Equity Portfolio generation and optimization using Machine learning

[1]Manav Mulchandani, [2]Divesh Harchandani ,[3]Vanshika Makhija, [4]Tarun Chawla, [5]Rupali Hande

[1,2,3,4]*Undergraduate student, Department of Computer Science Engineering, Vivekanand Education society's Institute of Technology, Chembur, Maharashtra, India.*
[5]*Assistant Professor, Vivekanand Education Society's Institute of Technology, Chembur, Maharashtra, India.*

Abstract—In today's booming equity markets a well-constructed portfolio is very important to wealth creation. The equity markets in India are generating great returns. The returns generated by equity seem to dwarf the returns generated by all other classes of assets. To achieve high returns while maintaining the risk to minimum value is the goal of any equity investor when investing their capital in the market. A portfolio refers to the grouping of stocks in which the capital among stocks is invested in such a way that the profit is maximum and the risk is minimum. Equity Portfolio generation and optimization is designed to be a highly extensible project applying machine learning to create a portfolio and optimize it that will give higher returns than the major index and is also less risky and diversified.

I. INTRODUCTION

A portfolio is termed as a set of investment assets, is a basket of securities which helps in reducing risk. Portfolio management means the process of investment decision making based on customized investment strategies to get maximum returns for each investing time horizon. There are two mostly used approaches to manage the investment portfolio: traditional and quantitative. These approaches follow some particular common characteristics such as investigating a small set of factors of stock values, analyzing past historical data to estimate these key drivers, adopting criteria for eligibility for stock-selection decisions, and evaluating the performance over time. However, the traditional approach to portfolio management is based heavily on judgment, depth analysis, important key characteristics, and qualitative factors, quantitative portfolio management mainly focuses on exploration of the universe, discipline, verification, risk management, and lower fees.

Our project 'Equity portfolio generation and optimizations' main aim is to provide a well optimized portfolio that gives returns higher than existing indices in a risk efficient way. Our project will help the investor to achieve high returns by analyzing the fundamental aspects of various companies.

The following points will help us understand the need of equity portfolio generation and optimization:

- ***Minimizing risk*** *:* Our software will help to minimize risk by investing in various equities. Our software will help to analyse how individual equities in the portfolio relate or correlate to one another, the main reason behind this is that each and every investor have different risk and reward goals based on their capital invested and the period of investment which can be short term or long term based on the equity preference. For example individuals, companies, trust funds, banks, insurance companies and governments and retail investors.

- ***Simplifying investing:*** Only 1.5% of population of India invests in the equity market main reason being not every individual has the time nor do they have the appropriate knowledge to invest in the equity market.
- Our
- software "equity portfolio generation and optimization" will help to provide the appropriate knowledge by analysing fundamentals of the company so that the individual investors can get proper idea where to invest their capital.

● *Portfolio management process:*

Planning- Planning will help us in understanding the client's needs of investment policy statements.

Execution- This software will help in asset allocation equity analysis and portfolio construction.

Feedback- The feedback from the retail investors and customers will help us in improving monitoring and rebalancing the portfolio it will also help us in performance measurement and reporting.

II. LITERATURE REVIEW

Machine Learning techniques are considered to be much more accurate and fast when they are compared to traditional prediction techniques. It can be observed from the literature survey that the application of machine learning techniques to stock market prediction is being practiced thoroughly throughout the world.

The regression techniques by the author [1], data set that has been used for testing is taken out of the data that was initially collected and the sample size of data used is very less as compared to the original raw data. If the same analysis is done on large datasets, the outcome may be different.

A Data Mining algorithm for predicting changes in the Stock Market was developed[2]. This paper gave us insight whether a stock trend can be predicted using the abnormalities of the historical financial data (in this case, tick data) or not. The data that was taken as input in this was pre-processed to find the anomalies and further a clustering algorithm was applied to predict the stock trends.

