Status Report 1

Sam Burgess

Alex Fortescue

Insider Trading Group

May 2020

So far, we have mostly been compiling data on cases of insider trading. We have compiled a few dozen cases with ticker symbols and dates the trading occurred. We have narrowed down the stock data being parsed to windows of about a month during which we know insider trading was occurring. The data gathered so far is from the years 2014 to 2011, but we will continue getting more data from years earlier than this. Most of this data comes from reported cases from the SEC's website. We just have to finish up gathering this data and we can start to build our models. We have started parsing the data into numpy arrays, which we know from lab will convert to Pytorch tensors easily. We have been using the python moduel 'yfinance' to obtain stock data from yahoo finance's API relevant to our cases.

For the next check point we plan to have finished collecting data from the SEC's website and to have started testing out training different models. The model we plan to start with, our baseline model, is taken from "Mining Illegal Insider Trading of Stocks: A Proactive Approach". This model uses an LSTM RNN that predicts stock volume and use these predictions to find anomaly's in the actual stock volume. This is our goal for the next check point, but in the check-in after that, we plan to have started testing new and original models in addition to our baseline with our new data. (Our data is new and different than the article we are using as our baseline). An example of a plan we have is to test for anomalies between different companies' volumes in addition to the volumes of that are predicted by our model.

So far, we have been working together gathering data and Sam did a bit of work on the parsing of the data into numpy arrays.

The main obstacle we have been dealing with are that the data is a bit more difficult to compile than we originally thought it would be. Some of the stocks are no longer active and the old data from when they were active is not saved on yahoo finance. Another obstacle is some of the stocks are international stocks and we want to stick to NYSE stocks.