Project Proposal: Anomaly Detection in Financial Markets Using Deep Learning

Team Name: Pump-and-Dumpbusters/Dumpsters

Team Members: Ronak Agarwal, Kushal Dudipala, Rashmith Repala

Date: February 14, 2025

1. Project Summary

Financial markets can be unpredictable, and oftentimes people take advantage of this. Scams like pump-and-dump schemes and spoofing can manipulate prices, making it harder for regular investors to make informed decisions. Our goal is to build a system that can spot unusual trading patterns and anomalies, potentially flagging fraud-adjacent movements to help us manage our own risk in the market.

We will use deep learning to analyze market data and look for signs of manipulation across different areas like stocks, cryptocurrencies, and forex. The system will recognize patterns that don't follow normal trading behavior and identify when something suspicious is happening.

On top of that, we will also explore whether these anomalies can create investment opportunities. Since prices often return to normal (long-term mean reversion theory) after big swings, we'll look into how spotting these events could help with smarter trading decisions.

2. Motivation

- **Market Manipulation:** Financial markets, especially in crypto, can be like the Wild West. They often lack clear rules/regulations, which make them easy targets for manipulation. Some traders take advantage of this by artificially driving prices up, only to sell off quickly and leave others with the losses.
- **High Volatility Risks & Lack of Robust Detection Models:** On top of manipulation, prices can swing wildly, making it tough for investors to predict what will happen next. A single news event or big trade can cause huge price jumps/crashes. Traditional methods for spotting fraud and manipulation aren't keeping up, and newer machine learning approaches could do a better job at catching these patterns early.
- **Investment Opportunities through Mean Reversion:** At the same time, not all anomalies are bad. Sometimes, a sudden drop or spike is just the market overreacting, and prices eventually return to normal. If we can spot these patterns, there could be opportunities to make smart investments when prices are likely to bounce back.

3. Methodology

a. Data Collection

- To train our system, we need the right data, so we plan to gather:
  - **Crypto price and trading volume** from platforms like Binance, CoinGecko, and Yahoo Finance
  - **Historical records of pump-and-dump schemes** to use as real-world examples.
  - **Social media sentiment** from Twitter and Reddit (optional), since online hype can drive sudden price changes.
- Our model will look at:
  - **Price movements** (Open, High, Low, Close) to track trends.
  - **Trading volume spikes** that could signal unusual activity.
  - **Order book data** (if available) to see how buyers and sellers are behaving.
  - **Social media buzz** to check if online chatter influences market moves.

b. Deep Learning Models

We'll test a few different AI models to detect unusual trading activity:

- **Autoencoders**: Learn what "normal" trading looks like, then flag anything that doesn't make sense.
- **LSTM (Long Short-Term Memory) networks**: Track sequences of price changes to detect suspicious trends.
- **Variational Autoencoders (VAEs) & GANs**: Simulate market conditions to improve how well our system detects anomalies.
- **A hybrid CNN-LSTM model**: Combine different AI techniques to get the best of both worlds by spotting patterns in both time-series data and broader trends.

c. Experimental Setup

We'll first train our models on normal trading data, then test them against known pump-and-dump events to see how well they perform. We'll also check how adaptable the system is by applying it to different markets, like stocks and commodities.

To take things a step further, we'll study whether detected anomalies create good investment opportunities. If prices tend to recover after a sudden drop, the model could help spot potential entry points based on the **mean reversion strategies.**

d. Evaluation Metrics

To measure how well the system works, we'll look at:

- **Precision, Recall, and F1-score** to see how accurately the model flags fraud.
- **AUC-ROC Curve** to check how well it separates normal activity from manipulation.
- **False Positive Rate (FPR)** to ensure it's not over-alerting on harmless trades.
- **Investment performance**: If we use this to make investment decisions, we'll track **ROI (Return on Investment)** and **Sharpe ratio** to measure profitability.

4. Related Work & Background

Individuals have been long trying to detect unusual financial monopolies in financial markets to profit from market inefficiencies and anomalies. Over time, with the rise of deep learning, researchers have begun to explore ways to train models to spot manipulation and fraud. Below are studies that influenced our approach:

1. **Detecting Anomalies in High-Frequency Trading:** The researchers in this study developed a model to analyze order book data, and detect anomalous trading patterns that appear unnatural compared to typical market behavior. They utilized deep learning to uncover these activities in high-speed trading environments.
2. **Identifying Crypto Pump-and-Dump Schemes:** The researchers in this study looked at market trends and social media influences price spikes into cryptocurrencies. They focused on how online discourse shaped trading activity, and were then able to predict when a pump-and-dump event was about to happen.
3. **Using LSTMs for Anomaly Detection in Stock Markets:** The researchers in this study used a Long Short-Term Memory (LSTM) network to detect anomalous movements in stock prices. The network was trained on previous data on market anomalies.

Our project builds on these three previous studies, but for general use. Instead of making the model specific to stocks or cryptocurrencies, we are developing a general tool across financial sectors. We aim to make the model general enough for forex and commodities as well. One additional factor we will incorporate is mean reversion, which is the idea that stock prices will return to normal after periods of unstable change. With the current economic climate, mean reversion seems to be one prominent factor that will define market behavior in the next term. Ultimately, we hope to spot anomalies that highlight investment opportunities.

