**Proposal**

This project is one of the T5 Bootcamp requirements it's talking about [CVE (Common Vulnerabilities and Exposures)](about:blank) .

[**CVE (Common Vulnerabilities and Exposures)**](about:blank) is a list of publically disclosed information security vulnerabilities and espousers.CVE was launched in 1999 by the MITRE corporation.

**Data**

###### Each entry represents a vulnerability registered in the CVE database. The dataset is contain 89660 unique values . The goal of this project is analyzes vulnerabilities and then uses the Common Vulnerability Scoring System (CVSS) to evaluate the threat level of a vulnerability. A CVE score is often used for prioritizing the security of vulnerabilities. Also, It will explore the relationship between exploits and products. It will focus on how to cluster those products according to the exploits that affect them. Each product will be represented as a vector of exploit types.

This dataset can be found at Kaggle and contain the following features:

|  |  |  |
| --- | --- | --- |
| Attribute | Data type | Description |
| mod\_date | Date time | The date the entry was last modified |
| pub\_date | Date time | The date the entry was published |
| cvss | float | Common Vulnerability Scoring System (CVSS) score, a measure of the severity of a vulnerability |
| cwe\_code | categorical | Common Weakness Enumeration (CWE) code, identifying the type of weakness |
| cwe\_name | categorical | The name associated with the CWE code |
| summary | String | A text summary of the vulnerability |
| access\_authentication | categorical | {NONE, SINGLE, MULTIPLE} |
| access\_complexity | categorical | {LOW, MEDIUM, HIGH} |
| access\_vector | categorical | {LOCAL, NETWORK, ADJACENT NETWORK} |
| impact\_availability | categorical | {NONE, PARTIAL, COMPLETE} |
| impact\_confidentiality | categorical | {NONE, PARTIAL, COMPLETE} |
| impact\_integrity | categorical | {NONE, PARTIAL, COMPLETE} |

**Tools**

There are tools that will be used to achieve the goal of this study, such as: Numpy, matplotlib, pandas, keras and seaborn for discovering the data and train a model. The work will be done through Jupyter notebook.

**TO DO:**

Explore the data and come up with EDA phases then use a model to fit the data.

**The used features may be increased or changed and the model as well.**