

1 Pager Synopsis

Dibyendu Mandal(EE18B108)

Start of Project:

6th JUNE 2022.

Meetings: Met **Prof. Gaurav Raina** regarding various topics to start with, plans for the couple of days.

Revision of few concepts:

JULY 2022 - SEPTEMBER 2022.

Works done: revision on Probability, Baye's theorem, Data engineering, Python Packages, Data Structure and Algorithms, Basics on R.

Start of Internship:

10th OCTOBER 2022.

Meetings: Meeting with **Prof. Gaurav Raina**, Team lead **Aviral Verma** under company **Cyber Security Works** joined as intern and discussed with the project topic and getting to know each others, setting platforms to work on.

Works done: Topic Selected : **Classification of Vulnerability using NLP(Natural Language Processing).**

Familiarization :

11th OCTOBER-LAST WEEK OCTOBER 2022.

Works done & Meetings: Weekly meeting started on every Wednesday, got familiarize with CVE,CWE and vulnerability and their types, Databases: NVD, cvedetails, etc., little bit of web-scraping

Dataset Readiness :

1st WEEK NOVEMBER 2022.

Works done & Meetings: learned web scraping and with help retrieved 190K data for the use on the project from www.cvedetails.com.

Model Comparison :

TILL LAST WEEK NOVEMBER 2022.

Works done & Meetings: Pre-processing of data using NLP, data visualization, basic feature engineering, comparing performance of supervised models and building the setup for multi-mapping of vulnerabilities with 73% accuracy.

Summarization and improving accuracy :

MID DECEMBER 2022.

Works done & Meetings: Summarize all the works done till that date and discussed with expert on how to improve the model performances. Stratified Sampling, use of n-grams was proposed.

Data Engineering :

LAST WEEK OF JANUARY 2023.

Works done & Meetings: use of Word clouds to understand key words, relatedness, feature importance for different vulnerabilities.

Wordcloud and use of Bigrams :

MID FEBRUARY 2023.

Works done & Meetings: wordcloud from all vulnerabilities analysed and was decided to use of bigrams.

Bigram report :

MID MARCH 2023.

Works done & Meetings: Word counts , Bar graph and Wordcloud of Bigrams for each vulnerability types are reported and best bigrams for each vulnerability type was to be selected by team lead for further purposes.

Analysis on Bigrams :

5th APRIL 2023.

Works done & Meetings: Team lead shared the important bigrams for each vulnerability types and I was supposed to filter them and remove rest others.

Limitation on Hardware(RAM) :

8th APRIL 2023.

Works done & Meetings: All useful bigrams are aggregated manually. Started vectorizing the bigram data but due to high computation ram limit exceeded, which caused crashing of the data generated. Team Lead due to hardware limitation had to conclude the project...