ANALYSIS OF THE MUTUAL FUND NET ASSET VALUE(NAV) BASED ON PORTFOLIO HOLDINGS USING MACHINE LEARNING

Submitted by:
Vipul Khandelwal
Enrollment No. 2020MTCSE015

Thesis Supervisor:
Dr. Tarun Kumar

A thesis submitted to the School of Engineering and Technology in partial fulfillment of the requirements for the award of degree of

Master of Technology
in
Computer Science and Engineering

Department of Computer Science and Engineering
School of Engineering and Technology

Central University of Rajasthan

May 25, 2022

DECLARATION BY THE CANDIDATE

I the undersigned solemnly declare that the project report "ANALYSIS OF THE MUTUAL FUND NET ASSET VALUE(NAV) BASED ON PORTFOLIO HOLDINGS USING MACHINE LEARNING" is based on my own work carried out during the course of our study under the supervision of Dr. Tarun Kumar, Assistant Professor, Department of Computer Science and Engineering, Central University of Rajasthan. I have not submitted the matter presented in this report to any other in- stitution for any other degree/certificate/diploma in Central University of Rajasthan or any other university of india or abroad.

Date: July 13, 2022
Place: Central University of Rajasthan, Ajmer

Vipul Khandelwal
2020MTCSE015

CANDIDATE'S UNDERTAKING

I, Vipul Khandelwal, understand that plagiarism is defined as any one or the combination of the following:

1. Uncredited verbatim copying of individual sentences, paragraphs or illustrations from any source, published or unpublished, including the Internet.

2. Uncredited improper paraphrasing of pages or paragraphs.

3. Credited Verbatim copying of major portion of the paper without clear definition of who did or wrote that. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not result of my work, are properly credited. Long phrases of sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

I affirm that no portion of my work in dissertation titled **"ANALYSIS OF THE MUTUAL FUND NET ASSET VALUE(NAV) BASED ON PORTFOLIO HOLDINGS USING MACHINE LEARNING"** can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand very well that the thesis advisor may not be in a position to check for the responsibility of such incidences of plagiarism in this body of work.

Date: May 25, 2022 — Vipul Khandelwal
Place: Central University of Rajasthan, Ajmer — 2020MTCSE015

CERTIFICATE

This is to certify that the dissertation entitled **"ANALYSIS OF THE MUTUAL FUND NET ASSET VALUE(NAV) BASED ON PORTFOLIO HOLDINGS USING MACHINE LEARNING"** done by **Vipul Khandelwal**, Enrollment No. **2020MTCSE015** is an authentic work carried out by him at Central University of Rajasthan, Ajmer under my guidance. The matter embodied in this work has not been submitted earlier for the award of any degree to the best of my knowledge and belief.

Date: July 13, 2022

Dr. Tarun Kumar (Supervisor)
Assistant Professor
Department of Computer Science and Engineering
School of Engineering and Technology
Central University of Rajasthan, Ajmer

ABSTRACT

Mutual fund analysis is sub-domain of algo trading in stock market where the analysis is focused on the extraction of stock portfolio price and analysis net asset values form its portfolios from unstructured stock price data with respect to its date. In this project, the machine learning algorithm is applied on mutual funds dataset taken from yahoo finance. Before it, many machine learning algorithms are taken, including multivariate regression, decision tree regression, and linear regression. By these algorithms NAV of the mutual fund is calculated with good accuracy to analysis the best result for the dataset. The various models are been created from different portfolio to analysis the NAV of the mutual funds. The multivariate regression model is applied in the last to analysis and predict the NAV. In this, proposed work the R square coefficient is taken out with 0.86 which shows that models is created with good accuracy and good performing dataset.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

At present majority of people are investing or planning for investment as future financial security. There are various way for investments such as real estate, gold, stocks etc. everyone is investing as per there capacity, easy access and ensure the return and risk.

Nowadays investments are rapidly increasing in Mutual funds. Due to its easy access and flexive there are many financial and investment decisions that can be done on the mutual fund. Within the mutual fund, there is a vast amount of fundholding which can be classified for mutual funds and can be used for predicting and analysing NAV values.

The mutuals fund are the type of funds, in which the investment from various investor are collectively invested into various securities such as stocks, short term debt, bonds etc by a facilizing organization. The record of the investment is classified as mutual funds portfolio. These portfolios are typically classified into four categories.

The four primary types of funds are:

Stock or equity funds Money market funds Bond or fixed-income funds Hybrid funds

In stock or equity portfolio, the dedicated shares of investment are invested into stocks or equity of various sectors listed into stock exchange market. The variation of the net asset value of any mutual fund is proportional depends to the current value of these portfolio. The risk factor of the mutual fund is also directly proportional to these portfolios. As majority share investment usually inversed into these.

