IFERP
ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Vol 10, Issue 6, June 2023

Maximizing Portfolio Performance using Hybrid Portfolio Construction

[1] Rajat Jaiswal, [2] Dr. Namita Srivastava, [3] Dr. Manoj Jha

[1] Ph.D. scholar, MANIT-Bhopal, Madhya Pradesh, India
[2] Professor, MANIT-Bhopal, Madhya Pradesh, India
[3]Assistant Professor, MANIT-Bhopal, Madhya Pradesh, India
Corresponding Author Email: [1] 213104003@stu.manit.ac.in, [2] namitas@manit.ac.in, [3] m_jha28@rediffmail.com

Abstract— ***The future performance of stock markets is an essential factor for portfolio creation. With the advancement of machine learning techniques, new possibilities have opened up for incorporating prediction concepts into portfolio selection. The paper proposes a hybrid approach, involving machine learning algorithms for stock return prediction and a mean-VaR model for portfolio selection, as a unique portfolio construction technique. Two machine learning regression models, XGBoost and linear regression, are used for stock prediction, and a novel optimizer, Grey Wolf optimizer, is employed for parameter optimization with both XGBoost and linear regression. The results show that the mean-VaR model with linear regression prediction produces better results than the mean-VaR model with XGBoost prediction.***

Keywords: XGBoost, Linear Regression, Grey wolf optimizer, Value at risk (VaR).

I. INTRODUCTION

Investment is defined as the investment of money in order to obtain additional or specific benefits in exchange for money. The investment has a risk that the investor must bear in addition to the benefits (returns). The higher the expected rate of return for an investor, the greater the risk to be covered by the investor. By forming an appropriate portfolio, the level of risk can be minimised at a certain rate of stock portfolio expectations.[1]. As a result, stock portfolio optimization is critical in determining investment portfolio strategies for investors.

Portfolio optimization is the process of finding the optimal combination of assets to maximize return while minimizing risk. It is a multi-objective optimization problem studied by researchers, investors and fund managers to achieve the best possible outcome. An optimal portfolio optimization model can help investors to achieve their desired returns with a lower amount of risk [2].

The Markowitz mean–variance model is the cornerstone of contemporary portfolio theory, seeking to maximize portfolio returns while minimizing investment risk[3]. This model produces an efficient frontier, which reveals a portfolio that reduces overall risk for a given expected return. To address the drawbacks of the MV model, various other models have been developed, mean-semi-variance model [4], mean absolute deviation model[5], mean semi-absolute deviation models [6].

In recent years, machine learning has been increasingly used in portfolio optimization in quantitative finance. The traditional Markowitz Model (MV Model) uses historical data to determine the optimal portfolio, but lacks predictive capabilities. Thus, machine learning algorithms have been employed to predict returns and volatility for future investments[7]. To maximize profits, investors must consider multiple aspects when making decisions in the stock market, and the incorporation of stock price prediction methods into portfolio optimization could prove beneficial.

In the machine learning domain, there are various ensemble learning algorithms that are used to reduce prediction bias and variance, thus increasing predictive performance beyond those of a single algorithm [8]. These algorithms include AdaBoost[9], GBDT, and XGBoost [10] which has been gaining much attention for its low computational complexity, high prediction accuracy, and remarkable efficiency. XGBoost is an improved GBDT, composed of multiple decision trees, and is used for both classification and regression. Recently, XGBoost has been implemented for forecasting in the financial sector.

This paper presents a unique portfolio construction technique that combines machine learning algorithms for stock return prediction and a mean–VaR (value-at-risk) model for portfolio selection. Extreme Gradient Boosting (XGBoost) and linear regression are used to forecast stock values for the upcoming period and Grey Wolf optimizer is used to optimize the hyperparameters of the XGBoost and linear regression. In the first stage, stocks with the most potential returns are chosen. Moving on, the mean–VaR portfolio optimization model is applied for portfolio selection in the second stage.

