aspects such as word vector dimensionality, number of words in a given entry, neural network batch size and number of neurons in each model.
* The results of this experiment prove that the accuracy and recall of each model was ~99%, but the SVM model had the highest detection and accuracy rate.
  + However, the paper notes that the SVM model produced a high number of false positives and lacks sufficient learning capability with average data containing large numbers of non-UNK words.
* Conclusions of this research:
  + Deep learning has a capability for generalization than traditional ML models.
  + LSTM had a better detection rate at the expense of accuracy.
  + Anomalous data was considered as XSS by the model
  + Deep learning models require more parameters and training data

**Outcome:**

This paper provides detailed data on the performance of various ML algorithms with XSS and SQL injection attacks, which are considered by OWASP as principle vulnerabilities. The research conducted showed that LSTM (in XSS) and SVM (in SQL injection) performed the best of the algorithms analyzed, however with some drawbacks.

**References** *(with citation)*

C. Chen, J Zhong and W. Chen, “Machine Learning-based web security intrusion detection system”, in 2021 3rd International Academic Exchange Conference on Science and Technology Innovation, Guangzhou, China, pp. 173-177, doi: 10.1109/IAECST54258.2021.9695557

**Evaluation of Report**

**Evaluation summary with justification.**

* In order to maintain web server quality, this paper evaluates machine learning-based intrusion detection systems against SQL injection and XSS threats. For SQL detection, seven methods are tested, with Adaboost attaining the highest accuracy (0.962). Despite high false positives, SVM outperforms for XSS. Deep learning, particularly LSTM, shows promise, but more data and parameters are required.

**The quality of the major result(s) with justification.**

* Adaboost achieves the highest accuracy in SQL detection; despite significant false positives, SVM excels in XSS detection. Deep learning provides better generalization.

**The usefulness of the paper to the overall project.**

* The study gives crucial insights for selecting appropriate intrusion detection methods, as well as improving project efficacy and security measures.

**Other comments**

**Evaluation Approval  
  
Evaluation by:** Yeshwanth Reddy Chennur **Date:** Sep 25, 2023

**Is the written report of the in-depth study complete with all the major result(s) of the paper(s)? If not, provide as many examples of the major result(s) missing in the written report as possible. (in bullet form). [Normally within 100 words]**

* Yes, the in-depth study is complete with all the major results.

**Is each section of the guidelines sufficiently completed? If not, point out what is missing. [Normally within 40 words].**

* Yes, each section of the guidelines are sufficiently completed

**Is the quality of this version of the written report satisfactory? If not, then why not? [Normally within 40 words]**

* Yes, The quality of this report is satisfactory

**Approval.  
  
Approved by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date: 09/25/2023  
  
Is the quality of this written in-depth study report and Evaluation report satisfactory? If not, then why not? (limit: 40 words)**

The quality of this written in-depth study report and Evaluation report is satisfactory. The evaluation report signifies correct evaluation and the report justifies the assigned research domain by providing detailed data on the performance of various ML algorithms with XSS and SQL injection attacks.