In this project, we explore the ISCXIDS2012 Cybersecurity Dataset provided by BigBear.Ai. Our goal is to build Machine Learning models to predict whether the incoming network traffic is “malicious” or “normal” to improve network security. We start by exploring the data to gain more insights followed by cleaning the data (converting categoricals to numerical values, dropping features that don’t provide any useful information, etc.). We then use various different classification models and cross-fold validation to see which has better performance. After finding the model that performs the best, we use grid search to optimize.

We plan to follow a similar process to build a model to predict applications used for network communication. If we have time, we will also try to visualize (heat map, or something similar) the network activity.