A User-Centric Machine Learning Framework for Cyber Security Operations Center

CREATIVE AND INNOVATIVE PROJECT REPORT

Submitted by

G. Santhosh(2016103057)

S. Ben Stewart(2016103513)

P. Udaykumar(2016103622)

College of Engineering, Guindy

ANNA UNIVERSITY: CHENNAI 600 025

OCTOBER 2019

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report "**A User-Centric Machine Learning Framework for Cyber Security Operations Center**" is the bonafide work of "G.Santhosh (2016103057), S.Ben Stewart (2016103513) and P.Udaykumar (2016103622)" who carried out the project work under my supervision.

Dr. AR. Arunarani
SUPERVISOR

Teaching Fellow
Department of
Computer Science and Engineering,
College of Engineering, Guindy Anna University.

Dr. S. VALLI
HOD

Professor and Head
Department of Computer Science and Engineering
College of Engineering, Guindy
Anna University

ACKNOWLEDGEMENT

This project would not have been possible if not for the tireless efforts by the members of this team in all the divisions in which they were asked to work on. Right from the project ideation till the implementation and the documentation and presentation part. We would also thank our mentors and friends who helped us when the project hit a few roadblocks.

ABSTRACT

To assure cyber security of an enterprise, typically SIEM (Security Information and Event Management) system is in place to normalize security events from different preventive technologies and flag alerts. Analysts in the security operations center (SOC) investigate the alerts to decide if it is truly malicious or not. However, generally the number of alerts is overwhelming capacity to handle all alerts. Because of this, potential malicious attacks and compromised hosts may be missed. Machine learning is a viable approach to reduce the false positive rate and improve the productivity of SOC analysts. In this paper, we develop a user-centric machine learning framework for the cyber security operation center in real enterprise environment. We discuss the typical data sources in SOC, their workflow, and how to leverage and process these data sets to build an effective machine learning system. We use the system using the key repository of information regarding the vulnerabilities that allow intruders to breach computer networks is the National Vulnerability Database (NVD). NVD is a product of the U.S. National Institute of Standards and Technology's (NIST) Computer Security Division and is also sponsored by the U.S. Department of Homeland Security's Computer Emergency Readiness Team (US-CERT). We are implementing the below steps from data massaging, label creation, feature engineering, machine learning algorithm selection, model performance evaluations, to risk score generation.

The above implementation would help other teams with only knowledge of machine learning to get a better understanding of the domain of cyber security and the challenges it provides with the requirement of high accuracy models though the dataset is highly biased. It also helps the teams on the other side of the spectrum who are from cyber security to get to understand how machine learning models can be used to the greater benefit.

TABLE OF CONTENTS

LIST OF FIGURES

REFERENCES

1. A user-centric machine learning framework for cyber security operations center https://ieeexplore.ieee.org/document/8004902
2. Downloading and unzipping a .zip file without writing to disk https://stackoverflow.com/questions/5710867/downloading-and-unzipping-a-zip-file-without-writing-to-disk
3. Parsing JSON Dataset with Pandas https://www.geeksforgeeks.org/pandas-parsing-json-dataset
4. Filtering Pandas DataFrames on dates https://stackoverflow.com/questions/22898824/filtering-pandas-dataframes-on-dates
5. Beautiful Soup to parse url to get another urls data https://stackoverflow.com/questions/4462061/beautiful-soup-to-parse-url-to-get-another-urls-data
6. Adding a y-axis label to secondary y-axis in matplotlib https://stackoverflow.com/questions/14762181/adding-a-y-axis-label-to-secondary-y-axis-in-matplotlib
7. Rotate axis text in python matplotlib https://stackoverflow.com/questions/10998621/rotate-axis-text-in-python-matplotlib
8. Become a Machine Learning Engineer https://www.udacity.com/course/machine-learning-engineer-nanodegree--nd009t
9. National Vulnerability DatabaseCVE-2017-5638 Detail https://nvd.nist.gov/vuln/detail/CVE-2017-5638