JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Investment Portfolio Management System Using Machine Learning

[1]Arbaz Attar, [2]Shubham Narale, [3]Pranay Mule, [4]Piyush Kulkarni, [5]Prof. Ms. Jaitee Bankar

[1,2,3,4]Students, [5]Assitant Professor, Department of Information Technology, RMD Sinhgad School of Engineering, Savitribai Phule Pune University

Abstract : Portfolio management is the concept of determining the proportions of various assets to be held in a portfolio in order to maximize return while minimizing risk exposure. Portfolio optimization is a critical component in financial management and investment banking. Constructing an optimal portfolio by selecting the best possible combinations of different portfolios is a computationally difficult problem with exponential complexity. It is widely assumed that public sentiment is related to financial markets. Different types of machine learning models have recently been used with positive results for short-term financial market prediction. However, historical returns do not appear to follow the normal distribution hypothesis. However, when using long short-term memory networks in addition to historical data, sentiment analysis yields better results. In this project, we want to use AI/ML to predict portfolio risk and provide insights into how stocks will perform. We will train our model using datasets obtained from the Yahoo Financial API that include historical data from the top 100 companies (NIFTY 100) in the NSE and BSE from 2010 to 2021.

***Keywords* - Portfolio Management, Machine Learning Models (AI/ML), Historical Data, Financial Markets, Data Vector, ML algorithm, Efficiency, etc.**

I. INTRODUCTION

Artificial intelligence-based machine learning (ML) models are the latest technological innovation to hit the portfolio management industry. And while, at least so far, ML models have not been a panacea for active investment managers, they've proven themselves a valuable tool by augmenting human decision-making around important activities such as asset allocation, risk management, and portfolio construction.

Investment Portfolio Management is a concept where the risk correlated to the investment portfolio is reduced and has also tried to maximize the profit if it is withdrawn early from the particular stock. It has also been seen that public moods towards particular stocks are related with the financial markets to a greater extent.

We plan to use long short term memory networks, wherein we firstly do the user's savings,, incomes, assets, liabilities, etc. where we compare different datasets related to the stock, which helps in getting better predictions. Secondly, we compare the portfolios of different Superstars to get a better prediction. Lastly, we go through the historical data that are available and compare it, so we get predictions accurately.

In this project, we aim to build a system for predicting portfolio risk using AI/ML and provide insights on how the stocks will perform. We will train our model on datasets which include historical data and predict the outcome that where user can invest Portfolio management is the practice of choosing and managing various investments, with the twin goals of maximizing returns while minimizing risk, for an entity (be it an institution, a company, or an individual investor). While some prefer to manage their own investments – especially in the era of trading apps and retail market manias – it's very common for portfolio managers to oversee an entity's portfolio of assets.

II. METHODOLOGY

Index Construction: We have briefly introduced the KBB index tree structure, which assists us in introducing the index construction. In the process of index construction, we first generate a tree node for each document in the collection. These nodes are the leaf nodes of the index tree. Then, the internal tree nodes are generated based on these leaf nodes. The formal construction process of the index is presented in Algorithm 1. Note that the index tree T built here is a plaintext.

Search Process: The search process of the UDMRS scheme is a recursive procedure upon the tree, named as "Greedy Depth-first Search (GDFS)" algorithm. We construct a result list denoted as RList, whose element is defined as ⟨ RScore, FID⟩ . Here, the RScore is the relevance score of the document fFID to the query, which is calculated according to Formula (1). The RList stores the k accessed documents with the largest relevance scores to the query. The elements of the list are ranked in descending order according to the RScore, and will be updated timely during the search process.

Security analysis: We analyze the BDMRS scheme according to the three predefined privacy requirements in the design goals:

- **Index Confidentiality and Query Confidentiality:** In the proposed BDMRS scheme, Iu and TD are obfuscated vectors, which means the cloud server cannot infer the original vectors Du and Q without the secret key set SK. The secret keys M1 and M2 are Gaussian random matrices. The attacker (cloud server) of COA cannot calculate the matrices merely with cipher text. Thus, the BDMRS scheme is resilient against cipher text-only attack (COA) and the index confidentiality and the query confidentiality are well protected.

- **Query Unlink ability:** The trapdoor of query vector is generated from a random splitting operation, which means that the same search requests will be transformed into different query trapdoors, and thus the query unlink ability is protected. However, the cloud server is able to link the same search requests according to the same visited path and the same relevance scores.

- **Data Privacy:** In this scheme, the confidentiality of the index and query are well protected that the original vectors are kept from the cloud server. And the search process merely introduces inner product computing of encrypted vectors, which leaks no information about any specific keyword. Thus, the keyword privacy is protected in the known cipher text model. But in the known background model, the cloud server is supposed to have more knowledge, such as the term frequency statistics of keywords. This statistic information can be visualized as a TF distribution histogram which reveals how many documents are there for every TF value of a specific keyword in the document collection. Then, due to the specificity of the TF distribution histogram, like the graph slope and value range, the cloud server could conduct TF statistical attack to deduce/identify keywords.