5. Uniqueness

Most anomaly detection models focus on just one type of market: either stocks, crypto, or forex. Our approach is different because:

- **Generalization for all markets:** The model will be useful for stocks, crypto, forex, and commodities.
- **Social media Sentiment:** In today's world, market moves are driven by hype– especially in the crypto world. We hope to factor in online discussions, and use a multimodal approach to correlate our results. This is a secondary goal dependent on how the original model performs and how much time we have.
- **Not only can we track fraud, but also investment opportunities:** By combining anomaly detection with mean revision strategies, we can identify when prices will shift, and be able to make profitable trades.
- **Unique AI implementation:** Instead of relying on a single deep learning model as in the other papers, we use a hybrid approach combining *CNNs* and *LSTMs*. To detect unusual market patterns we use *CNNs,* and to track time based trends we use *LSTMs*.

6. Benchmark/Dataset/Setup

Dataset:

We will put data from multiple sources. These data sources include crypto exchanges with an abundance of information.

- **Crypto Market Data:** Binance, CoinGecko, and Yahoo Finance will provide historical price and trading volume data.
- **Historical Events:** There have been many infamous pump-and-dumps and flash crashes, such as GameStop, AMC, and many Crypto memecoins. We will compile data from these past situations to help train the model on a more diverse set of cases.

Benchmark:

To test how well our model performs, we'll compare it to what investors traditionally use in their analysis. Below are some of the analysis/metrics they use:

- **Z-score analysis:** This is a fundamental analysis metric to find extreme price movements. If the Z-score is large, it is farther away from the predicted.
- **Bollinger Bands:** This is a metric that shows the support and resistance of a stock, which helps analysts determine if a stock is overvalued or undervalued.
- **MACD (Moving Average Convergence Divergence)**: This is a metric that helps investors identify price trends.
- **Exponential Moving Average (EMA)** – This is a metric that helps identify trends by giving recent price changes more weight. This can be especially important in certain cases where the price drops/increases very fast.

Setup:

- **Development Tools:** The programming language we'll use to drive development is Python. Along with Python, we also utilize machine learning frameworks, like Tensorflow and Pytorch, for model development.
- **Computational Resources:** The main computational resources that'll be used include GPUs or other cloud-based computing platforms, such as Google Colab or PACE-ICE.

7. Risks & Roadblocks

Below are possible risks and roadblocks we may face when creating this project

- **Limited fraud data:** There are not a lot of publicly available datasets for fraud detection, so the model may use semi-supervised or unsupervised techniques.
- **False positives:** The model will need to be able to differentiate normal trading and fraudulent trading accurately. The detection threshold must be fine tuned, and the results must be validated. We hope to employ a training dataset and validation dataset for this case.
- **Data size:** If the tool is to be utilized in live trading, and if it accounts for social media, it will intake a massive amount of data. We must optimize how the model handles and analyzes incoming information.
- **Mean reversion is not always reliable:** Though an indicator for stock price, mean reversion is not 100% accurate. Market trends shift, and anomalies are sometimes random. The model will need to account for that, and thus we may need to adjust strategies as we go.

8. References

[1] B. Ning and K. Lee, "Advanced Statistical Arbitrage with Reinforcement Learning," arXiv preprint arXiv:2403.12180, 2024.

[2] Y. Chang, C. Lizardi, and R. Shah, "Optimizing Returns Using the Hurst Exponent and Q Learning on Momentum and Mean Reversion Strategies," arXiv preprint arXiv:2205.11122, 2022.

[3] S.-H. Moon, Y.-H. Kim, and B.-R. Moon, "Empirical investigation of state-of-the-art mean reversion strategies for equity markets," arXiv preprint arXiv:1909.04327, 2019.

[4] D. Evangelista and Y. Thamsten, "Approximately optimal trade execution strategies under fast mean-reversion," arXiv preprint arXiv:2307.07024, 2023.

[5] "Mean Reversion Trading Strategies," TrendSpider Learning Center. [Online]. Available: https://trendspider.com/learning-center/mean-reversion-trading-strategies/

[6] "Mean Reversion Trading Strategy That Works (86.84% Winning Rate)," TradingwithRayner. [Online]. Available: https://www.tradingwithrayner.com/mean-reversion-trading-strategy/

[7] "Mean Reversion Strategies: Introduction, Trading, Strategies and More," Interactive Brokers. [Online]. Available: https://www.interactivebrokers.com/campus/ibkr-quant-news/mean-reversion-strategies-introduction-trading-strategies-and-more-part-i/

[8] "Mean Reversion Trading Strategy: Your Ultimate Guide," TIOmarkets. [Online]. Available: https://tiomarkets.com/en/article/mean-reversion-trading-strategy-your-ultimate-guide

[9] "What is mean reversion in trading and how do you use it?" FOREX.com. [Online]. Available: https://www.forex.com/en/news-and-analysis/mean-reversion-strategies/

[10] "Mean Reversion Trading Strategies – Backtest With Mean Reverting," Quantified Strategies. [Online]. Available: https://www.quantifiedstrategies.com/mean-reversion-trading-strategy/