**Examining Stock Market Performance using Machine Learning**

Noah Chasse, Deepak Prabhu and Anuvrat Chaturvedi

**Introduction/Overview**

The goal of this project is to examine stock market portfolio performance when using a variety of machine learning models. The challenge for this project is going to be getting an adequate amount of data, creating meaningful models, and producing reliable and useful results. The limitations we foresee are the ability to put any one method into actual trading and finding meaningful results which can be used to help improve general knowledge of machine learning and the stock market.

**Similar Example**

*A Practical Machine Learning Approach for Dynamic Stock Recommendation* by Hongyang Yang, Xiao-Yang Liu, and Qingwei Wu used historical stock data containing all stocks within the Standard and Poors 500 index from 1990 to 2007. The research paper used linear regression, random forest, ridge regression, stepwise regression, and generalized boosted regression to predict the upcoming quarterly return for a stock and dynamically update stocks chosen for an investment portfolio. The paper chose the top 20 percent of stocks in the S&P 500 and used the minimized mean square error to choose which stocks to include in the portfolio.

Our proposed data will cover from (2000 to 2020) (2010-2020) and cover stocks in the (SP500 or NASDAQ or DOW JONES). In addition, we will consider different regression models including (Linear, GBM, CNN?,LSTM?). Our model will feature a similar approach for stock selection and how we value stocks. We will follow the research papers approach of choosing the top 20 percent of stocks and using the Sharpe ratio to grade which model produces the best results for investing given the risk and variability of the chosen stocks.

Part A - Supervised learning

We propose to gather stock price, public company financial information collected from the SEC (EDGAR) and Yahoo Finance/Google Finance. We will scrape the information from the above sources and create our own dataset. The features we will include are stock price, date, earnings per share, return on assets, return on equity, return on capital employed, net profit, equity ratio, debt to equity ratio, price to earnings, and several other financial ratios. We will employ (linear regression, random forest, and gradient boosting?) and compare their performance on stock prediction. We will use the Sharpe ratio to compare the performance of each model since it will help to account for the return of the stock portfolio while accounting for the risk undertaken in investing the money. For visualizations we will use a bar graph for feature importance and line graphs for stock portfolio performance over time.

Part B - Unsupervised learning

See the above description of the dataset and features. For data manipulation we will need to ensure that stock symbols and data are properly joined from multiple sources. In addition, depending on the type of unsupervised learning we may need to try and generate additional training data. We propose using (LSTM) for making predictions for stock predictions and using principal component analysis to decide which features to include in the models. We plan to incorporate bar graphs for feature importance and line graphs for model performance over time with the chosen stock portfolio.

Team Planning:

The team members specific contributions are outlined below:

|  |  |  |
| --- | --- | --- |
| **Task** | **Due** | **Lead member** |
| Draft proposal | W2 (5/12) |  |
| Data enablement | W3-W4 |  |
| Data & feature engineering | W4 |  |
| Revised proposal | W4 (5/24) |  |
| Supervised learning - model builds | W5-W6 |  |
| Supervised learning - visualizations | W6-W7 |  |
| Unsupervised learning - model builds | W5-W6 |  |
| Unsupervised learning - visualizations | W6-W7 |  |
| Project report | W7-W8 |  |
| Project management | W1-W8 |  |

Detailed Timeline:

Github project link: <https://github.com/users/anuvrat-umich/projects/3/views/4>

Week 1 (5/2-5/9)      - Complete Topic selection and Draft Project Proposal

Week 2 (5/10-5/16)  - Submit Project Proposal by 6/12

   - Research different modeling techniques and different ways to expand projects for Capstone?

Week 3 (5/17-5/23)  - Complete Revised Project Proposal

  - Complete Acquisition of Data and finish Cleaning Data?

Week 4 (5/24-5/30) -

Week 5 (5/31-6/6) - Complete First Slack Standup by 6/3

Week 6 (6/7-6/13) -

Week 7 (6/14-6/20) - Complete Second Slack Standup by 6/15

Week 8 (6/21-6/27) - Complete Final Report due on 6/27