Money market portfolios have comparatively lower risk as compared to other portfolios. Investment into these portfolios is carry out in the form of debt investments into various organization. These investments are revaluated by various agencies portfolio.

Bond or fixed-income portfolios have comparatively higher risk than money market portfolio. Fixed or bond portfolio are the type of portfolio in which the investor got paid fixed amount over invested amount. These

funds are issued by corporations and the government for raising money to expand their business or any project.

The NAV of mutual funds is largely depending on the stock's portfolio. However, due to dynamic nature of the stock prices, the probability of the risk and returns are very high. The prediction of NAV of the mutual funds is challenging tasks for the experts. This work proposes an approach for prediction of the NAV of the any mutual funds with the use of multiple dependent machine learning models. The approach presented in this work is also applied to a specific mutual fund as a case study.

1.1 MOTIVATION

As the major investment of any mutual funds is invested into the stock portfolio. So the analysis of the stock prices is important for any experts to mange the investments of the mutual funds. The intent behind this works into propose ML based analysis approach for better prediction of NAV as mutual funds. This will help exports to identify profitable and low risk stocks portfolios for the mutual funds.

Chapter 2

OVERVIEW

2.1 MACHINE LEARNING

A class of methods known as machine learning (ML) enables software systems to become increasingly accurate at predicting outcomes without being explicitly changed. Building algorithms that can receive input data, use statistical analysis to predict an output, and update outputs as new data becomes available is the fundamental goal of machine learning. Machine learning processes are similar to those used in data mining and predictive modelling. Both involve searching through data for examples and modifying system operations as necessary. Many people are familiar with machine learning thanks to online shopping and receiving offers that are relevant to their purchases. This occurs because suggestion engines almost constantly utilise machine learning to tailor how online advertisements are delivered. Extortion identification, spam separation, organising security danger location, predictive upkeep, and creating news sources are just a few examples of basic machine learning use cases outside of past tailored advertising.

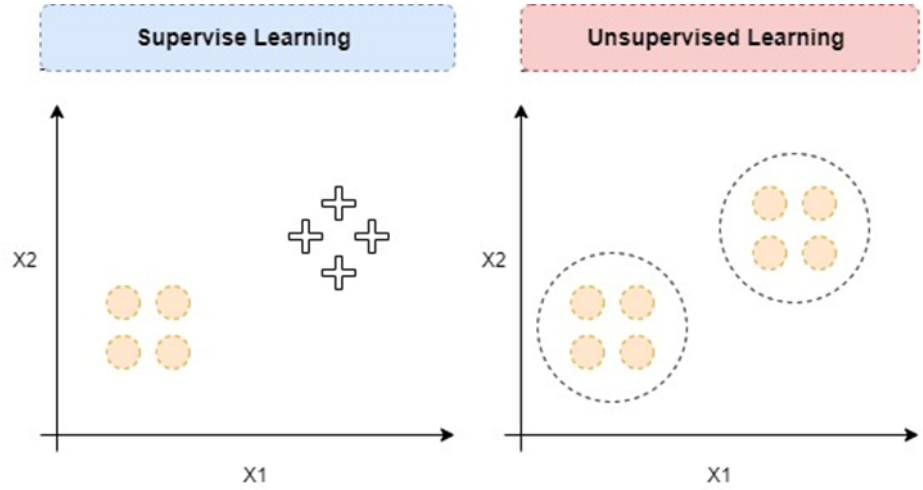

Figure 2.1: Machine Learning types

Three subcategories of machine learning exist: reinforcement learning, unsupervised learning, and supervised learning.

2.1.1 SUPERVISED LEARNING

So could say that while learning is overseen, an instructor is directing it. We have a dataset that functions as a teacher and is responsible for putting together the machine or model. When the model is trained, it can start making predictions or decisions based on new data.

1. Classification Classification is the sub categories of supervised machine learning. In which the model is trained over two or more classes.

2. Regression You have a regression issue when the output variable is a real or continuous value, such as "salary" or "weight." The most fundamental model, linear regression, is just one of many that can be used.

2.1.2 UNSUPERVISED LEARNING

Unsupervised learning models find patterns in the data by learning through observation. When a dataset is provided to the model, it automatically creates clusters in the dataset to identify patterns and linkages. It cannot name the group, just as it cannot name this collection of apples or mangoes, but it will separate every apple from every mango. Imagine that we showed the model images of apples, bananas, and mangoes. Based on certain patterns and linkages, the model would create clusters and divide the dat aset into those groups. Presently If the model receives further data, it adds it to one of the created clusters.

2.1.3 REINFORCEMENT

Reinforcement The ability to interact with the environment and determine the best outcome is known as learning. It pursues the concept of the hit-and-miss method. For a correct or incorrect response, the agent is rewarded or penalised with points, and the model trains the agent depending on the positive reward focus it learned. Additionally, after training, it gets ready to forecast the fresh data that is presented to it.