In paper[3], the prediction of the stock exchange using SVM was discussed. it's a decent idea to use SVM because it always gives unique results and works well even at local minima. The authors worked on a dataset from the Fed Bank of St. Louis over the information of 15 companies. there's no definitive way to define an honest or poor, so that they had considered it a decent investment if the stock price of a corporation surges over a period of your time. The study demonstrated that SVM produces results with good accuracy for the sample of knowledge which is outside the training sample.

The author has compared the LR model and Support Vector Regression model for data prediction supported by the historical data[4] that they had used LeastMedSq function and SMOReg function for the regression techniques respectively. The metrics for comparison were mean absolute error (MAE) and root mean squared error (RMSE). When these models were applied to the regression data, LeastMedSq function fits more for prediction of the values, although it had been taking longer than SMOReg.

In paper[5],the author used regression modelling to create predictions. The model was redefined anytime the degree of the changes in the problem. The experimentation was performed on Polynomial Regression, RBF Regression, Sigmoid Regression and LR. For prediction purposes, LR proved to be the foremost suited and that they were fitted using the least squares approach.

This problem discussed by the author is commonly considered to be of foremost challenging real-world applications for time-series prediction[6]. Unlike traditional repeating neural networks, LSTM supports time steps of arbitrary sizes and without the problem of vanishing gradients. The author considered both bidirectional and stacked LSTM predictive models in our experiments and also benchmark them with shallow neural networks and straightforward styles of LSTM networks. The evaluations are conducted employing a publicly available dataset for stock exchange closing prices.

III. PROPOSED SYSTEM

In this project we will clean and prepare the dataset of historical stock prices and fundamentals after which we will apply different machine learning algorithms like Regression, LSTM and various classification algorithms to discover the relationship between stock fundamentals such as PE ratio, debt/equity, float, etc and the subsequent annual price change of the stock. After conducting a simple backtest we will then generate predictions on simple data. Portfolio Optimization will help to allocate specific funds in predicted stocks for maximized returns. Optimization will let the investors know how much quantity to add a particular equity in their portfolio so as to achieve high returns by keeping the risk to minimum.

We are providing a web application that mainly focuses on the stock market so to provide the valuable information our website will have three parts,

1. Provide the information and daily news in the stock market world.

2. To predict the best stock for maximum returns using our LSTM model.

3. To optimize the capital or portfolio of an individual for its maximum returns with low risk involved.

Every person does not have time and skills to invest wisely in the equity market but they do want to invest and they don't have the appropriate knowledge or the skills to invest. Our application will help a normal person to get an edge of an investor or an investing firm, our system will provide and be the brain behind the valuable investing of an individual.

IV. METHODOLOGY USED

Our proposed architecture of methodology is based on the typical strategies used for investment

1. Raw data is collected from the internet as historical data is processed and dumped into the main database from the main database, data is segregated into fundamental database, technical database, index database.

2. Input data consist of these above mentioned databases, input data is then fed to the bootstrap dataset that is the part of random classifier algorithm with the help of sampling method and after that various decision trees are created based on the features on which different regression techniques are applied and output is predicted using ensemble learning method. The portfolio was constructed by selecting the outperform stocks from the predicted result in terms of the highest expected return and lowest risk.