II. LITERATURE REVIEW

Ta [11] highlighted how machine learning can be used for quantitative trading, employing linear regression and support vector regression for stock movement forecasting. Optimization strategies were applied to maximize gains and minimize risks. Both prediction models proved successful in the short-term, with linear regression proving more accurate and profitable. Including technical indicators in the dataset improved prediction accuracy.

Chen [12] introduced a new portfolio construction process that combined machine learning for stock prediction and mean–variance (MV) model portfolio selection. The hybrid model had two stages: the first employed a combination of Extreme Gradient Boosting (XGBoost) and an improved firefly algorithm (IFA) to predict future stock prices. The IFA was used to fine-tune the XGBoost's hyperparameters. In the second stage, stocks with higher potential returns were identified, and the MV model was used for portfolio selection. Experiments on the Shanghai Stock Exchange showed that the proposed method outperformed traditional methods (without stock prediction) and benchmarking in terms of returns and risks.

Sharma [13] suggested a hybrid deep learning method, DBRNN, for predicting portfolio returns, optimized using the hybrid meta-heuristic algorithm HH-DHO(Harris Hawks-Deer Hunting Optimization). The combination of standard algorithms was seen to be more successful than regular algorithms. With the predicted information, the HH-DHO algorithm was used to select companies with high returns. The results of the analysis showed that the proposed method was better than existing methods and benchmarks in terms of returns and risks.

[14]proposed a hybrid machine learning model combining R-CNN-BiLSTM and MV for predicting future stock closing prices and forming an optimal portfolio. The study addressed the gap in existing literature by proposing a new approach for portfolio formation and utilizing robust input for machine learning training. Three LSTM-based machine learning models were used for comparison, and the method included a stock selection process to guarantee the quality of stock inputs.

[15] used Support Vector Machine (SVM) for stock price prediction and MV model for portfolio selection, selecting 85 active companies out of 450 listed on Tehran Stock Exchange. MOGWO and NSGA II were used for optimization in the MV model. The results showed that MOGWO algorithm had a higher return rate of 133.13% with a risk of 3.346%, as compared to the portfolio return of 107.73% and risk of 1.459% obtained from the NSGA-II algorithm. Comparison of the solutions indicated that MOGWO algorithm was more effective in stock portfolio optimization than NSGA-II.

[16] proposed a novel portfolio optimization model based on prediction, using AE for feature extraction and a LSTM network for stock return prediction. The worst-case omega model was used to construct the portfolio optimization model using predicted and historical returns. Empirical results showed that the proposed model had better performance than an equally weighted portfolio and achieved a satisfactory return even after deducting transaction fees.

[17] proposed a hybrid approach combining machine learning algorithms and a mean-VaR portfolio selection model for constructing portfolios. Regression models such as Random Forest, XGBoost, AdaBoost, SVR, KNN, and ANN were used to forecast stock values, while the mean-VaR model was used for portfolio selection. The research sample included datasets from the Bombay Stock Exchange, Tokyo Stock Exchange, and Shanghai Stock Exchange. It was found that the mean-VaR model with AdaBoost prediction had a better performance than other models.

III. PREDICTION MODELS

a) XGBoost (Extreme Gradient Boosting)

XGBoost is a distributed gradient boosting toolkit that has been tuned for performance[18]. It uses a recursive binary partitioning strategy to obtain the optimal model by choosing the best partition at each step. XGBoost's tree-based nature makes it insensitive to outliers, and like many boosting methods, it is resistant to overfitting, which makes model selection much easier[19]. Eq. (1) depicts the XGBoost model's regularized objective at the t^{th} training step, where $l(y^{(t)}{}_{pred}, y_{act})$ denotes the loss, which refers to calculating the difference between the simulated value $y^{(t)}{}_{pred}$ and the associated ground truth y_{act}.

$$L^{(t)} = \sum_i l\left(y^{(t)}{}_{pred}, y_{act}\right) + \sum_k \Omega(f_k) \qquad (1)$$

Where $\Omega(f_k) = \Upsilon\mathrm{T} + \frac{1}{2}\lambda||\omega||^2$ represents the complexity of the kth tree, in which T signifies the number of leaves and $||\omega||^2$ specifies the $\ell 2$ norm of all leaf scores for training examples. When searching the tree, the parameters γ and λ regulate the degree of conservatism.