2.2 DECISION TREE

The decision tree is a sort of supervised learning algorithm (having a predefined target variable) that is for the most part utilized in classification issues. It works for both categorical and continuous input and yield variables. In this strategy, we split the populace or test into at least two homogeneous sets (or sub-populaces) in light of the most noteworthy splitter/differentiator in input variables.

Decision tree is of two types that are as follows: - Categorical Variable Decision Tree: It is called a categorical variable decision tree when it has a categorical target variable. Model: - In the above situation of understudy issue, where the target variable was "Understudy will play cricket or not" for example Indeed or NO. Continuous Variable Decision Tree: It is called a categorical variable decision tree when it has a categorical target variable.

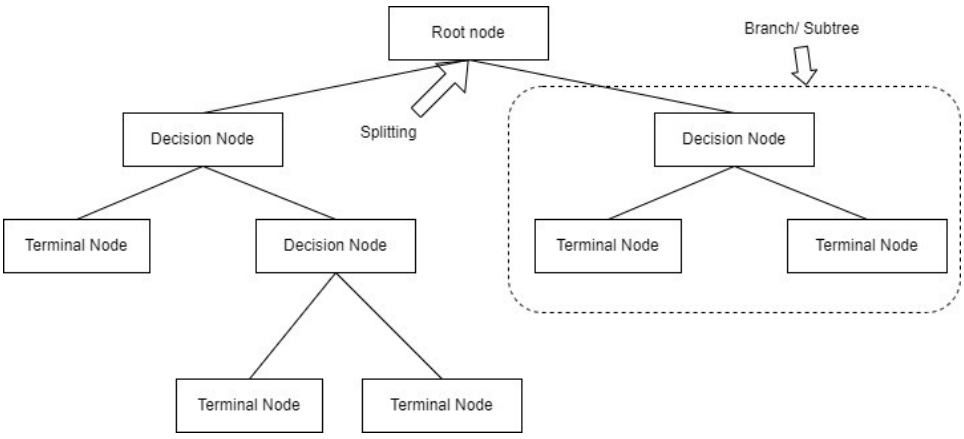

Figure 2.2: Diagram of Decision Tree

The understanding dimension of the Decision Trees algorithm is so natural contrasted and other classification algorithms. The decision tree algorithm endeavours to take care of the issue, by utilizing tree representation. Each internal node of the tree relates to an attribute, and each leaf node compares to a class label.

2.3 LINEAR REGRESSION

Linear regression is a sort of supervised learning algorithm which is the easiest and most efficient way to analyse the regression models of machine learning. Linear regression models are mainly used in machine learning for predicting and analysing numeric variables and continuous variable such as price, salary, age etc.

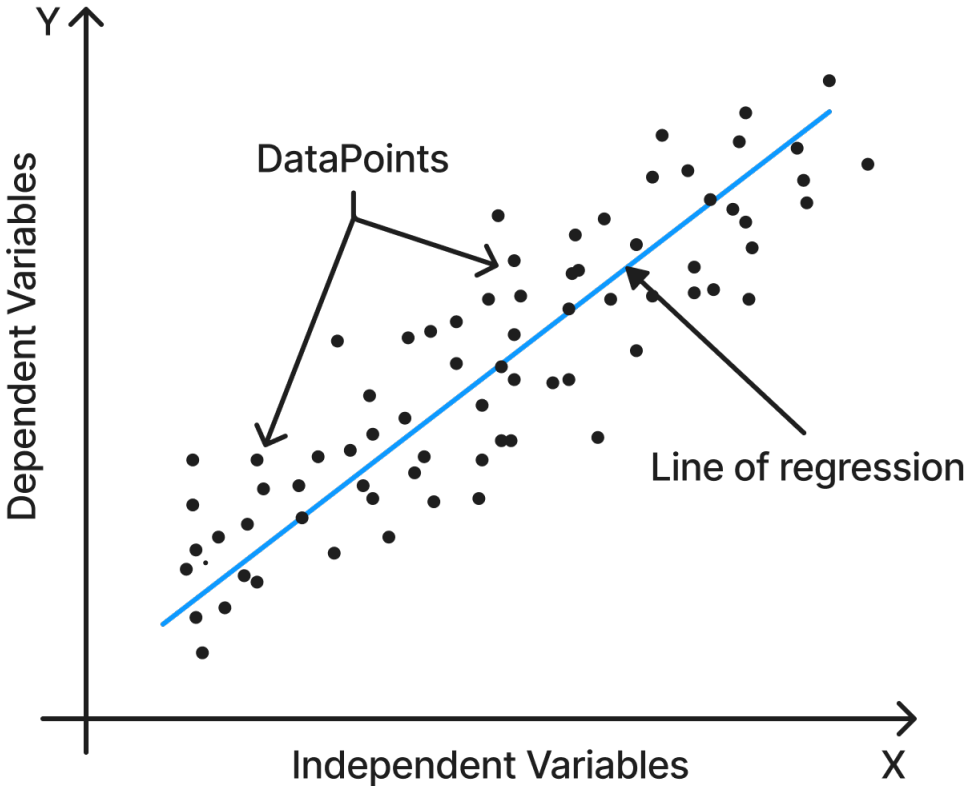

Figure 2.3: Diagram of Linear regression

The link between the variables is represented by a slanted straight line in the linear regression model. Consider the following illustration in figure 3.