3.Stock allocation that was optimized for the constructed portfolio was measured and checked by simulation and optimization modeling. Simulation modeling Monte Carlo simulation (MCS), Equal-weights allocation (EQ), and mean-variance optimization (MVO) were used to evaluate the optimal stock allocation weights giving us the stock portfolio that was optimized. The overview of the methodology architecture is shown in the below figure.
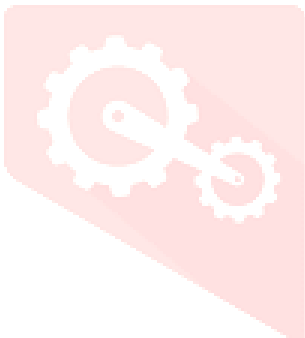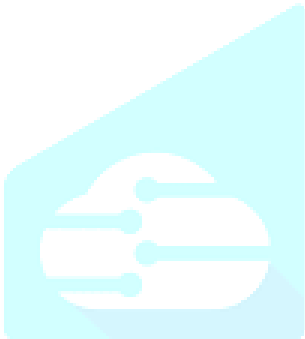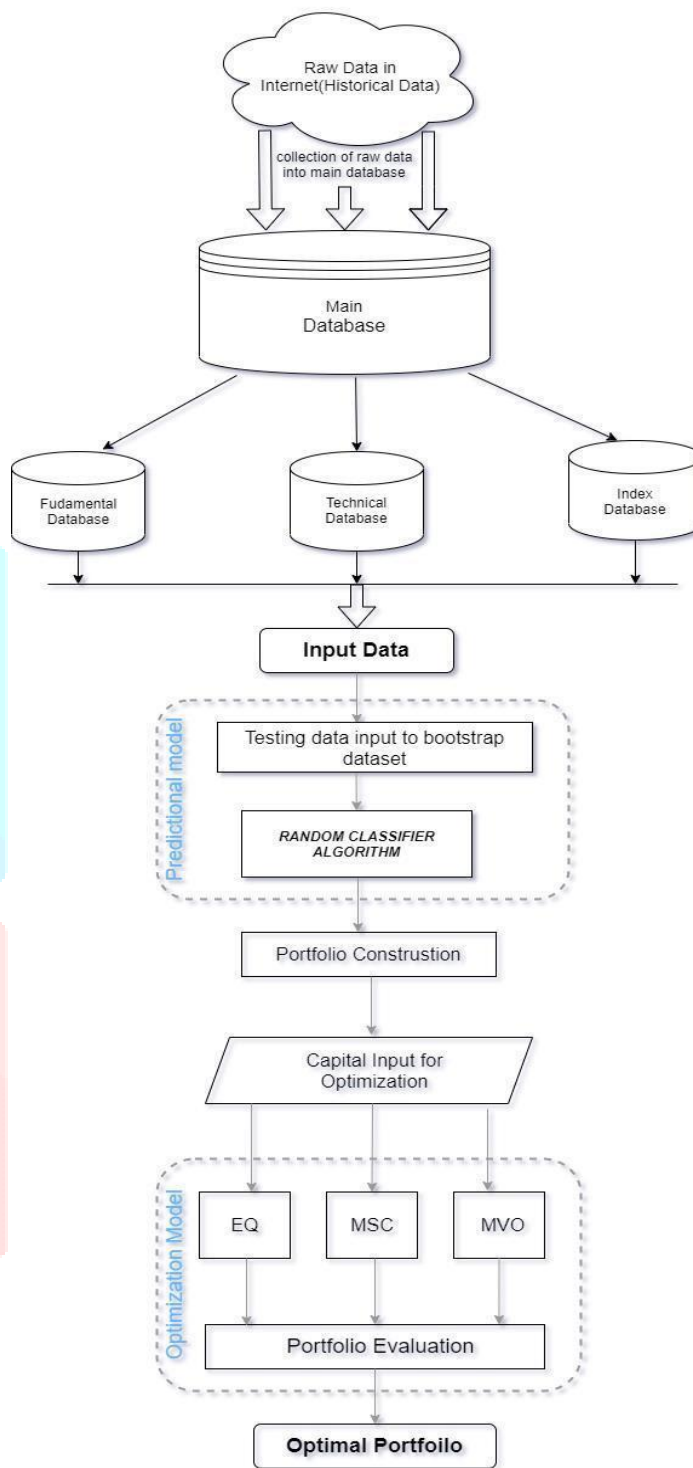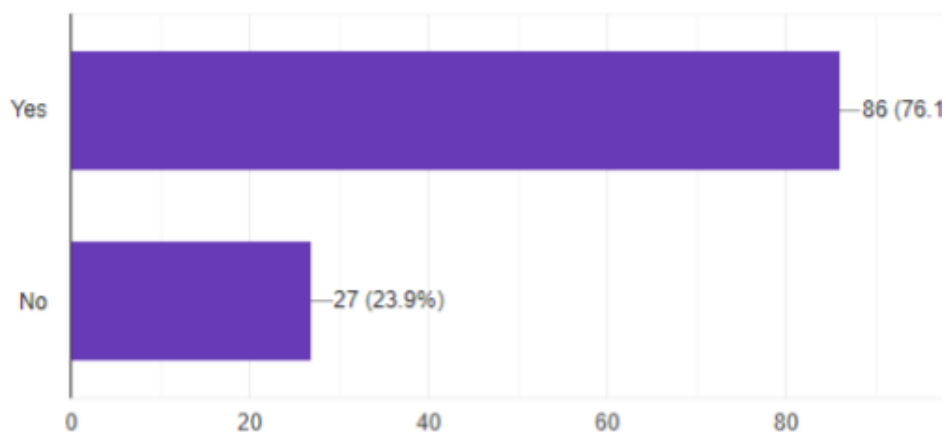
V. PUBLIC SURVEY

