Insiderbot

Sam Burgess
Alex Fortescue

Insider Trading

➢ Trading based on non-public information

➢ Sometimes Legal

➢ Illegal in the U.S. when:

   ○ Trade made by corporate directors, officers, or stockholders owning more than 10% of a firm's shares

   ○ Profits are made within 6 months of trade

Finance Jargon

➢ Volume

 ○ Amount of some security traded on one day

➢ Press release

 ○ Public announcement of some significant information
 ○ Earnings reports, merger announcement, etc
 ○ Marked by spike in volume

Who Cares?

➢ Generally agreed to be bad for the economy

➢ Securities Exchange Commision (SEC)

➢ FBI

Current Detection Methods

➢ Tip-offs
➢ Obvious sketchy trading (human readable)
➢ SEC quantitative methods are classified
➢ Some publicly available research applying deep learning / nlp techniques
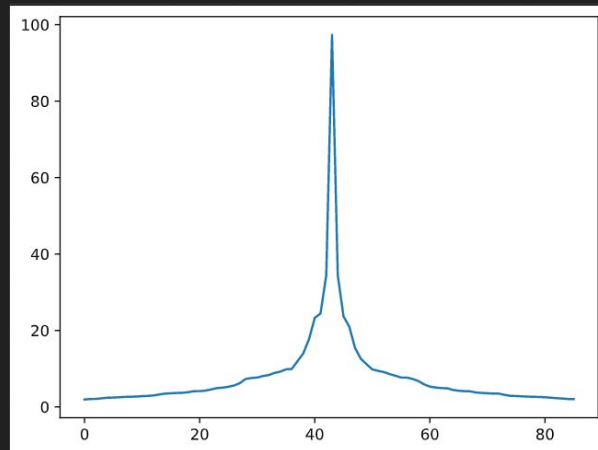
Our Approach

➢ Train a model to be able to classify windows of stock data as insider trading occuring or not

➢ Semi-Supervised Learning

Data

➢ SEC records
➢ Stock Exchange datasets (Kaggle)
➢ Parse into n day windows
  ○ Based on detected volume spikes

Hard Example Mining

➢ Limited number of positive (insider trading is happening) cases
➢ Unlimited number of negative cases
➢ Scan negative data for 'hard' windows that look similar to positive cases

Unsupervised Anomaly Detection

➢ Regression with an LSTM

➢ Sliding Window Approach

Supervised Classifier

➢ LSTM is given positive labelled inputs
  ○ Known to be insider trading cases
➢ Negative inputs hard examples
➢ Tries to learn to classify as positive or negative

Results

➢ Classifier at ~90% accuracy with 'easier' examples

➢ Ongoing work creating 'harder' inputs

Potential for Expansion

➢ More (positive) Data

➢ Better Anomaly Detection

➢ COVID-19 presents rich data

  ○ U.S. Senate