2.4 MULTIVARIATE REGRESSION

Multivariate regression is the one of simplest machine learning algorithms. It comes from the family of supervised learning algorithms. Multivariate regression is mainly used for calculating the various measures by which independent variable is dependent to dependent variable.

Equation of multivariate regression is used in the analysis of relationships between dependent and numerous independent variables.

Overview of multivariate regression equation: -

$$Y = b_1X_1 + b_2X_2 + ... + b_nX_n + C \tag{2.1}$$

Where,
Y= Regression's dependent variable
b(1,2,3..n) = Regression's coefficients
X(1,2, 3..n)= Regression's independent variables
C= Regression's constant

In this multivariate regression models, there are as many as independent variables that can be used to predict the dependent variable with good accuracy.

2.5 R2_SCORE OR R2 - COEFFICIENT OF DETERMINATION

R2 or r2 is the coefficient of determination, which is used in statistics to determine how much regression model data fits under the observed dataset. For instance, if the coefficient of determination is 20, it means that 20 of our training data fits into the model. The higher the r square value, the better the dataset fits the model.

$$R2 = 1 - \frac{RSS}{TSS} \tag{2.2}$$

Where,
R2 = coefficient of determination
RSS = sum of square of residuals
TSS = Total sum of square

Chapter 3

LITERATURE REVIEW

There has been impressive research done and as yet going on the mutual fund and nav prediction or forecasting. Analysis of mutual funds turned out to be a piece of research towards the start of the 20th century. This proposed work did analysis on internal holding problems of the portfolios of the mutual fund.

Houssainy et al. [1] use various time series model for forecasting the net asset value of the mutual fund. In this work author compared various models like ARIMA (Autoregressive integrated moving average) , decision tree , Gradient Boosted Trees (GBT) and random forest for predicting gold price fund. In this model 90 data is used for training proposes. For data testing, performance and RMSE for comparing statistical measure. In which they have found that based on RMSE= 38.52, RF outperformed DT, GBT, and ARIMA in predicting future gold prices.

Chen et al. [2] propthe goal of the research is to apply machine learning algorithms, namely the Gaussian hybrid clustering algorithm, to improve the accuracy and timeliness of fund classification. Xiaofei Chen, Shujun Ye, Chao Huang have comparison to the classic K-means and network clustering methods, there step-two GMM method can generate probability that funds belong to a specific category.

In Nayak et al. [3] used machine learning to predict the trends of the stock prices. The approach is use historical prices of stocks along with sentiments data. The apprch need 70 accuracy in prediction of daily stock prices.

Adil Moghar et al. [4] proposed LSTM based approach for forcasting the stock price. As the prices trends of the stock flows with time series like LSTM mode. In this model it has use input layer, hidden layer for modelling the prediction and last one is output layer. In this model data of google and

Nike stock data has been used for prediction. The prediction has been done in different epochs to find out the most relevant result with minimum loss value. The epochs in which models have been created are for 12,25,50,100 epochs. After getting result from each epochs the authors have compared the values and find out best epochs model for predicting future price trends for each stock.

Promptsook et al. [5] proposed a machine learning model using decision tree alogn with XGboost model. The work analysis the equity funds factors of mutual funds. The approach achieve 80 accuracy in predicting the NAV of the mutual fund. The Random Forests Algorithm with a Group offactors inside the type of average of Equity General is the moving average gave the best approach.

C. M. Anish and B. Majhi [6] have work on prediction NAV value of mutual fund using neural network for good fanatical forecasting accuracy. The dataset used in this research paper are NAV of Birla sun life equity fund, ICICI pro top100 , SBI bluechip fund and UTI equity fund. The author has used FFLANN feedback functional link artificial neural network for prediction NAV assets value. Author have used RMSE Root mean square error and MAPE Mean absolute percentage error for calculating accuracy of the model.

Hiransha M et al. [7] have used four different predictive model for stock from three different sector which are automobile, IT sector and banking sector. In [4] they have applied recurrent neural network (RNN),‘‘ Convolutional Neural Network(CNN), A multilayer perceptron (MLP) and long sort time term memory (LSTM) approach for prediction model. In which DL models are outperforming. Hence it was not good model to predict time series interval prediction. For automobile sector stock which is TATA motors it's best results in CNN models where as stock for banking sector which bank of America give its best model in LSTM.