1. **Number of people that actively invest in financial market**

Do you invest in financial markets?
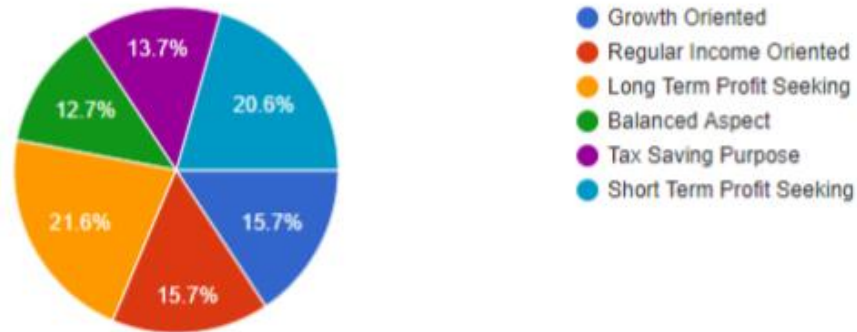
113 responses

From the above figure it can be seen that from 113 respondents, only 76.1% of the respondents invest in the financial market where 23.9% of the respondents do not invest in the financial market.

2. **Following table shows the Objective while investing of the Respondent**

Objective while investing	Percentage of the Respondent
Growth Oriented	15.7
Regular Income Oriented	15.7
Long Term Profit Seeking	21.6
Balanced Aspect	12.7
Tax Saving Purpose	13.7
Short Term Profit Seeking	20.6

Objective while investing?

102 responses

- Growth Oriented
- Regular Income Oriented
- Long Term Profit Seeking
- Balanced Aspect
- Tax Saving Purpose
- Short Term Profit Seeking

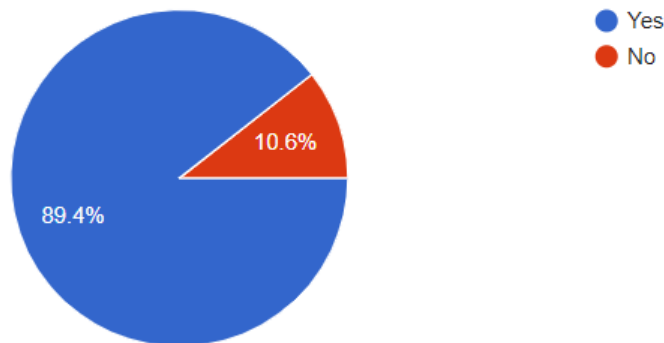**3. Chart to depict the percentage of people that experience trouble selecting equities for their portfolio**

Do you experience trouble selecting equities for your portfolio?

104 responses

- Yes
- No

The above Figure shows that 89.4% of the respondents have trouble selecting equities for their portfolio and the rest of the 10.6% of the respondents can manage on their own.

4. Expectations of people from the 'portfolio generation and optimization application'

What do you expect from a portfolio generation and optimization application?

