

Detecting Security Vulnerabilities using Machine learning models

Team members: Rohit Vishwakarma, Sreekanth R Gunishetty, Shamanth R Rao, Anil M S

Guide Name: Prof. Vinay Joshi

Department Of Computer Science and Engineering, PES University.

Problem Statement

- 1) Our Problem statement is finding the security vulnerabilities and cyber frauds such as fraud transactions, account takeover etc at a large scale that is humanly impossible.
- 2) Our project tries to minimise this by machine learning all the logs and looking for anomalies and reports few suspected logs to the human for inspection.


Background

- 1) Survey of Fraud Detections
- 2) A Novel Web Fraud Detection Technique using Association Rule Mining
- 3) Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models
- 4) A Dynamic Rule Creation Based Anomaly Detection Method for Identifying Security Breaches in Log Records

Results and Discussion

CBLOF Outliers-274 Inliers-27181
 HBOS Outliers-270 Inliers-27185
 KNN Outliers-172 Inliers-27282
 Isolation Forest Outliers-271 Inliers-27184
 Random Forest 98.27%
 Isolation Forest Outliers-550 Inliers-54939

Design Approaches / Methods

- a) Event Driven Architecture
- b) 
- c) Methodology
 - 1) **Breach of Terms of Use**
Cluster Based Local Outlier Factor (CBLOF) Histogram-Based Outlier Detection (HBOS)
K-Nearest Neighbours
Isolation Forest
 - 2) **Self Delivery Fraud** Decision Trees
Random Forest
 - 3) **Credit Card Fraud** Isolation Forest

Dataset and Features / Project Requirements / Product Features

According to the project requirement we need to produce the data from the customised logs of the application backend server in a transaction oriented website **Dataset Fields**
 DATE,IP,TYPE(GET/POST),HTTP CODE,URL_LINK,SHORTNAME_OF_URL,IF_LOGGED,USER_ID,EMAIL,CARDNUM,PRODUCT_ID,MERCHANT,PRICE,QUANTITY,AMOUNT,RATING,ZIPCODE

Summary of Project Outcome

- 1) The First model can be implemented by outlier detection of user activities over a period of time. The model thus trained by this is again trained when it gets a fresh set of data from the website on the next day or another week.
- 2) The Second model detects merchants ordering for themselves and giving high ratings. The plan to detect this is by looking at the deliveries and rating for the different products from the same merchant. All high ratings and same delivery locations raise suspicion.
- 3) The Third model detects card frauds where stolen cards are used for an unusually high number of purchases.

Conclusions and Future Work

Our Goal is to reduce the amount of logs a human needs to go through to detect anomalies in the user accounts. The Website can facilitate to take an input as customised logs and give an output of various detected fraud transactions in the form of csv. As future work, We could look at extending the number of frauds detected in Our Website Analyst.io

References

"Anomaly detection from log files using data mining techniques.", "Credit card fraud detection using machine learning models and collating machine learning models."