The tuned parameters for XGBoost are n-rounds (to determine a maximum number of iterations), max-depth (to control the depth of the tree), learning rate, and gamma (to control regularization to prevent overfitting).

b) Mean-VaR portfolio optimization model

Suppose there are $n \geq 2$ number of risky assets that the investor decides to invest for a fixed time period T. Suppose r_i represents the asset's return rate i. Then the expected rate of return of the asset i is given by

$$\mu_i = \frac{1}{\mathrm{T}} \sum_{t=1}^{n} r_{it}. \qquad (2)$$

Let x_i be the proportion of the wealth invested in asset i in such a way that $\sum_{i=1}^{n} x_i = 1$.[20]

Then the rate of return of the portfolio is

$$r_p = \sum_{i=1}^{n} r_i\, x_i \qquad (3)$$

The expected return of the portfolio is

$$\mu_p = \sum_{i=1}^{n} \mu_i\, x_i \qquad (4)$$

And the variance of the portfolio is

$$\sigma_p{}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}\, x_i x_j. \qquad (5)$$

Further, the definition for VaR of the portfolio is provided as follows:

Definition

Let β^* be the β-quantile of standard normal distribution, $\beta \in(0.5,1]$. [21]Then, for a fixed period of time, the portfolio's VaR may therefore be expressed as follows:

$$VaR = \beta^*\sigma_p - \mu_p = \beta^*\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\,x_i x_j} - \sum_{i=1}^{n}\mu_i x_i \tag{6}$$

This paper has considered the mean–VaR portfolio optimization model, which is adapted from (Sheng et al., 2012). Following is the mean–VaR portfolio optimization problem.

$$\text{Min } VaR = \beta^*\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\,x_i x_j} - \sum_{i=1}^{n}\mu_i x_i. \tag{7}$$

$$\text{Subject to } \sum_{i=1}^{n}\mu_i x_i = \mu_{fix}, \tag{A}$$

$$\sum_{i=1}^{n} x_i = 1; 0 \le x_i < 1 \;\forall\; i = 1,2,3 \ldots\ldots, n. \tag{B}$$

This mean–VaR model is constructed by allying VaR as a risk measure to the mean–variance portfolio optimization model. In the model, the objective is to minimize the VaR of the portfolio. Meanwhile, as presented in constraints (A), (B), the investor wants that the expected return of the portfolio must achieve the preset minimum expected return value μ_{fix}, and the investment proportion x_i must sum to one.

c) Linear Regression

Linear Regression is a supervised machine learning algorithm used for regression tasks[22]. It is used to predict a target value based on independent variables. It is often used to determine the relationship between variables and forecasting. The dependent variable in regression can be referred to as an outcome variable, criterion variable, endogenous variable, or regressand, while the independent variables can be called exogenous variables, predictor variables, or regressors. To create a regression model, a set of X and Y values are used to learn a function, which can then be used to predict Y from an unknown X. The Y value is referred to as the Criterion Variable and the X value is referred to as the Predictor Variable.

Hypothesis function for linear regression

$$y = \theta_1 + \theta_2 \cdot x$$

The model gets the best regression fit line by finding the best $\theta_1$and θ_2 values.

θ_1: intercept, θ_2: coefficient of x.

Once we find the best θ_2 and θ_2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x. [11]

Cost Function: By achieving the best fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the $\theta_1$and θ_2 values, to reach the best value that minimize the error between predicted y value and actual y value.

$$\text{Minimize} \quad \frac{1}{n}\sum_{i=1}^{n}(y_{pred} - y_{act})^2 \tag{8}$$

$$J = \frac{1}{n}\sum_{i=1}^{n}(y_{pred} - y_{act})^2 \quad \text{(Cost Function)} \tag{9}$$

d) Grey Wolf Optimizer

Grey Wolf Optimizer (GWO) is an optimization algorithm that is based on the leadership hierarchy and hunting mechanism of grey wolves[15]. It was developed to solve optimization problems in a wide range of fields, including portfolio optimization. GWO has been found to be suitable for optimization problems that have multiple local optima and large search spaces. It uses multiple leaders and follows the same hunting and exploration strategy used by grey wolves to search for prey. It has been proven to be an effective and efficient portfolio optimization algorithm.

