FINAL REPORT: Team 01 - Stock Price Prediction

Team Members:

1. Kartik Khare (MT21034)
2. Shanagh Chakravarty Jayaprakash (MT21047)
3. Ankit Kumar Jha (MT21051)
4. Bibhuti Bhushan Sharma (MT21061)
5. Mihir Chaturvedi (MT21065)

1. Introduction

Asset management is a rapidly growing industry. Significant growth has taken place in recent years. According to a PwC article at least 10% of the worldwide market will be managed by automated robo advisors by 2020. Machine learning and artificial intelligence will be heavily utilized in this industry. Quantitative hedge funds already employ teams of rocket scientists and are using cutting edge algorithms of most sophisticated nature. Renaissance Technologies, a quant hedge fund, is the most successful hedge fund of all time. Its Medallion fund has generated average annual gross returns in the high mid-teens for decades. Even the lower return Institutional fund managed by the firm, which has been approximately 5% in little less than a decade, has shown great outstanding as a compact hedge fund.

Our Problem Statement

For a firm looking to make investments, our solution uses past data from the stock market to predict the future performance of a list of stocks and on the basis of predicted performance, recommend portfolios to the firm to optimize returns from the stock market while keeping the investment risk in check.

Initial Hypothesis

Our initial hypothesis was that technical analysis of stock price trends along with its ratios if it is possible to identify stocks with low volatility and high returns. Based on this assumption, we will predict future stock prices and determine the optimal stock allocation to the user for a balanced portfolio.

2. Literature Survey

Research was conducted on the topic to see what work has been done by earlier or former students or other people in this domain. Following are the key takeaways from the literature review:

- Stock price prediction using LSTM, Machine Learning, and Deep Learning models - Shunrong Shen, Haomiao Jiang, Tongzhang Zhang - 2018
- Stock Price Prediction Using Linear and Polynomial Regression Models - Ayushi Chahal, Nishtha Gupta, Deepti Gupta - 2020
- A Hybrid Stock Trading Framework Integrating Technical Analysis and Machine Learning Techniques - Saeed Sadeghi Reineh, Saeed Sarfjoo, Nasrollah Moghadam Charkari - 2020
- Stock Price Prediction using Support Vector Regression - Vineet Sahu, Neha Mishra, Saurabh Suman Chaudhary - 2020
- Long Short Term Memory Neural Network for Financial Time Series - Oliver Kraus, Matthias Feuerriegel - 2019
- Decision Tree and SVM-based Data Analytics for Predicting Ownership Changes - Ali Tarhan, Ali Vefa Bengi - 2019

Solution Design

The objective of this project is to provide a functionality model for stock price prediction based on both fundamental and technical factors. It achieves this by using machine learning algorithms on a dataset of 5 years of data for the stocks in the NIFTY 50 index. We have used Jupyter Notebook and Google Colaboratory for our implementation. Our solution provides a recommended investment allocation for the performing stocks, analysis which returns and high accuracy in forecast of future stock prices and risk assessment. Our solution is divided into 3 parts, Technical Analysis, Fundamental Analysis and Stock Price Prediction.

1. Data Collection and Cleaning

The data for our project was obtained from the internet in various ways as well as other fundamental factors like earnings, expense, assets and liabilities. Data has technical variable including Open, Close, Adjusted Close, Volume, Dividends and Stock Splits for each stock along the years. Data for fundamental factors consists of various financial ratios such as Price to Earnings, Price to Book, Debt to Equity, Dividend Payout Ratio, Return on Equity, Return on Assets, etc.

2. Clustering Algorithm

Clustering is used to group similar stocks together in a cluster for ease of analysis and making it easier to process stocks. Therefore, it identifies groups of stocks that have similar price attributes. Pearson Correlation Coefficient was determined. K-means Clustering was then performed by grouping similar stocks together. Dimensionality reduction was performed using PCA for better visualization of the clusters.

3. Technical Analysis

Technical Analysis is a method of evaluating securities by analyzing statistics generated by market activity. Therefore, it identifies groups of stocks that have similar price attributes. Pearson Correlation Coefficient was determined. K-means Clustering was then performed by grouping similar stocks together. Dimensionality reduction was performed using PCA for better visualization of the clusters.

4. Fundamental Analysis

Fundamental Analysis is a method of evaluating a security in order to estimate its intrinsic value. It includes qualitative and quantitative analysis. For our analysis, we have used only quantitative analysis. Hypothesis tests were then formed to identify factors affecting Expected Stock Returns. Variable reduction was done using Principal Component Analysis. Multiple Linear Regression was then performed using the reduced variables to predict Expected Stock Returns.