92 responses
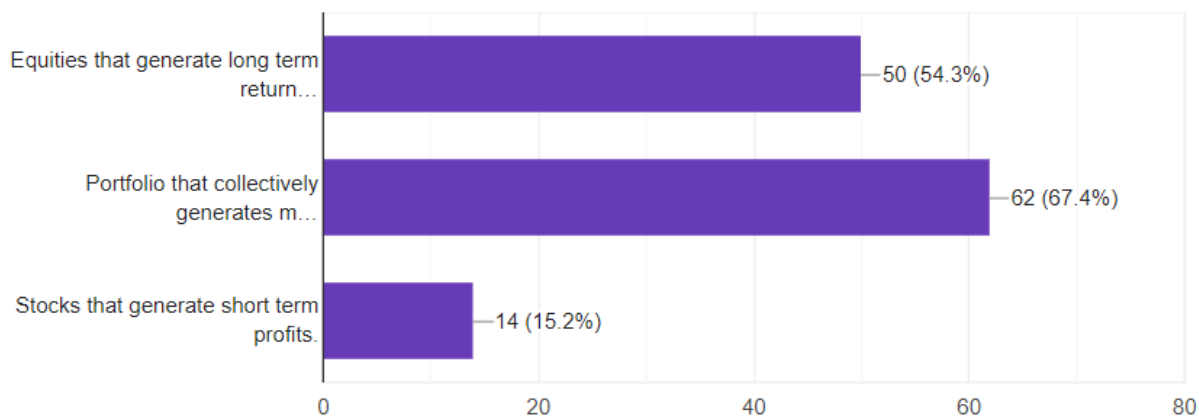
The above figure shows that the 67.4% respondents expect the portfolio generation and operation application to help create a 'Portfolio that collectively generate more returns than the index' whereas 54.3% of the respondents wants 'Equities that generate long term returns and dividends' and only 15.2% of the respondents wants 'Stocks that generate short term profits'.

VI. COMPARATIVE ANALYSIS OF ALGORITHMS

What is the Random forest classifier algorithm? Random forest is a flexible, simple to use machine learning algorithm that generates, without hyperparameter tuning, a good result many of the time. It is also one of the most used algorithms, because of its ease, simplicity and diversity (it can be used for both classification and regression tasks).

- Random forest is a supervised learning algorithm. The "forest" it builds is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. Let's look at random forest in classification, since classification is sometimes considered the building block of machine learning.

- ARIMA model concentrates on the present data and works on present data, accuracy of ARIMA model is more than LSTM but it is not ideal for our project as previous data is the most crucial part for our project.

- CNN or Convolutional Neural Network is known as a class of deep neural networks, it is commonly applied to analyzing visual imagery. Prediction of stock using CNN is a bit different approach CNN is applied to visualize images though it can be used to predict the stock but it will be only dependent on charts and graphs of the index or stock, but in our project we will also consider fundamental aspects of the company i.e management, growth, depts, etc.

VII. CONCLUSION

Machine learning is considered to be a very powerful tool and as necessary as it has some great applications. We have seen till now that machine learning is very much dependent upon the data given as input. Thus it is important to understand that data is quite important and as simple as it may sound, data analysis is not an easy or light task.

Our Project will create a website that will help to develop a strong and optimized portfolio for the investors by using machine learning techniques, it will also keep the viewers of our website updated about the latest financial news and it will also help the investors to provide them information on how to allocate their funds so as to minimize the risk and maximize their returns.

VIII. REFERENCES

[1] D. Bhuriya, G. Kaushal, A. Sharma and U. Singh, "Stock Market Prediction Using A Linear Regression," in International Conference on Electronics, Communication and Aerospace Technology, 2017.

[2] L. Zhao and L. Wang, "Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm," in IEEE Fifth International Conference on Big Data and Cloud Computing, 2015.

[3] Z. Hu, J. Zhu and K. Tse, "Stocks Market Prediction Using Support Vector Machine," in the 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013.

[4] S. Kavitha, S. Varuna and R. Ramya, "A Comparative Analysis on Linear Regression and Support Vector Regression," in Online International Conference on Green Engineering and Technologies (IC-GET), 2016.

[5] A. Sharma, D. Bhuriya and U. Singh, "Survey of Stock Market Prediction Using Machine Learning Approach," in International Conference on Electronics, Communication and Aerospace Technology, 2017.

[6] K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 151-156.

[7] M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.

[8] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSBE), Langkawi, 2012, pp. 1- 5.

[9] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[10] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.