Developing algorithmic Trading strategies using Machine Learning Methods

Rajeev Kumar Singh

Research Proposal

February 2020

**Abstract**

There has been an attempt for decades to intelligently and safely operate in Equity markets. Equity Markets are a vital component of market economy. Traditionally various techniques like Mean Variance Optimisation have been tried by portfolio managers for portfolio optimization. However with the advent of Machine Learning and Artificial Intelligence in this field there have been multiple strategies that have become popular.

This research aims at exploring the existing portfolio optimisation techniques and compare them with certain modern techniques like Asset Graphs,ML based portfolio optimisation to start with.Algorithmic Trading algorithms like Deep Q Learning (DQN) and variants of Reinforcement Learning will be tried and models will be developed, Performance of these Algorithms will be compared for metric like sharp Ratio, Annualised Return Percentage etc.

Models will be developed to have a capability to provide as early warning signal with an emphasis on saving the portfolio on Crisis like situations: Financial Credit crisis, COVID-19.

**List of Figures**

Figure 1 Research Methodology Flowchart

Figure 2 Research Plan and Timelines

**List of Tables**

Table 1 Related work list

Table 2 **Data snapshot**

Table 3 Attribute Description

Table 4 Few Features.

**Background**

Equity markets are a popular as well as risky investment hubs for both retail and institutional investor. Certain major Stock exchanges of the have market capitalization in Trillions of Dollars. Given the direct and indirect involvement of masses in equity market they are an important Area of Study.

Investments in the stock markets have been on the rise due to competitive economy. Mutual Funds, which put certain percentage of their investment in stock markets, have become popular with small investor recently. If we compare the data from Statistical Abstract of US from previous years we can see that not only direct ownership but also indirect ownership in form for Retirement accounts rose from 39% in 1992 to 52% in 2007.These suffice to say given these trends stock markets will increasingly become important parts of our financial investment and planning.

The growth in Information and Technology specially digital connectivity have provided platforms in this Era of globalisation where world markets are gradually being integrated in their movement, global citizens have access to all the market and opportunities around the globe.In this vein this area has witnessed keen studies early on, seminal paper of Harry Markowitz is considered by many as a systematic analysis of Portfolio Optimisation.

As per Markowitz, “The process of Selecting a Portfolio may be divided into two stages.The first stage starts with observation and experience and ends with belief about the future performances of available securities. The second stage starts with the relevant belief about future performances and ends with the choice of portfolio”.Now traditional and Modern Portfolio optimisations theories and even some Algorithmic trading models try to solve the above problem.Multiple efforts have been made to come up with accurate prediction of future based on historical data study through Statistical methods and analysis like Mean Variance optimisation. However these efforts are often academic in nature which difficult to reproduce for real future scenarios or too complex to implement.Thus the current development in machine learning and deep learning frameworks provide an opportunity to efficiently model these scenarios and come up with predictive models.The same is being attempted in this research.

Related Research

The field of Portfolio optimization and Algorithmic trading is very dynamic.

Literature review done for the research revolved around exploring the traditional as well as contemporary portfolio optimisation techniques.(Doering et al., 2019) to start with.

Traditional Methods like Mean Variance Optimization were compared with a host of methods from cross functions like Heuristic and Meta Heuristic methods.(Doering et al., 2019).It was found that the drawbacks of traditional methods in terms of solving the problem in real setting as well as their intractability(Perrin, 2019) left the need to come up with newer methods. Optimization algorithms like Coordinate descent, the alternating direction method of Multipliers, the proximal gradient method and the Dykstra’s algorithm were studied.(Perrin, 2019).

Work on Asset Graphs and Trees(J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, 2003) that based on correlation of asset returns was studied this scheme of analysis could also benefit in providing a cue to act as early warning signal to Financial Crisis and minimising losses.

