Project 4 Proposal:
CryptoGuard: Unveiling Ransomware Transactions

Jakob Long, Ethan Colley
COE379L

Introduction:

We have decided to investigate advanced classical algorithms for tabular data. We would like to compare these algorithms to those given in unit 2, and instead of using the dataset from project two we have chosen a dataset featuring Bitcoin information over several years. The ultimate goal is to find out if these advanced methods perform better than the methods we explored in class, and to perform multiple classification methods on this new dataset.

Dataset References:

As of now we will be using the following Bitcoin dataset that investigates specific features on the BitCoin network to identify payments that were likely ransomware.

- **Dataset:** https://archive.ics.uci.edu/dataset/526/bitcoinheistransomwareaddressdataset

The dataset above is from the UC Irvine Machine learning repository. A site that holds many repositories for the purpose of analyzing different machine learning models or algorithms over data that's been cleaned, organized, and scraped for practical and modern use if necessary.

Techniques/Methods Considered:

Some methods/techniques we've anticipated in using for this classification are the following:

- **Logistic Regression**
- **Naive Bayes**
- **KNN**
- **Random Forest**
- **Decision Tree**
- **Support Vector Machines (SVM)**
- **Gradient Boosting Machines: XGBoost or LightGBM**
- **Development of a Neural Networks**
- **TabTransformer**

The list above is quite intensive, and may not be representative of the final product. As some models may become too complex. One such model that may be too complex is the new TabTransformer model that has only been around for a short period of time, as such could pose

some issues. But with the following documentation: https://paperswithcode.com/method/tabtransformer, results are likely possible.

Products to be Delivered:

The products that will be delivered by the end of our project are the following:

- **Jupyter-Notebook:** A notebook containing the construction, testing, and validation of our model(s) that were constructed.
- **Project Writeup:** A pdf file containing the purpose behind our project, overall objectives, methodology & steps taken to complete the project, and the resulting conclusions made post results.
- **Docker Files:** Docker files, including a yml file that will be used to easily build our image hosting our model on a flask service that will allow the user to use http requests to both post data to our model(s), and get information.