Is there a way to systematically identify a company as being a potential cybersecurity target from online attacks? Cybersecurity is the practice of protecting systems, networks, and programs from digital attacks. Attacks on online platforms usually attempt to access, change, or destroy sensitive information; extort money from users; or interrupt normal business processes. Cybersecurity attacks also have the capacity to suppress certain social and political activity, to steal intellectual property, and to harm regional and international adversaries.

This paper starts with our collaboration with Hacking for Defense (H4D) Spring 2022. H4D is a class offered at Georgetown University sponsored by the Department of Defense that teaches students to work with the Defense and Intelligence Communities to rapidly address the nation’s emerging threats and security challenges. Students work with sponsors to address government concerns in solving specific, mission critical problems while allowing students to investigate their interests in the given field. This semester’s class has 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 National Security Agency (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 target. The H4D project plans to scale the services the NSA currently conducts on a small-scale cyber security-as-a-service pilots to DIB companies. There is currently no standardization evaluation metric of business cybersecurity risk while little is known about the business risk posed by companies not in the insurance industry. Thus, the class's proposal and our research question both plan to address the complexity of the problem posed by the NSA in identifying companies that would benefit from additional cybersecurity support. This support will contribute to critical technology suppliers 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, which have been designated by the Pentagon as instruments that could dictate the outcomes of future conflicts.

In order to answer our primary research question, we examine contracts and the corresponding awards given by the United States government. With these contracts can we understand the significance of a company within the development of an emerging technology. We start our analysis by focusing on hypersonic technologies. This technology has the potential to further destabilize nations with the speedy deployment of nuclear missiles, is uniquely situated as a “lagging” technology with respect to the United States amongst its foreign adversaries, and is a personal interest to the authors with respect to advanced aerospace technologies and applications. These hypersonic awards and all contracts in general are made available by USAspending which is the official open data source of federal spending information. We focus on published subawards as these provide further insight to all developers outside of the main contributors. By evaluating at the subaward scale, we expand our scope of companies involved in hypersonic technology development outside the main developers such as Raytheon, Northrop Grumman, and Lockheed Martin.

We can supplement the USAspending data by including other features to the main dataset. In our attempt to collect additional data to map to our data from USAspending, we have two options: to match on a company's Data Universal Numbering System, or D-U-N-S, number which is a proprietary identifier given by the firm Duns & Bradstreet or on a company's name. Matching on a company's D-U-N-S number relies on using the Duns & Bradstreet Application Programming Interface (API), with which we did not have any success. Matching on a company's name relies on consistent naming conventions across multiple sources, which we cannot accurately trust. Thus, our efforts lay in the use of subaward contract information provided by USAspending. We can also understand our analysis as an effort to stretch the information we can learn from subaward contract data. We evaluate data from USAspending from the calendar year 2021 so that we have the latest full years’ worth of contract information.

Working with H4D, we were tasked as data scientists to data wrangle and develop machine learning models to identify companies that could be cybersecurity targets and to develop a risk evaluation metric as a continuously monitoring tool. Data wrangling provides an initial effort to identify at-risk companies. Machine learning will help us develop a model to continue our identification of at-risk companies and to hopefully predict attacks.

However, without additional information outside USAspending, our effort to create a cybersecurity risk evaluation metric stalls as we do not have enough identifiers to predict a potential attack since we do not have data on past cybersecurity attacks. Without past cybersecurity attacks we cannot create labeled prediction models to anticipate these attacks since we do not have a specific dependent variable with which to train our models. Furthermore, if we were to have a list of past cybersecurity attacks across the companies included in USAspending, there would be a select number of companies that had been victims of attacks. We would have an extreme imbalanced class distribution amongst those companies who had been attacked versus those who had not. We could address this imbalance by oversampling the positive class (those who had been attacked) but would have these observations repeated many times over, resulting in overfitting for some models.

With this obstacle came our decision to pivot our focus to identifying companies that would first be prone to cyber security attacks, dictating a reason to support these companies with government response and then to find any anomalies in the data. Our subsequent results and analysis focus on comparisons between all subaward contracts and hypersonic subaward contracts. We develop unlabeled machine learning models to gain further insight to the USAspending data.

Ultimately, we were able to identify DIB companies that could be cybersecurity targets. These companies exhibited a lacking focus on Information Technology infrastructure and attracted specific attention regarding their accomplishments in hypersonic technology across domestic and international media outlets.

Even though our initial goal of creating a cybersecurity risk evaluation metric was not successful, we identified companies that serve as critical contributors to hypersonic technology development that could act as cybersecurity targets from American adversaries. Our analysis focuses on data wrangling to find specific companies that could favor from additional government support within the digital sphere. Of the companies we identified, the ones that stand out the most with respect to the data analysis, machine learning, and associated media coverage are Optical Sciences Corporation, Bendix Commercial Vehicle Systems LLC, and HySonic Technologies, LLC. Of the companies that are identified and assumed to have a critical role in hypersonic development and should already have significant infrastructure in place to handle cybersecurity attacks, Lockheed Martin Corporation shares slightly under 20% of the entire aerospace market as of the end of the calendar year 2021. Our analysis was successful in both identifying companies that could benefit from additional support and companies that should have cybersecurity measures since they represent large portions of the respective market.

Even though we focused our analysis to hypersonic subaward contracts, our findings show that our approach can be extended to any other emerging technology the United States government has shown interest. Such other technologies include artificial intelligence, quantum information technologies, biotechnologies, and space technologies and systems.

To continue in our initial effort to create such a cybersecurity risk evaluation metric, I aim to meet with NSA officials who sponsored the Hacking for Defense team. I hope to gain additional insight as to new ways to supplement the USAspending data and to explore other venues of analysis. Their insights could contribute to developing a dashboard where the emerging and critical technology could be selected and companies, descriptions, business types, etc. are highlighted as those with increased risks of cybersecurity attacks. Such a dashboard could reduce unnecessary efforts to analyze large amounts of data in favor of a readily available system at the disposal of the United States government and other cybersecurity participants.

I would like to thank Professor Gregory Lyon for his introduction to the Hacking for Defense program and his support throughout this project. I would also like to thank Thomas Keelan who was my partner within the Hacking for Defense program. Thomas helped narrow my focus and motivated my approach to the question at hand.