III. MODELING AND ANALYSIS

The model proposes to take various inputs from different sources to build the effective risk manager. When input is given, all the relevant news from credible and available sources will be fetched along with top few tweets from twitter, Open interest (if available) and FIIS and DIIS data of current day, markets closing point price will be stored in the database. After the first stage of information collection is done the processing of data is initiated by sending NEWS and Tweets to the classifier to predict the general sentiment of the people and to classify them into Positive to Negative sentiments. The sentiments, current day market closing price of stock and FIIS and DIIS buying and selling activity will be given as an input to the LSTM Neural network. The predicted output will be shown to the user along with General Market sentiment and Open interest of the underlying asset.

Our contributions are summarized as follows:

- Portfolio optimization is a hot research topic, which has attracted many researchers in recent decades. Better portfolio optimization model can help investors earn more stable profits. This paper uses classifiers like Decision Tree (DT) to build prediction-based portfolio optimization models which own the advantages of both deep learning technology and modern portfolio theory. These models first use DT to predict each stock's future return. Then, predictive errors of DT are applied to measure the risk of each stock. Next, the portfolio optimization models are built by integrating the predictive returns and semi-absolute deviation of predictive errors. These models are compared with three equal weighted portfolios, where their stocks are selected by DMLP, LSTM neural network and DT respectively. Also, two prediction-based portfolio models built with support vector regression are used as benchmarks.
- This system applies component stocks of China or similar country securities 100 index in Chinese stock market as experimental data. Experimental results present that the prediction-based portfolio model based on DMLP performs the best among these models under different desired portfolio returns, and high desired portfolio return can further improve the performance of this model. This paper presents the promising performance of DT in building prediction based portfolio models.

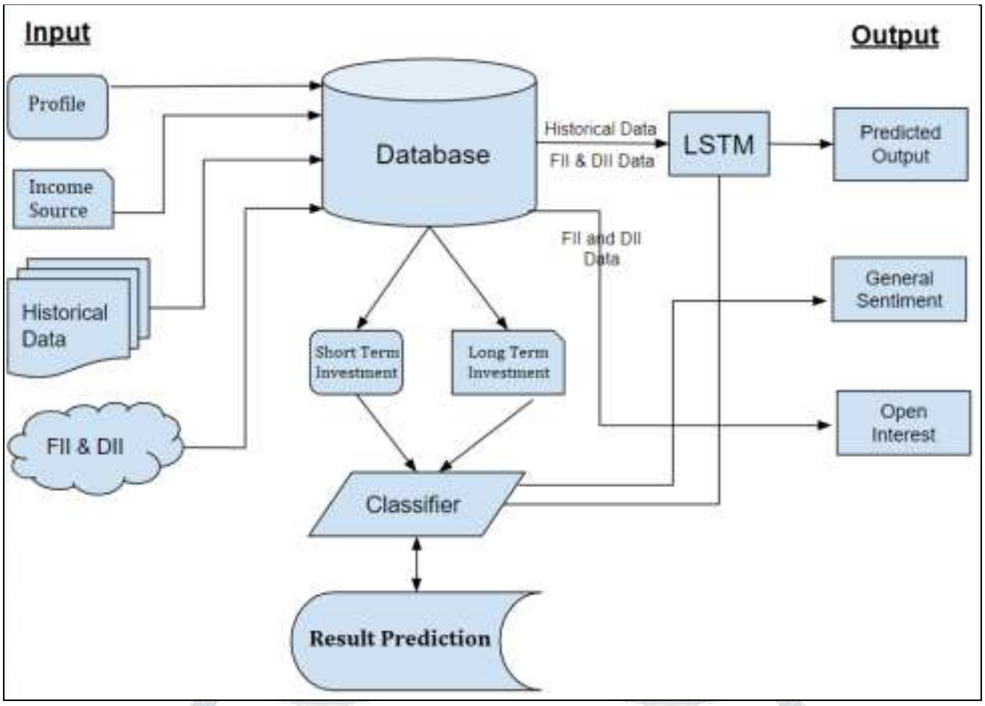


Figure 1: Architecture of Proposed System

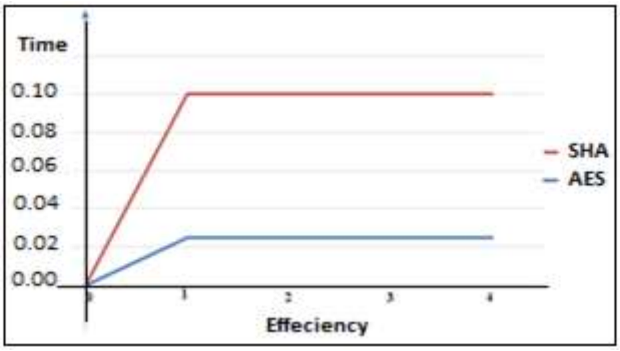
IV. RESULTS AND DISCUSSION

- Due to the special structure of our tree-based index, the proposed search scheme can flexibly achieve sub-linear search time and deal with the deletion and insertion of documents.
- We design a searchable encryption scheme that supports both the accurate prediction and flexible dynamic operation on document collection.
- Due to the special structure of our tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic. And in practice, the proposed scheme can achieve higher search efficiency by executing our algorithm.
- High Accuracy & great performance.
- Proposed System uses different algorithms to increase accuracy rate.
- It is user friendly application.
- Easy to use.
- Time saving.

