CSE 494: Artificial Intelligence for Cyber Security: Spring 2019

Lab 4: Vulnerability analysis using NIST and Rapid7 data

Objectives of the lab:

- See where to extract vulnerabilities data – we will use NIST for this purpose
- Examine various attributes of the JSON report that can be used to characterize vulnerabilities
- Obtain the CVE exploit data that lists whether a CVE has a POC exploit available we will use Rapid7 data

The first part of the demo would show how to extract reports on CVEs using the NIST data (20 min.)

1. *CVE report from NIST* – the CVE data is available from NIST publicly through their website https://nvd.nist.gov/vuln/data-feeds which maintains JSON feeds
   a. Parse the report and extract the attributes that can be turned into a feature for CVEs
   b. Understand what a CVSS score is and categorize other attributes.

The second part of the demo would show how to analyze the CVE attributes obtained from the report and clean the data for input to scikit learn (40 min.)

2. *Feature curation (10 mins.)*
   a. Categorize each of the attributes into categorical, ordinal, textual, binary attributes.
   b. Conceptual demo on feature vectorization, normalization, encoding, n-grams(optional in case of textual descriptions), n-grams for hex dump based features.
   c. Convert the features into a feature matrix that can be fed to various data mining algorithms – cases of missing data/dropping variables.

3. Obtain the list of exploits that have POC available as indicated in Rapid7 data (10 mins.)
   a. Analysis of samples based on the classes and their attributes