CYBR proposal suggested outline

Our team includes students ____[names]_____, and we want to work on _____[project]_____, which has the machine learning goal of _____.

It is related to cyber network security because ____[sell it]_____.

Our overall deliverable will follow CRISP-DM format -- a paper writeup with sections [list the sections of CRISP-DM]. We anticipate that this project will have significant contributions for the _____ [... and _____] steps of CRISP-DM. [Mention if anticipated large contribution in a particular area. For example, data prep if feature engineering / data wrangling will be important, modeling if test-train split strategies is anticipated to be important, model evaluation if SHAP etc is of interest, deployment if anticipated deliverable that others can use, perhaps to AWS / GCP / heroku etc?].

We will each have our own clearly defined deliverables. We will use GitHub for collaboration and code review etc.

CYBR GROUP SET

CYBR 0: Chidi; Ty; Spencer

CYBR 1 Nick; Kenny, Xiaosong

CYBR 2 Kent (hindom matrix)

CYBR 3: Neil

CYBR 4

CYBR 5 Jordan and Jack (Phishing)

CYBR 6: Brandon

Sort-of replicate CTU-13, predict whether an ip address is a member of a botnet (or any botnet-detection-ML paper that this paper builds on) **| CYBR 3**

5:57

Figure out how to replicate the bread-and-butter algorithm of ***Hybrid Analysis***, the site I showed you that ran against Wannacry (https://www.hybrid-analysis.com/) | **CYBR6, CYBR1**

5:58

Sort-of replicate any of the more traditional ML papers that HinDom builds on -- predict whether a domain name is malicious

Phishing URLs -- detect whether a given URL is phishing. Maybe pull in more information than just the one dataset I linked to, because this one would be relatively small-scale on its own, would just be URL parsing **| CYBR 5**

Sherlock mobile phone dataset -- using phone sensor data, detect when an app is doing something malicious | **CYBR 0**

MY CLASS GROUP SET

MC 1: Chidi; Ty; Spencer
MC 2: Nick; Kenny, Xiaosong
MC 3: Neil
MC 4: Kent (interest in sherlock)
MC 5: Jordan and Jack (Phishing)
MC 6: Dria, Sam, Renee
MC 7: Brandon

Sort-of replicate CTU-13, predict whether an ip address is a member of a botnet (or any botnet-detection-ML paper that this paper builds on) **| MC 3**

5:57

Figure out how to replicate the bread-and-butter algorithm of ***Hybrid Analysis***, the site I showed you that ran against Wannacry (https://www.hybrid-analysis.com/) | **MC7, MC2**

5:58

Sort-of replicate any of the more traditional ML papers that HinDom builds on -- predict whether a domain name is malicious

Phishing URLs -- detect whether a given URL is phishing. Maybe pull in more information than just the one dataset I linked to, because this one would be relatively small-scale on its own, would just be URL parsing **| MC 5**

Sherlock mobile phone dataset -- using phone sensor data, detect when an app is doing something malicious | **MC1**