The mathematical model of encircling behaviour is modelled as:

$$\vec{D} = |\vec{C}\cdot\vec{X_p} - \overrightarrow{X(t)}| \tag{10}$$

$$\vec{X}(t+1) = |\vec{X_p}(t) - \vec{A}\cdot\vec{D}| \tag{11}$$

Where t= current iterations. $\vec{X_p}$= position vector of the prey. $\vec{X}$=position vector of Grey wolf. $\vec{A}$,$\vec{C}$ are coefficient vectors.

$\vec{A}$,$\vec{C}$ vectors are calculated as

$$\vec{A} = 2\cdot\vec{a}\cdot\vec{r_1} - \vec{a} \tag{12}$$

$$\vec{C} = 2\cdot\vec{r_2}$$

where $\vec{r_1}, \vec{r_2}$ are random vectors in [0,1] and component $\vec{a}$ is linearly decrease from 2 to 0 over the course of iterations.

Hunting:

In each iteration, omega wolves update their positions in accordance with the positions α, β, and δ because α, β, and δ have better knowledge about the potential location of prey.

$$\vec{D_\alpha} = |C_1\cdot\vec{X_\alpha} - \overrightarrow{X(t)}| \tag{13}$$

$$\vec{D_\beta} = |C_2\cdot\vec{X_\beta} - \overrightarrow{X(t)}| \tag{14}$$

$$\vec{D_\delta} = |C_3\cdot\vec{X_\delta} - \overrightarrow{X(t)}| \tag{15}$$

$\vec{X_\alpha}, \vec{X_\beta}, \vec{X_\delta}$ are position vectors of α, β, δ.

$$\vec{X_1} = |\vec{X_\alpha} - A_1\cdot\vec{D_\alpha}| \tag{16}$$

$$\vec{X_2} = |\vec{X_\beta} - A_2\cdot\vec{D_\beta}| \tag{17}$$

$$\vec{X_3} = |\vec{X_\delta} - A_3\cdot\vec{D_\delta}| \tag{18}$$

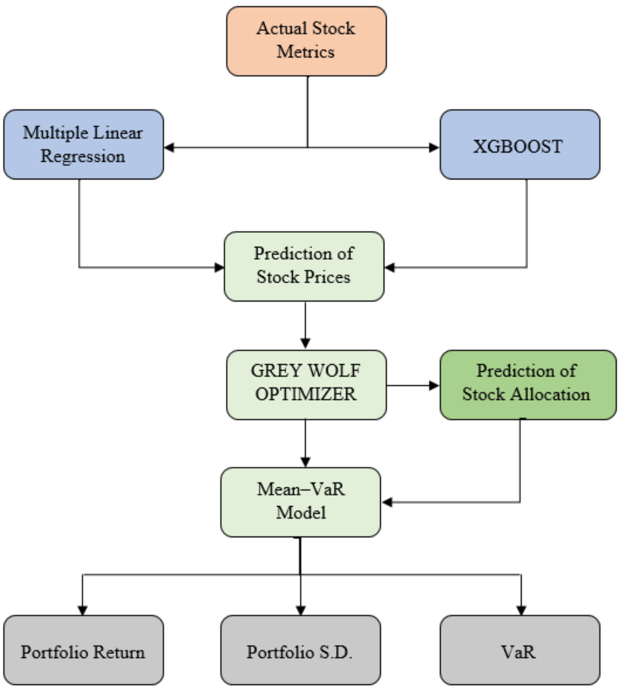
$C_1, C_2, C_3, A_1, A_2, A_3$ are the coefficient vectors.