The Literature review with references is detailed below:

|  |  |  |  |
| --- | --- | --- | --- |
| Reference | Problem(s) | Purpose | Algorithm |
| Jana Doering, Renatas Kizys, Angel A. Juan, Àngels Fitó, Onur Polat | Most of real life problem in computational finance are complex and NP hard in nature. These can not be solved as in the given framework and are often solved with simplifying assumptions. | Metaheuristic methods try to overcome drawbacks of traditioanal as well as Heuristic methods. Group of Methods like single solution and Population based are good alternative to Markowitz method. | NA |
| Sarah Perrin,Thierry Roncalli | Poor Control on Academic Portfolio optimisation methods in practical settings in spite of their rigorous theorisation. | Exploration of alternate Portfolio Optimsation methods in Machine learning,Coordinate descent,The alternating Direction method to state a few. | Coordinate descent, the alternating direction method of  multipliers, the proximal gradient method and the Dykstra’s algorithm |
| J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kant | Asset Trees and Assets graphs find correlation between asset returns but both are not suited for all sceanarios.Study probes on the scenarois where each could be suitably used. | Dynamic asset graphs based on Correlation between Asset returns. | Spearman correalation,Pearson Correlation. |
| Ghali Tadlaoui | Compare the traditional trading strategy returns with ML based algorithmic technique. | Develpoing a predictive algorithm using Machine learning to Find Stock Direction ,Model Volatility of returns and compare the returns over time. | Random Forest,Time Series |
| Hongyang Yang, Xiao-Yang Liu, Shan Zhong and Anwar Walid | 1.Inherent complexity and cost of imlementation of traditional strategy involving maximising returns and minimising risk.2.Stratageis like markov decision Process that use Dynamic programming to derive optimal strategy have scalabiltiy issues. | Training a Deep Re-enforcement Learning Agent further obtain an ensemble trading strategy. It has benefit of Adjustablity,Light weight in memory consumption. | Three actor-critic based algorithms: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C), and Deep Deterministic Policy Gradient (DDPG). |
| Li,Yang Zheng,Wanshan Zheng,Zibin | Compare the traditoinal Portfolio allocation strategies with Deep Learning Methods to optimise Sharpe Ratio | Deep Learning Methods for portfolio optimisation to outperform traditional strategies.FourUS Market Indices Total stock index (VTI),aggregate bond index (AGG), commodity index (DBC) ,Volatility Index (VIX). | LSTM with 64 Units |
| Thibaut Theatea, , Damien Ernsta | Zeroing in on Optimal Trading position in time. | Problem Specific Deep Re-inforcement learning Agent. Developed using Artifical Trajectories of Limited stocks. | Deep Q Learning. |

Table 1: Related work list

In the second half of literature survey an attempt was made to understand how the Trading environment in Theorised and modelled through traditional machine learning models. It was found that a mix of Mean Variance optimisation with Feature Engineering is used to create Random Forest and time series forecasting models.(Ghali Tadlaoui, 2018).

Certain advanced techniques in the area of deep learning were analysed. These techniques were more accurate and real in Trading environment reconstruction. They revolved around the use of Re-enforcement learning in general, however three actor critic based ensemble strategy(Li et al., 2019) was found of particular merit. Sequential LSTM model was evaluated to enhance trading predictions(Li et al., 2019).Deep Q Network that aimed at optimal trading position prediction were found value adding to research.(Théate & Ernst, 2020)

**Aims and Objective**

2.5 cm

1.25 cm

The aim of this research is to develop Algorithmic trading strategies using Machine learning frameworks, improving upon the existing methods to generate profit and provide early warning signal to save losses under crisis situations like the Credit Crisis of 2007,Market crash During COVID-19.

The research Objectives arrived in congruence with the aim of research are below:

* Data Collection from Yahoo Finance and manipulating it to arrive at important features for analysis.
* Performing Statistical analysis to come up with selected list of stocks.
* Comparing Traditional and Current Portfolio optimization techniques
* Creating Algo-Trading models in Trading Environment based on Machine learning Algorithms.
* Evaluating the efficacy of Models through popular Testing Strategies and Mertics inclding the scenarois arising of credit crisis,COVID.

**Significance of Study**

Equity Markets are an important component of market economy. It’s a meeting ground for Buyers and sellers of different capacities be it retail whose capacity is few thousand Dollars to institutional players whose capacity is in millions.

The promise of reward at the market also comes with a huge risk of losses. Study of these markets have been more than half a century old. One of the seminal works in this area of Markowitz is appreciated because it tries to balance the risk in process of maximizing reward.