Lawrence K.D. et al. [8] work on to create time series model for forecasting NAV net asset value for asset allocation of the mutual fund. In this research paper the author has use the dataset of Vanguard Wellington Fund. Author have use various type of models like linear tread model, quadratic trend model, exponential model and many more.

Roondiwala et al. [9] haves used LSTM approach to predict the value of nifty50 from NSE. They have used five years of data set which is divided into two parts mainly for validation and training part. In their methodology they are trained the dataset into five stages which are Raw data, data processing feature extraction, training neural network and output genera-

tion. For analysis they have used root mean square method (RMSM).In this research paper the training RMSE of 0.00983 is obtained which predicting the price treads of stock. which is efficient way to find out error matric for numerical prediction.

Chapter 4

PROBLEM STATEMENT

There have been several problems with analysing net asset value of a mutual fund. There are various approaches to analysis the mutual fund. However, Today's analysis and prediction of mutual funds are totally based on forecasting NAV with time series models and technique. First, They are prediction is totally based on net asset value commonly doesn't tell any information about internal portfolio holding. Second, prediction and analysis of net asset value using portfolio have not been done so far.

Chapter 5

PROPOSED WORK

This work proposes a analysis approach for prediction of NAV of mutual funds. As, the mutual fund NAV is depending on the portfolio profiles in NAV the investment is shared. In general trends, the majority of the mutual funds invested are made towards the stock-based portfolios.

This approach propose to analysis the major stock hold and stocks portfolios of the mutual funds. In this the individual machine learning model is propose of each stock portfolio. This machine learning model will predict the future stock trends of the portfolio mostly based on historical prices data. Laster, a decision tree model is proposed to combine the prediction obtain from the individual ML model of each stock price.

This model will be used to predict the future NAV of the mutual funds. The propes approach can be utillied for better investment plans for risk assumed. The workflow of the proposed approach is also illustrated in fig 4.1

5.1 CASE STUDY

A case study data is used for training the model verification and testing of the propose hypothesis. The proposed case study discussed as follows:

Case study, This project is focusing on finding the internal fund holding of the axis bluechip fund that will help us provide the prediction values of the stocks which come under the fund. Stocks like Infosys ltd, Bajaj Finance Ltd, ICICI bank ltd, HDFC bank ltd, Tata consultancy services ltd and Avenue Supermarts ltd holds more than 50 of the fund portfolio holding of axis bluechip fund [10].

Dataset is taken from highly equity holding portfolio fund (Axis Bluechip fund) which invested portfolio into various different sectors which are Banking, IT, NBFC and retailing sectors from yahoo finance. Stock symbol, stock

Table 5.1: Detailed portfolio overview according to its percentage of portfolios holding

Stock Invested In	**Sectors**	**Percentage of fund holding**
Infosys Ltd.	software	9.77%
Bajaj Finance Ltd.	NBFC	9.69%
ICICI Bank	Bank	9.09%
HDFC Bank	Bank	7.47%
TCS Ltd.	software	7.00%
Avenue supermart Ltd.	Retailing	6.97%

series, stock date and previous closing, opening, high, low, final, closing, and average prices and total traded volume. We extract only the day-by-day closing price and the date of each portfolio from these datasets because day-by-day stock prices are preferable because investors make decisions on which stocks to purchase or sell based on the market's closing price. Alone with the closing price we have taken net asset value of the particular fund in which its portfolios hold all the stocks which we have extracted.

In the NAV value prediction, we have face various challenges which deals with various type of classification.