$$\vec{X}(t+1) = (\vec{X_1} + \vec{X_2} + \vec{X_3})/3 \tag{19}$$

IV. METHODOLOGY

Investors in the financial market typically seek to construct an optimal investment portfolio with the greatest return potential and least amount of risk. To achieve this, the primary focus of this paper is on selecting stocks which have the highest returns in order to build a portfolio from a predictive perspective. This strategy involves two stages:

- **Stock return prediction:** For stock return prediction, machine learning regression techniques such as XGBoost and linear regression are taken into consideration. To assess the accuracy of the models, one metric, root-mean-squared error, is utilized in this work. Stocks with superior performance obtained from various models are taken into account for the subsequent stage.
- **Portfolio optimization:** The objective of this phase is to ascertain the proportion of wealth that is allocated to each stock. To achieve this, the mean–VaR model is employed, with top-performing stocks identified from different models to construct the asset allocation of the portfolio.



Working flowchart of methodology used

V. RESULTS

In this paper one-year stock data of 10 random companies namely AAPL, ADBE, C, DIS, F, MSFT, MS, GME, TSLA and AMZN was taken from Yahoo Finance(from 1/1/2021 to 15/9/2022) to test the efficiency of our model. Two machine learning models, linear regression and XGBoost were used for stock prediction. Further Grey Wolf Optimizer was used to reduce the Mean Squared Error in both XGBoost and Linear Regression models. As a result, it is found that the Linear Regression model with GWO is more efficient than the XGBoost model with GWO. Further Mean-VaR model was applied for portfolio selection and obtained a return of 112%, a VaR of 3.99% at 5% confidence level, and a standard deviation of 2.35%. Based on the ranking of the stocks, an optimized portfolio can be formed with a consideration of both the return and risk measure.

Comparison of Mean squared error using XGBoost and Linear Regression:

Company	XGBoost	XGBoost+GWO	Linear Regressor+GWO
AAPL	0.2122	0.4040	0.1201
ADBE	7.7876	7.2932	1.4306×10^{-25}
C	0.4562	0.4399	0.1288
DIS	0.5911	0.5781	1.08508×10^{-16}
F	0.0317	0.0441	0.01067
MSFT	1.9247	1.8429	0.9673
MS	0.6110	1.2702	0.2933
GME	0.5234	0.4843	9.4487×10^{-23}
TSLA	2.5891	3.1026	6.63510×10^{-17}
AMZN	0.5431	0.3145	2.4024×10^{-20}

VI. CONCLUSION AND DISCUSSION

Investors need an effective strategy in order to make sound financial decisions in the current market. This practice can motivate them to invest their capital into the financial markets and optimize their investments efficiently. To aid in making more profitable and wise decisions, machine learning algorithms have become essential tools, particularly when it comes to investing in the financial market. This paper presents the fundamentals of portfolio optimization in terms of stock prediction and stock selection. In the experiment, both linear regression and XGBoost (Extreme Gradient Boosting) models are used to predict the stock price. The results showed that both regression prediction models were effective in prediction with high accuracy on average. The linear regression model with the optimizer GWO (Grey Wolf Optimizer) performed better than XGBoost regression model with GWO. This experiment was conducted to check the efficiency of the model. With the suggested models, this work may help with portfolio optimization to get better results by increasing the objective functions and applying them to larger data sets.

REFERENCES

[1] V. D. Vasiani, B. D. Handari, and G. F. Hertono, “Stock portfolio optimization using priority index and genetic algorithm,” *J. Phys. Conf. Ser.*, vol. 1442, no. 1, 2020, doi: 10.1088/1742-6596/1442/1/012031.

[2] F. Yang, Z. Chen, J. Li, and L. Tang, “A novel hybrid stock selection method with stock prediction,” *Appl. Soft Comput. J.*, vol. 80, pp. 820–831, 2019, doi: 10.1016/j.asoc.2019.03.028.

[3] H. Markowitz, “Portfolio selection,” *J. Finance*, vol. 7, pp. 77–91, 1952.

