

CS5831 Final Project: Preliminary Idea

February 2025

1 Who We Are

The Mining Avengers: Blake Krouth, Ganesh Vannam, and Nandhika Rajmanikandan.

2 What We Want to Do

We hope to explore market regime and trend detection since financial data is abundant, easily accessible, and techniques like Hidden Markov Models (HMMs), clustering, LSTMs, and Transformers have shown success in financial applications. The biggest challenge will be dealing with the noisy nature of financial markets, which requires careful feature selection and hyperparameter tuning. We aim to automate the detection of market conditions (such as bull, bear, or sideways trends). Additionally, we wish to explore differences in supervised and unsupervised techniques, since financial data is not always easily classified into discrete categories.

3 Problem and Goal

Traditional market trend detection relies on lagging indicators like moving averages or the RSI, which can lead to delayed and sometimes inaccurate predictions. Our goal is to develop models that classify market conditions in real time, offering faster and more reliable insights than traditional methods. This will enable traders to adjust their strategies dynamically in response to market shifts.

4 Problem Formulation

Using stock price data, technical indicators (e.g., moving averages, RSI, MACD), and macroeconomic data, we plan to experiment with and compare different classification methods (e.g. SVMs, Decision Trees, Random Forest) and clustering methods (e.g. K-Means, DBSCAN).

The output will be both predefined labels for market phases like bull, bear, or sideways; and unsupervised clusters. The results of both the supervised and unsupervised techniques will be compared.

5 Extended Topics

We would like to explore some other techniques used in finance such as HMMs. They are widely used in finance since they can be used to classify time-series data. In addition, we would like to attempt to extend the clustering methods we will learn in class to apply to time-series data, like stock trends. Time permitting, it would be interesting to also explore transformers.

6 Literature Search

Conferences/Journals:

(<https://www.quora.com/What-is-the-recommended-conference-or-journal-for-publishing-a-paper-in-the-field-of-data-mining-and-machine-learning>)

- Conference on Neural Information Processing Systems (NeurIPS)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- Journal of Machine Learning Research (JMLR)

Papers:

- <https://www.imperial.ac.uk/media/imperial-college/faculty-of-natural-sciences/department-of-mathematics/math-finance/McIndoe.pdf>
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3406068
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4955958

7 Data Resources

We have access to several open-source datasets:

- Yahoo Finance API – Stock market data that can be accessed with Python
<https://pypi.org/project/yahoo-finance/>

- NASDAQ Data link - offers tabular data updated daily that can be accessed with Python
<https://github.com/Nasdaq/data-link-python>
- Federal Reserve Data – Economic indicators
<https://fred.stlouisfed.org/>

8 How Can We Evaluate the Project?

We will have to first gather and filter the market data, likely starting with a limited number of stocks at first. Then, we can train various classifiers and clustering methods. As discussed in class, we can use accuracy & F1-Score for classification models. For clustering methods, we can use metrics like the Silhouette Score or Davies-Bouldin Index. Comparing our results to those from our literature search will give us a good benchmark.