Ryan Ripper
Prof. Lyon – PPOL 565
4/1/22

Final Project Plan: An Effort to Create a Cybersecurity Risk Evaluation Metric

0. Will you be doing option 1 or option 2 for the final project?

I will be doing option 2 for the final project.

1. What is your plan for your project (brief description)?

I have been working with a team of fellow Georgetown students currently enrolled in the Hacking for Defense class where they have been tasked with creating a standardized evaluation metric to address the cybersecurity risk posed by defense companies working with the government. This metric will work to prioritize mitigation methods and to prevent cybersecurity incidents for the NSA Defense Industrial Base (DIB).

The NSA hopes to alleviate the burden of evaluating over 100,000 DIB companies by identifying those who present the greatest cybersecurity threat. The project plans to scale the services the NSA currently conducts on small-scale cyber security-as-a-service pilots to DIB companies. There is currently no standardized evaluation metric of business cybersecurity risk while little is known about the business risk posed by companies not in the insurance industry.

This project plans to address the complexity of the problem posed by the NSA in identifying companies that would benefit from additional cybersecurity support by first examining a set of contracts awarded by the United States government for hypersonics and then expanding to include other critical technologies that would seemingly pose as financial, political, technical, and/or military targets.

The policy implications related to this project involve the proper identification and allocation of resources from government entities to protect the interests of American development for critical and emerging technologies.

2. What is the goal of the project?

The ultimate goal of the project is to construct a best performing machine learning classifier in creating a model/standardized evaluation metric to predict whether a company is prone to cyber security attacks, dictating a reason to support these companies with government response. We can take the scope of the project a step further by creating a model that predicts the probability of a cyber security attack for a given company.

3. What data will you use for the project?

The project will use the USAspending Application Programming Interface (API) to collect awarded contract information for companies across the United States. In conjunction with the results from the API, I aim to use financial information relating to these companies, specifically market level information to better identify the capacities of companies and media discussion to identify company exposure. I will also use a list of past cybersecurity attacks by foreign entities. These cybersecurity attacks will act as our target feature.

Ryan Ripper
Prof. Lyon – PPOL 565
4/1/22

The data that will pose the most difficult to collect will be the list of past cybersecurity attacks where companies may be hesitant to publicly announce they have been attacked, let alone targeted. Additionally, collecting/scraping the data will be difficult due to the manner in which the data have been published (not in accessible formats). We can expand the list of features to include in our classifier to company level considerations that could be scraped from Wikipedia and other online resources (i.e. size, founding date, associations to other companies, etc.).

4. What steps do you need to make enough progress on the project to be prepared to present it by April 27?

I know the proposed goal is lofty but I aim to at least develop a model that would determine whether or not a company would be a likely target for cybersecurity attacks. In building this model, I will need to collect data from the proposed API and company specific information. I may need to limit the considered scope of companies to those that are only traded on the NYSE to simplify data collection. Therefore, in order to make enough progress on the project to be prepared to present by April 27, I will need to collect an appropriate amount of data to then feed into a machine learning pipeline in developing a risk evaluation metric.