[4] H. M. and H. M. M. Markowitz, “Portfolio selection: efficient diversification of investments.,” *J.Wiley*, 1967.

[5] H. Y. Hiroshi Konno, “Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market,” *informs*, vol. 37, 1991.

[6] M. G. S. Luca chiodi, Renata Mansini, “Semi-Absolute Deviation Rule for Mutual Funds Portfolio Selection,” *Ann. Oper. Res.*, vol. 124, pp. 245–265, 2003.

[7] B. M. Henrique, V. A. Sobreiro, and H. Kimura, “Literature review: Machine learning techniques applied to financial market prediction,” *Expert Syst. Appl.*, vol. 124, pp. 226–251, 2019, doi: 10.1016/j.eswa.2019.01.012.

[8] L. R. Omer Sagi, “Ensemble learning:A survey,” *Wiley Interdiscip. Rev.*, vol. 8, no. 4, 2018.

[9] X. dan Zhang, A. Li, and R. Pan, “Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine,” *Appl. Soft Comput. J.*, vol. 49, pp. 385–398, 2016, doi: 10.1016/j.asoc.2016.08.026.

[10] J. Nobre and R. F. Neves, “Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets,” *Expert Syst. Appl.*, vol. 125, pp. 181–194, 2019, doi: 10.1016/j.eswa.2019.01.083.

[11] V. D. Ta, C. M. Liu, and D. Addis, “Prediction and portfolio optimization in quantitative trading using machine learning techniques,” *ACM Int. Conf. Proceeding Ser.*, pp. 98–105, 2018, doi: 10.1145/3287921.3287963.

[12] W. Chen, H. Zhang, M. K. Mehlawat, and L. Jia, “Mean–variance portfolio optimization using machine learning-based stock price prediction,” *Appl. Soft Comput.*, vol. 100, p. 106943, 2021, doi: 10.1016/j.asoc.2020.106943.

[13] M. Sharma and H. S. Shekhawat, “Portfolio optimization and return prediction by integrating modified deep belief network and recurrent neural network,” *Knowledge-Based Syst.*, vol. 250, p. 109024, 2022, doi: 10.1016/j.knosys.2022.109024.

[14] A. Chaweewanchon and R. Chaysiri, “Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning,” *Int. J. Financ. Stud.*, vol. 10, no. 3, 2022, doi: 10.3390/ijfs10030064.

[15] N. B. Mazraeh, A. Daneshvar, M. Madanchi Zaj, and F. R. Roodposhti, “Stock Portfolio Optimization Using a Combined Approach of Multi Objective Grey Wolf Optimizer and Machine Learning Preselection Methods,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/5974842.

[16] Y. Ma, W. Wang, and Q. Ma, “A novel prediction based portfolio optimization model using deep learning,” *Comput. Ind. Eng.*, vol. 177, no. October 2020, p. 109023, 2023, doi: 10.1016/j.cie.2023.109023.

[17] J. Behera, A. K. Pasayat, H. Behera, and P. Kumar, “Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets,” *Eng. Appl. Artif. Intell.*, vol. 120, no. August 2022, p. 105843, 2023, doi: 10.1016/j.engappai.2023.105843.

[18] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016, doi: 10.1145/ 2939672.2939785.

[19] H. Kim, “Mean-variance portfolio optimization with stock return prediction using xgboost,” *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 55, no. 4, pp. 5–20, 2021, doi: 10.24818/18423264/55.4.21.01.

[20] J. Wang, “Mean-Variance-VaR Based Portfolio Optimization 1 Introduction,” *October*, pp. 1–19, 2000.

[21] S. Dedu and C. Fulga, “Value-at-risk estimation comparative approach with applications to optimization problems,” *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 4, no. May, 2010.

[22] Z. Peng and X. Li, “Application of a multi-factor linear regression model for stock portfolio optimization,” *Proc. - 2018 Int. Conf. Virtual Real. Intell. Syst. ICVRIS 2018*, no. 3, pp. 367–370, 2018, doi: 10.1109/ICVRIS.2018.00096.