DUAL DEGREE PROJECT

-Guided by Prof. Gaurav Raina

Categorization of Vulnerability using NLP

-Cyber Security Works

Dibyendu Mandal

Final Year DD

Electrical Engineering

IIT Madras

EE18B108

30th October 2022

Weekly Status Update

This report contains works done on the project and gets updated weekly on what's the progress and goals for the next weeks.

Overview

This Project is part of my Dual Degree Project for the FY 2022-2023. It's been guided and graded by **Prof. Gaurav Raina**. It is being assigned by **Cyber Security Works** and is done under the company as an internship.

The Project is to satisfy the need for an automated framework to identify "Vulnerability type" or the "CWE" from the CVE description of Vulnerability of publicly disclosed official databases such as NVD/MITRE.

Goals

Implement Natural Language Processing (NLP) algorithms on the CVE Description to identify and tag one or more Vulnerability Type(s) as well as predict the most accurate CWE that can be assigned to the CVE.

DUAL DEGREE PROJECT

-Guided by Prof. Gaurav Raina

Schedule and Changes

- Weekly Meeting once every Wednesday from 10:00 a.m. to 10:30 a.m.
- Weekly Sync-up Link: https://meet.google.com/zma-dgdd-zeb
- Leaded by Aviral Verma, Sandeep Challa and Prasanth Bharadhwaaj Sairaj

Last Week's Accomplishments and Project Activities

- Definitions and understanding of important terms such as CWE, CPE, CAPEC, Threat & Risks in Cyber Security are done.
- Relationships between Vulnerability, Threat & Risks or CVE & NVD are noted.
- Went through all the Vulnerability Types e.g., DoS (Denial of Service) or Code Execution etc. and learned basics on how the attacks are done.
- **Data Extraction Site-** https://www.cvedetails.com/. This site holds the information on all the Vulnerabilities reported and stores related CWE and description of attack type. This will be our database and this data will be used as training dataset for our model.
- Learned basic Web-Scraping in order to extract the data in excel file to use it as our training data.

This Week's Planned Project Activities

- Using the HTML file from the Data Extraction Site (CVE details site) using Web-Scraping storing all the vulnerabilities of all year since 1999-2022 based on type of Vulnerability.
- Writing a clear & clean code in python to extract those data one by one from every page and storing all in a file containing CVE ID, CWE ID, CVE Description and assigned Vulnerability Type.

Attachments

- **https://docs.google.com/document/d/1oPYOlVYxwR92znW4vHJg8OL0Vf0iI_aS/edit?usp=sharing&ouid=114650313572869184127&rtpof=true&sd=true**