5.2 WORKFLOW

This is working flow chart,

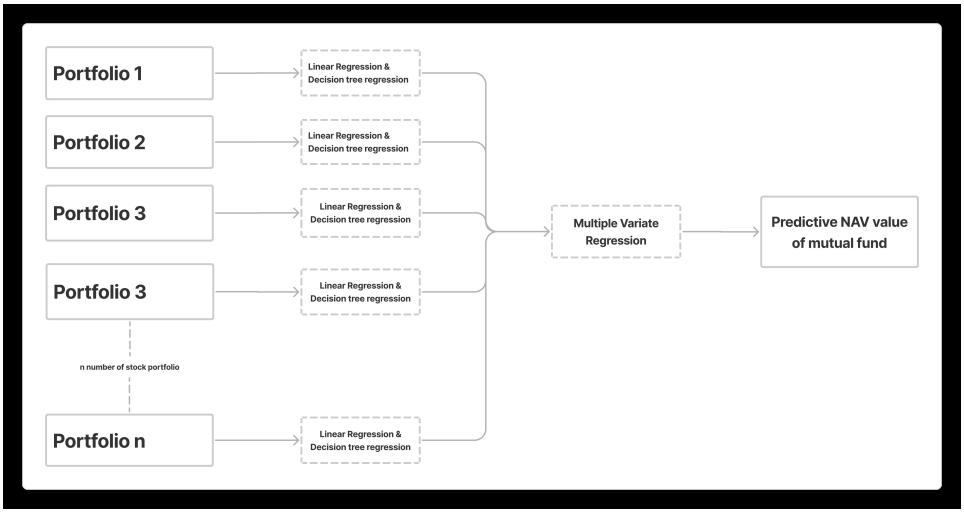

Figure 5.1: Flow chart of the working models to analyse and predictive the nav value of mutual fund

In this propose work the mutual fund is decide according to it equity percentage of its stock portfolio, The stock portfolio is been selected from the internal holding of the mutual fund. There can be n number of stock portfolio in the mutual fund which can choose on there based of equity share holding. After selecting s tock portfolio dataset is trained using linear regression to analysis the stock price trends. Decision tree regression will be used for prediction of the closing price of the stock portfolio. When training part is done our model will be ready for the testing propose. In both model linear regression and decision tree regression testing the model be preforms to check and analyse portfolio performance. As, the predict closing stock portfolio price is been predicted, a new model of multivariate regression is created to analyse the NAV of mutual funds. In Multivariate regression model import the predicted dataset the portfolio into the model. For every date a predicted and analyse the NAV of the mutual fund is taken out from the model. At the end test the model using r square error to analyse the model which shows coefficient of determinant.

Chapter 6

RESULT AND ANALYSIS

6.1 DATASET OVERVIEW

While analysing the portfolio of a mutual fund the investor of the mutual fund can analyze the future prediction value of the fund so that it is a good idea about return of investment.
We are having 6 datasets in which the dataset of all the 6 portfolios merge into a common table to perform the analysis and predict the result.
The data range has been taken out from (2018-01-01 to 2021-12-31) and all the dataset come from the different locations by which we can create models on the dataset.
While creating models, the datasets are divided into the ratio of 20% and 80% of the data set. In which 20% will be used for testing purposes and 80% of our data will be used for training the models. In this dataset all portfolio holding stocks values belong to an independent variable and NAV net asset value comes under the dependent variable.

	Date	INFY	Baj	DMART	HDFC	ICICI	TCS	NAV
1	02-01-2018	514.85	1724.65	1160	936.175	309.7	1315.6	17.78
2	03-01-2018	510.65	1715.3	1184.55	926.325	315	1319.325	17.79
3	04-01-2018	507.7	1754	1203.1	929.95	314.7	1328.55	17.83
4	05-01-2018	506	1814.5	1243.4	931.8	312.9	1344.6	18
-	-	-	-	-	-	-	-	-
980	28-12-2021	1888	6906.65	4721.95	1460.8	735.8	3706.55	23.84
981	29-12-2021	1885.55	6911.65	4683.95	1453.85	735.7	3694.7	23.82
982	30-12-2021	1892.85	6871.1	4639.35	1461.5	735.7	3733.75	23.79
983	31-12-2021	1887.75	6977.3	4671.45	1479.4	740.15	3738.35	24.01

Figure 6.1: Final dataset ready to create models for analysing the NAV of the axis blue chip fund with its internal portfolio holding.

6.2 DATA PREPROCESSING MODELS

This model shows the training and testing dataset of the Infosys equity in the portfolios which give a brief idea about the regression models about the mutual fund. Here, while creating the model the dataset is divided into 2 parts. Where 20% of the dataset is used for testing and 80% is used for creating the regression models over portfolios prices and date rage. In the data pre-processing model we have used mainly two models: linear regression and decision tree regression model to analyze the portfolio's price value.
In this decision tree regression, models are created using the past dataset of the stock portfolio, by which future price of the particular stock is forecasted to train the previous dataset of the stock price. While creating the model we will get the original price predicted forecast price and the valid price which is the actual closing price of the particular portfolio.

6.2.1 MODEL 1 - STOCK OF INFY FOR LINEAR REGRESSION AND DECISION TREE REGRESSION

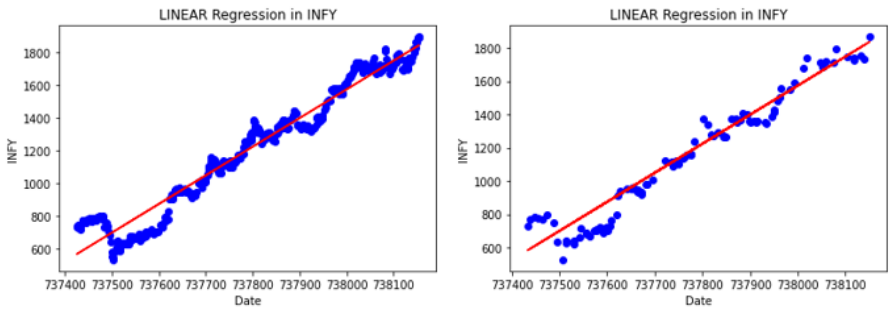

Figure 6.2: Linear regression over closing price of Infosys portfolio (b). testing of Linear regression model over closing price of Infosys portfolio

