Law + AI

Case Study

December 2, 2019

Project Overview

The U.S. Stock market is the largest in the world and there is an increasing need to protect it from malicious activities. Our goal is to increase the detection capabilities by incorporating AI to have better surveillance coverage so that any suspicious activities can be flagged quickly.

Challenges

There are a few challenges we faced during the project:

1) Difficult to create a supervised ML model - As the number of red flagged activities (abnormal, insider trading etc.) are so few (as compared to the entire trading activities history) that the labeled dataset is highly imbalanced. Even though we use many techniques such as undersampling / oversampling etc. to make it balanced, it is very difficult to train a supervised ML model with good accuracy.
2) Very difficult to prove the wrongdoings - Even we raise the flags from our analysis, it is very difficult to prove if any malpractice really took place unless we leverage other sources of information for further investigation (from Regulators' perspective). For private companies, internal emails history and transaction history may be required for corroboration.
3) Not many company level information available beforehand - Insider Trading moves generally took place many months before the final execution, which are often pretty quick (during press conference, public disclosure, etc).

Dataset

Yahoo Finance : https://finance.yahoo.com/

Solution

Demo 1.0

The Idea is to check if there is any significant volume transaction took place on a single day and also if there is any spike on the share price of the concerned company on the same day, then we highlight (red flag) that day for more scrutiny.

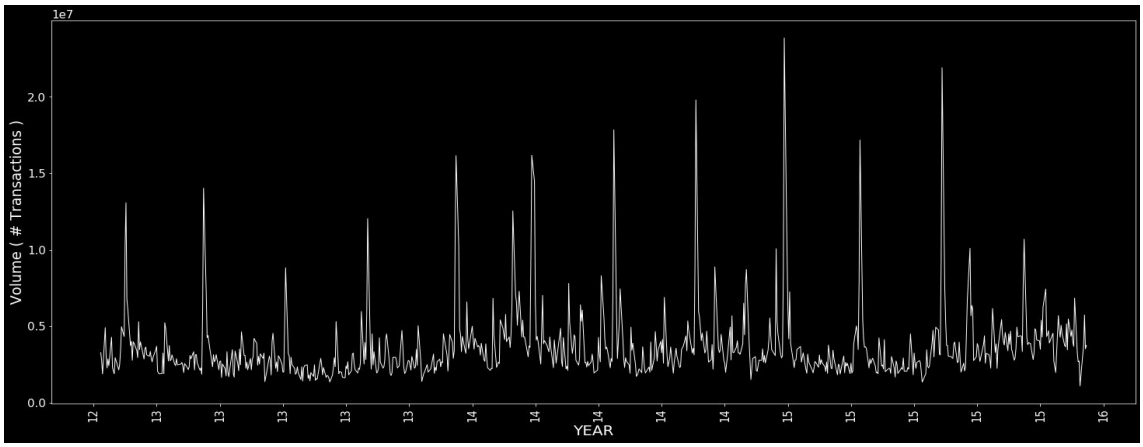

Figure 1: Transaction Volume of Amazon, Inc. from 2013 to 2016

Demo 2.0

We implemented some of the techniques mentioned in the paper *Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation* (https://onlinelibrary.wiley.com/doi/full/10.1111/0022-1082.00265) to detect certain patterns from geometric shapes of trading activities using nonparametric kernel regression. In short, the 8 patterns we identified are the following: Head and Shoulders, Inverse Head and Shoulders, Broadening Top, Broadening Bottom, Triangle Top, Triangle Bottom, Rectangle Top, Rectangle Bottom.

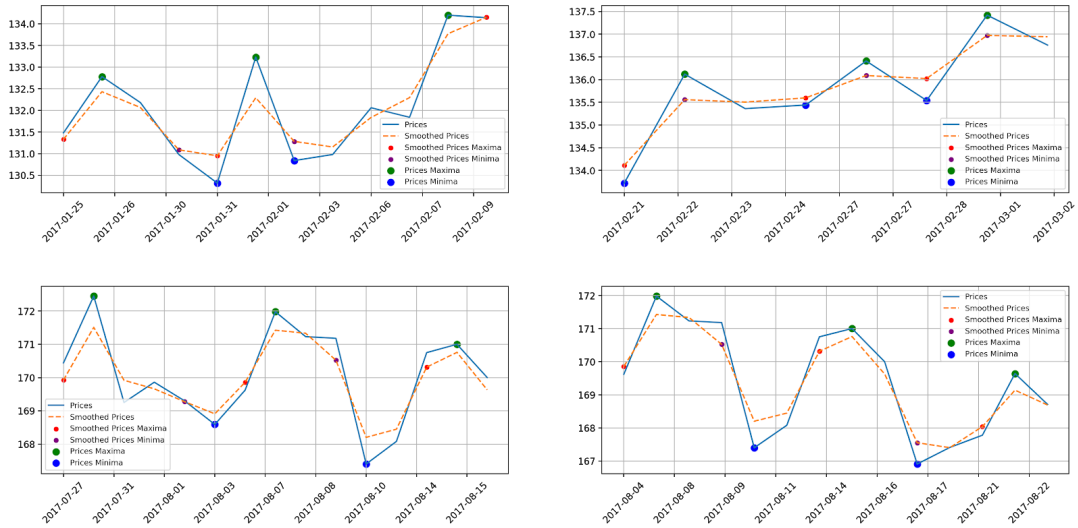

Figure 2: Some of the Rectangle Top Patterns Identified

Both Demos and our slides are open-sourced at https://github.com/Albert-Z-Guo/Law-and-Governance-of-AI

Future Work

1) **Incorporating other Data Pipeline as Metric**

   The idea is to incorporate other features such as organization's ranking on the news/twitter etc. It would help us to corroborate the red flags that we raise using our unsupervised techniques.

2) **Incorporating internal organization metrics**

   If the tool is used internally, the organization can use other internal metrics like internal employees transaction data such as emails, chats, meetings etc. to improve the analysis.