Current research aims at bringing more intelligence in the process through use of traditional as well as advanced Mathematical and Machine Learning models.The Research is aimed at developing models using Random Forest, Time Series, Re-enforcement Learning to bring in better decision for market positions.It will be tested for scenarios that cause loss of life savings and frustration of millions of investors during financial crisis situations.The research will try to explore and develop a scheme that could act as an early warning signal to investors to take safeguarding decision in case of an impending crisis.

2.5 cm

Centered

3 cm

2 cm

**Scope**

The Research will be aimed at analysing and developing models based on stock market data.

Combinations of Index and individual stocks will be tried.

It will tried to create a stock portfolios that will have characteristics like Diversified in terms of sectors, High Growth Stocks Traditional as well as new age stocks.

Index price like Nifty 50 will also be tried for modelling purposes.

Algo-Trading models have varying are significant as they provide consistent returns.

Unlike manual trading strategies they are run by statistical decision making in turn providing high returns, given the high volume of trading data generated, which is impossible to analyse through traditional methods Algo-Trading provides an efficient solution to steady returns.

Given the customisable nature of models it is possible to generate returns on the basis of risk appetite of the portfolio.

**Limitation**

Financial Markets are uncertain in their nature of being, the information that is available for public use is often too high level and aggregated.

Thus the research would work not work on a ticker/Minute level data and provide recommendations that are satisficing(Simon).

7. Research Methodology

Research Steps are detailed below:

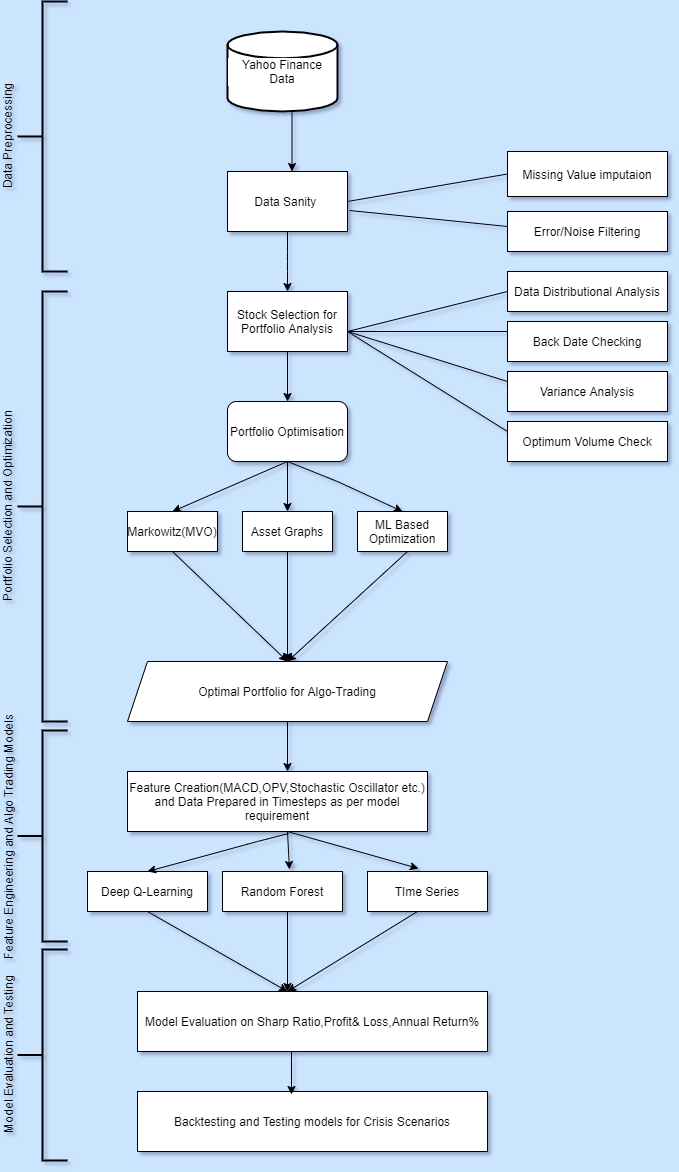


Figure 1: Research Methodology Flowchart

7.1 Introduction

The research is aimed at exploring different portfolio optimization techniques which are traditional like Mean Variance Optimization of Markowitz or the new techniques which have gained acceptance lately.