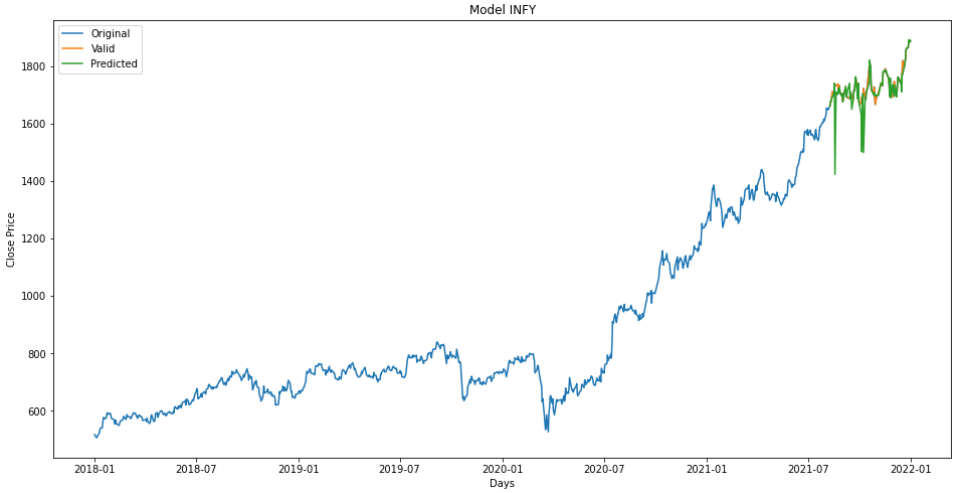

Figure 6.3: Decision tree regression on Infosys

In fig 6.2.(a) The linear regression model is used to analyse the past trend and volatility of the portfolio stock Infosys which holds 9.77% of the portfolio of invested amount. Whereas in 6.1.(b) Model is tested on this portfolio by selecting 20% of the data for training from the raw dataset.

In fig 6.3 The closing price value of the portfolio stock Infosys is predicted using the decision tree regression model. In this model, the Infosys portfolio's forecasted closed price values and actual closing prices for the previous 100 days are displayed.

6.2.2 MODEL 2 - STOCK OF TCS FOR LINEAR REGRESSION AND DECISION TREE REGRESSION

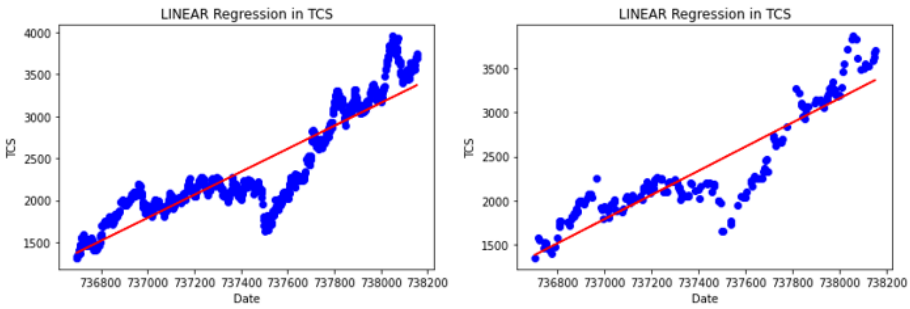

Figure 6.4: Linear regression over closing price of TCS portfolio (b). testing of Linear regression model over closing price of TCS portfolio