The goal of the present system is to develop a web-based approach for the portfolio network toward use in order to avoid number of investment plans on users along with the storage and transmission of confidential data records.

The following parameters are used in the analysis of the results:

- Time consumption
- Response Time
- Computation Cost
- Performance accuracy



The Figure 2: Time & Efficiency Graph

Present entire setup get captured the greatest number of qualities or input data parameters, although the focus is mostly on Performance of Operation and its Time.

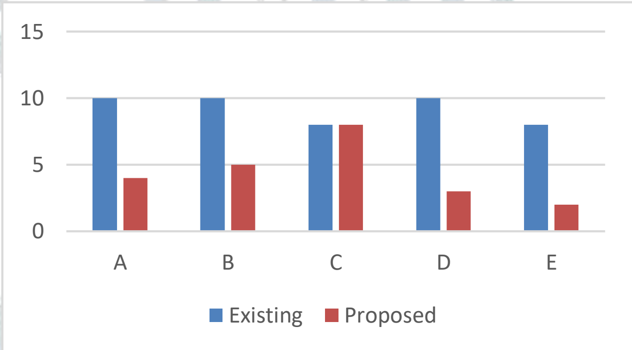
Here to receive the following analytical conclusion for our proposed system if we support a few of attributes.

Table1. Comparison of displacement of all 5 cases

Sr. No.	Parameter	Existing	Proposed
1	A	10	4
2	B	10	5
3	C	8	8
4	D	10	3
5	E	8	2

Where,

A = Time Consumption.
B = Response Time.
C = Computation Cost.
D = Performance accuracy.
E = Scalable & User Friendly.



The Figure 3: Time line chart of Result Analysis

V. CONCLUSION

In this proposed system we explore the properties of machine learning factor-based portfolios. We then examine whether factor-implied covariance matrices based on machine learning dimensionality reduction techniques can benefit minimum-variance portfolios comprised of individual stocks. Overall, our findings indicate that machine learning can help improve factor-based portfolio optimization.

ACKNOWLEDGEMENTS

We would prefer to give thanks the researchers likewise publishers for creating their resources available. We are conjointly grateful to guide, reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

VI. REFERENCES

[1] Markowitz, H.M. (March 1952). "Portfolio Selection". The Journal of Finance. 7 (1): 77–91. doi:10.2307/2975974. JSTOR 2975974.
[2] Sharpe, William F. (1964). "Capital Asset Prices – A Theory of Market Equilibrium Under Conditions of Risk". Journal of Finance. XIX (3): 425–442. doi:10.2307/2977928. hdl:10.1111/j.1540-6261.1964.tb02865.x. JSTOR 2977928.
[3] The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, John Lintner, 1965, Review of Economics and Statistics. 47:1, pp. 13–37.
[4] "Equilibrium in a Capital Asset Market", Econometrica, 34, 1966, pp. 768–783.
[5] "A fast and elitist multiobjective genetic algorithm: NSGA-IIK", Deb, A Pratap, S Agarwal, T MeyarivanIEEE transactions on evolutionary computation 6 (2), 182-197, 2002
[6] Anagnostopoulos, K.P. & Mamanis, Georgios. (2010). A portfolio optimization model with three objectives and discrete variables. Computers & Operations Research. 37. 1285-1297. 10.1016/j.cor.2009.09.009.
[7] Ye Wang, Bo Wang, Xinyang Zhang,A New Application of the Support Vector Regression on the Construction of Financial Conditions Index to CPI Prediction, Procedia Computer Science, Volume 9, 2012, Pages 1263-1272,

[8] Felipe Dias Paiva, Rodrigo Tomás Nogueira Cardoso, Gustavo Peixoto Hanaoka, Wendel Moreira Duarte, Decision-making for financial trading: A fusion approach of machine learning and portfolio selection, Expert Systems with Applications, Volume 115, 2019,Pages 635-655.

[9] Rapach, David & Zhou, Guofu, 2013."Forecasting Stock Returns," Handbook of Economic Forecasting, in: G. Elliott & C. Granger & A. Timmermann (ed.),Handbook of Economic Forecasting, edition 1, volume 2, chapter 0, pages 328-383.

[10] RAPACH, D. E., STRAUSS, J. K., & ZHOU, G. (2013). International Stock Return Predictability: What Is the Role of the United States? The Journal of Finance, 68(4), 1633–1662.

[11] Harvey, Campbell R. and Liu, Yan, Detecting Repeatable Performance (January 21, 2018).

[12] Li Bin, SHAO Xinyue, LI Yueyang. Research on fundamental quantitative investment driven by machine learning [J]. China Industrial Economics,2019(08):61-79.

[13] Jigar Patel, Sahil Shah, Priyank Thakkar, K Kotecha, Predicting stock market index using fusion of machine learning techniques, Expert Systems with Applications, Volume 42, Issue 4,2015, Pages 2162-2172