5. Portfolio Recommendation

Our performing stocks with high confidence on predicted value were selected for this step. Different combinations of these top performing stocks were considered to form portfolios. Various portfolio optimization techniques were used to determine the optimal allocation of investment across the selected stocks. Our solution provides a recommended investment allocation for the performing stocks, analysis which returns and high accuracy in forecast of future stock prices and risk assessment.

4. Data Import, Cleaning and Exploration

Our raw Stock Data consists of **9 data sets** extracted from the **Yahoo Stock Market Exchange** in.txt, both, and fundamental details about the companies listed on the NIFTY 50 index. Technical data is for 5 years (2016-2021) for the NIFTY 50 stocks and the NIFTY 50 index. The data has the following features:

Ticker, Open, Close, Adj Close, High, Low, Volume, Dividends, Stock Splits

The data scans listed above has 2 releases. 1. Release 21 The data from 2016 to 2016 is included in the release for companies that among stocks in the NIFTY50. The stocks include the data for the stock exchanges with at least 1000 days. Tags for use for the stocks are included in the release. 2. Release 22 The data from 2016 to 2021 is included in the release for companies that among stocks in the NIFTY50. The stocks include the data for the stock exchanges with at least 2000 days. The data from 2016 to 2016. 22 This point with adjusted dividend and splits is included for change in the price of each stock for the announced stock splits. 23. The dividend payment for each stock is included in the release. 24. The data for the NIFTY 50 index is included in the release. 25. The data for the NIFTY 50 index is included in the release. Tags are included for the stocks. The data for the NIFTY 50 index is included in the release. 26. The data for the NIFTY 50 index is included in the release. 27. The data for the NIFTY 50 index is included in the release. 28. The data for the NIFTY 50 index is included in the release.

DATE	OPEN	HIGH	LOW	PRICE	VOLUME
2019-01-02	12.0000	12.0000	11.8000	11.8000	1.0000
2019-01-03	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-04	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-05	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-06	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-07	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-08	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-09	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-10	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-11	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-12	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-13	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-14	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-15	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-16	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-17	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-18	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-19	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-20	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-21	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-22	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-23	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-24	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-25	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-26	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-27	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-28	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-29	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-30	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-31	12.0000	12.0000	12.0000	12.0000	0.0000

Prediction Model and Performance

Four standard Keras models are used for training predictions:

- Dense Feed Forward network
- Recurrent Predictive Long Short Term Memory

Prices prediction based on returns produces relatively linear Mean Squared error. As Mean squared error increases significantly when prediction window increases, the Long Short Term Memory network is used for long window for all stocks and portfolios.

Key Variables

Fundamental and Technical Plan Prices Stock price data based on both fundamental and Technical are a significant factor while training predictions. Technical indicators such as Simple Moving Average, Exponential Moving Average, Momentum, Relative Strength Index, Stochastic Oscillator, Williams %R, Rate of Change, Average Directional Index, Commodity Channel Index, Ultimate Oscillator, **Trix, Aroon Up, Aroon Down, and Aroon** Oscillator, **Central Jurkat Ball Index**, Vortex, Linear, Log, etc.

Based on Predictive model results for 10 stocks,

Predicted Value	Standard Deviation	Average Value	Percentage Change
Amazon	0.0000	1.8000	0.0000
Facebook	0.0000	0.0000	0.0000
Google	0.0000	0.0000	0.0000
Microsoft	0.0000	0.0000	0.0000
Netflix	0.0000	0.0000	0.0000
Tesla	0.0000	0.0000	0.0000

Using standard regression models Sector prices were predicted for QF-ETF stocks. **Prediction results from Standard Models for mid, or a 30 day window** and long windows show that prices are at record highs as an example.

Sector Index

- XRL American Airlines Group Inc
- LNG Cheniere energy partners
- CCL Carnival Corp & plc
- CHK Chesa energy corp

DATE	OPEN	HIGH	LOW	PRICE	VOLUME
2019-01-02	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-03	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-04	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-05	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-06	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-07	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-08	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-09	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-10	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-11	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-12	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-13	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-14	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-15	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-16	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-17	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-18	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-19	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-20	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-21	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-22	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-23	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-24	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-25	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-26	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-27	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-28	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-29	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-30	12.0000	12.0000	12.0000	12.0000	0.0000
2019-01-31	12.0000	12.0000	12.0000	12.0000	0.0000