CAPSTONE SPRINT 1

ANOMALY DETECTION OF CONGRESS TRADES

Tiffany Chu

Problem Area and Opportunity

Insider Trading

- US Congress members create policies and can leverage non-public information for financial gain,
- The STOCK act requires them to report their transactions

Opportunity

- There is a need to identify deviations that could suggest insider trading activities and stock manipulation

Parties Impacted:

Retail Investors

Regulatory Bodies

Financial Analysts/ investors

The Proposal

Anomaly Detection using Unsupervised ML Algos:

Clustering

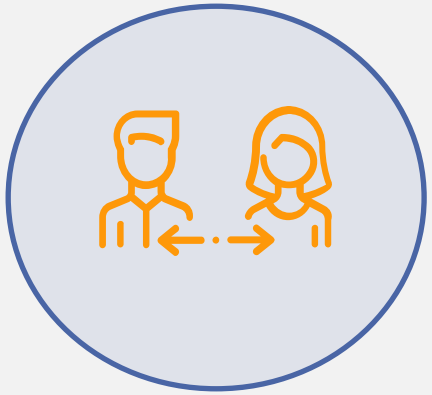
- Find patterns in unlabelled data and groups them into clusters

Isolation Forest

- Recursively splits data into subsets, anomalies can be isolated with fewer partitions

Impact on Stakeholders

1



Transparency

Level the playing ground for all investors, ensures fairness.

2

Safeguard Investments

Helps investors make informed trading decisions and avoid stock manipulation

3

Integrity and Ethics

Upholds the integrity of the stock market, encourages punishment for unethical financial tactics

Data to be Analyzed

Congress Trades

- 01 Transaction Type (buy, sell) and gains
- 02 Transaction Date mm/dd/year
- 03 Industry and Sector
- 04 Stock owner and Congress name

Financial Events (SNES)

- 01 Event date time frame
- 02 Industry and companies affected
- 03 News sources
- 04 Sentiment and Scale of event

Next Steps

Normalize / Scale

This is to ensure that they are on a similar scale of measurement, and it's **especially important when using algorithms sensitive to feature measurements** like our clustering method

Market Index

Shows the **overall performance of a segment of the financial market** (S&P 500, Dow Jones Industrial Average (DJIA), or NASDAQ Comp). They aggregate the performance of highly traded, significant basket of stocks, **providing insights into the broader market trends and sentiment**

Iterative Improvement

Improve the anomaly detection system's performance and effectiveness by **repeatedly refining hyperparameters** every run (ex. Changing depth of trees in isolation forest, or the learning rate in LSTM autoencoders).

Thank you!

Github:
https://github.com/tiffchu

Stock News Dataset:
https://www.kaggle.com/datasets/parsabg/stocknewseventssentiment-snes-10

Congress Trades Dataset:
https://senatestockwatcher.com/api