Using the publicly available Data from yahoo Finance an attempt will be made to compare the portfolio optimization techniques and their resulting returns.

Asset Graphs and Asset tree based portfolio correlation (J.-P. Onnela, A. Chakraborti, K. Kaski, 2002) will be analysed so will be other challenger techniques of optimization like Coordinate descent, the alternating direction method of multipliers, the proximal gradient method and the Dykstra’s algorithm (Perrin, 2019)

Generating an active trading model will be the next part of research wherein feature based prediction model where technical indicators like Open balance Volume, Stochastic Oscillator, Moving Average Convergence Divergence etc will be used in a Random Forest algorithmic setting or Time series(Ghali Tadlaoui, 2018) setting to maximise returns.

In the Later stages Variations of Deep learning Algorithms will be used to create algorithmic models.

This area of research is wide and dynamic thus as per current information and expertise Deep Q Learning (DQN) based models will be generated.

Also the current research on Deep Reinforcement learning will be tried to generate algorithmic trading model.(Théate & Ernst, 2020).

7.2 **Dataset Description**

Data set used for modelling will be Yahoo Finance Data.

Data will look like below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dt. | Open | High | Low | Close | Adj Close | Vol. |
| 1/1/20 | 121 | 122 | 120 | 120 | 120 | 538 |
| 1/1/20 | 120 | 121 | 119 | 119 | 119 | 771 |

Table 2: Data Snapshot

Input Data Attributes:

|  |  |
| --- | --- |
| Attribute | Description |
| Date | Date of Transaction |
| Open | Opening Price of stock at the beginning of Trading Day. |
| High | Highest traded price of the Stock in the day |
| Low | Lowest traded price of the Stock in the day |
| Close | Closing Price of a stock at end of Trading Day. |
| Adj Close | Adjusted close is the closing price after adjustments for all applicable splits and dividend distributions. Data is adjusted using appropriate split and dividend multipliers, adhering to Center for Research in Security Prices (CRSP) standards |
| Volume | These are the Physical number of shares traded on that stock in a Day. |

Table 3: Attribute Description

Past 20 years NIFTY data will be downloaded for Index as well as stock relating to Technology(Infosys,TCS,HCL),Manufacturing(Tata Motors, Ashok Leyland, Mahindra) ,Banking(SBI,ICICI,HDFC)

**Data pre-processing and Feature Engineering**

After evaluating the sample Data it was found that the Data from Yahoo Finance mostly clean. Few Null Values Rows have to be filtered from the Data.

Feature Engineering will be done to derive features like:

|  |  |
| --- | --- |
| Feature | Description |
| On Balance Volume | OBV is a momentum indicator relating the traded volume in the stock market to the price. |
| Stochastic Oscillator %K | %K compares the closing price with a high-low range of the price over a given period of time. |
| Moving Average Convergence Divergence | It is a Momentum and Trend following indicator |

Table 4: Few Features

**Model Building and Model Evaluation**

Host of models that are planned to be built are: Random Forest, Time series Model (GARCH), Deep Q Learning Model other variants of Re-enforcement learning models.

Widely Accepted Metric like sharp Ratio, Profit and Loss, Annualised Return etc. will be used to evaluate the model performance. Model will be back tested to check for robustness, out of sample testing will also be performed. Model will be tested additionally for the credit crisis induced crash of 2008 and Pandemic related crash of 2020.

**Required Resources and Hardware**

* Open source software like R,Python.
* Visualisation tools like Microsoft power BI, Tableau.
* High Computation Power GPU

Research Plan

Research plan is illustrated below in the Gantt chart.

2.5 cm

1.25 cm

2.5 cm

1.25 cm

2.5 cm

1.25 cm

2.5 cm

1.25 cm

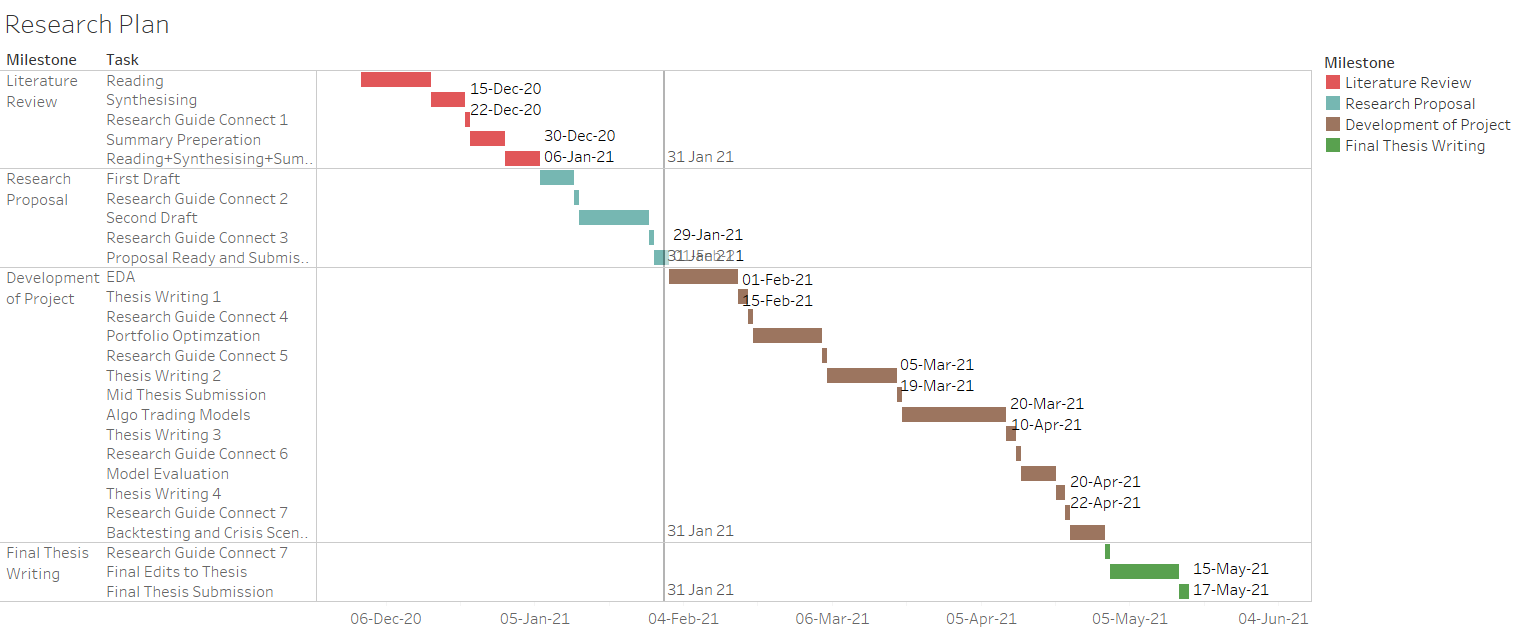


Figure 2 Research Plan and Timelines

References

Doering, J., Kizys, R., Juan, A.A., Fitó, À. & Polat, O. (2019). Metaheuristics for rich portfolio optimisation and risk management : Current state and future trends. *Operations Research Perspectives*. [Online]. 6 (February). p.p. 100121. Available from: https://doi.org/10.1016/j.orp.2019.100121.

Ghali Tadlaoui (2018). *Intelligent Portfolio Construction : Machine-Learning enabled Mean-Variance Optimization*. p.pp. 2017–2018.

J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A.K. (2003). Asset Trees and Asset Graphs in Financial Markets. *Physica Scripta T106*. 48.

J.-P. Onnela, A. Chakraborti, K. Kaski, J.K. (2002). Dynamic asset trees and portfolio analysis. *Eur. Phys. J. B*. 30.

Li, Y., Zheng, W. & Zheng, Z. (2019). Deep Robust Reinforcement Learning for Practical Algorithmic Trading. *IEEE Access*. 7. p.pp. 108014–108021.

Perrin, S. (2019). *Machine Learning Optimization Algorithms & Portfolio Allocation*. p.pp. 1–66.

Théate, T. & Ernst, D. (2020). *An Application of Deep Reinforcement Learning to Algorithmic Trading*. [Online]. Available from: http://arxiv.org/abs/2004.06627.

1.25 cm

2.5 cm

Centered