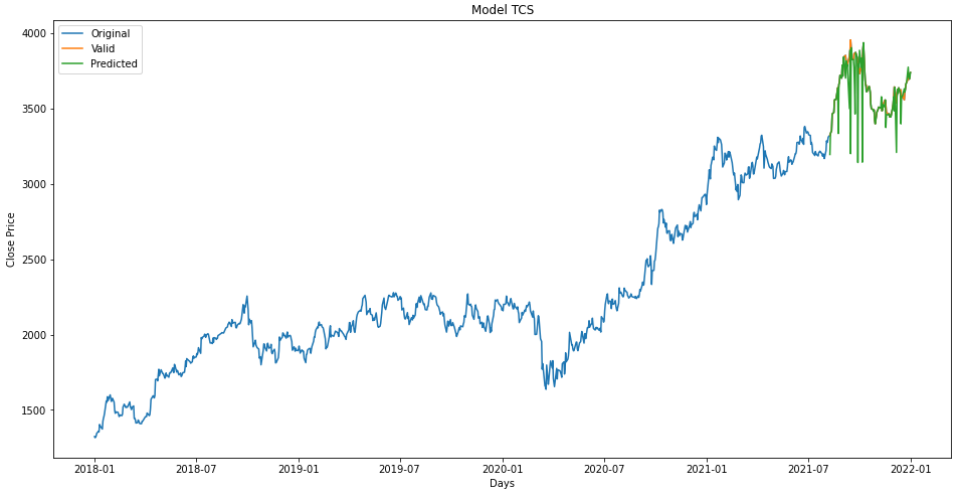

Figure 6.5: Decision tree regression on TCS

In fig 6.4.(a) The linear regression model is used to analyse the past trend and volatility of the portfolio stock Infosys which holds 7.00% of the portfolio of invested amount. Whereas in 6.4.(b) Model is tested on this portfolio by selecting 20% of the data for training from the raw dataset.

In fig 6.5 The closing price value of the portfolio stock TCS is predicted using the decision tree regression model. In this model, the TCS portfolio's forecasted closed price values and actual closing prices for the previous 100 days are displayed.

6.2.3 MODEL 3 - STOCK OF ICICI BANK FOR LINEAR REGRESSION AND DECISION TREE REGRESSION

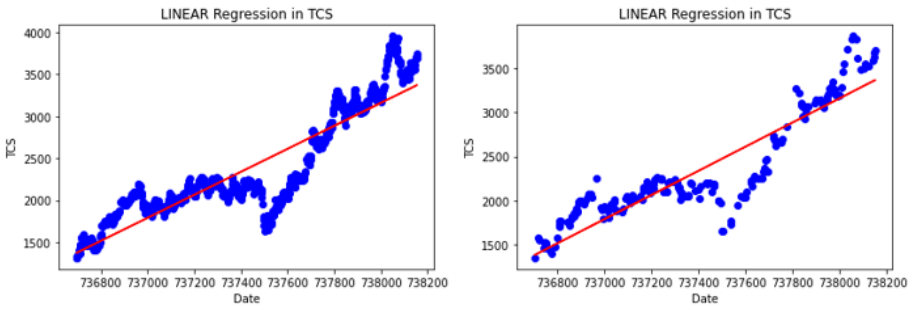

Figure 6.6: Linear regression over closing price of ICICI Bank portfolio (b). testing of Linear regression model over closing price of ICICI Bank portfolio

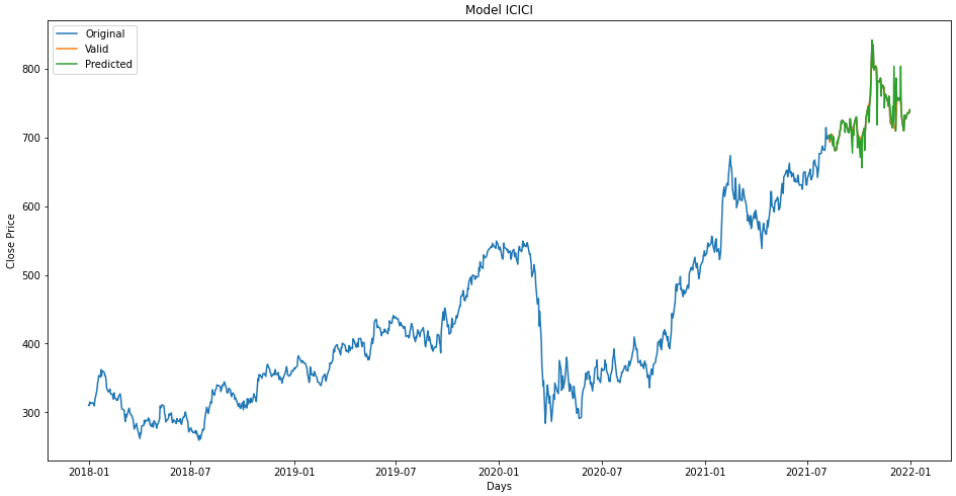

Figure 6.7: 6 Decision tree regression on ICICI

In fig 6.5.(a) The linear regression model is used to analyse the past trend and volatility of the portfolio stock ICICI BANK which holds 9.09% of the portfolio of invested amount. Whereas in 6.5.(b) Model is tested on this portfolio by selecting 20% of the data for training from the raw dataset. In fig 6.6 The closing price value of the portfolio stock ICICI BANK is predicted using the decision tree regression model. The ICICI BANK portfolio's actual closing price and predicted closed price values for the past 100 days are displayed in this model.

6.2.4 MODEL 4 - STOCK OF DMART FOR LINEAR REGRESSION AND DECISION TREE REGRESSION

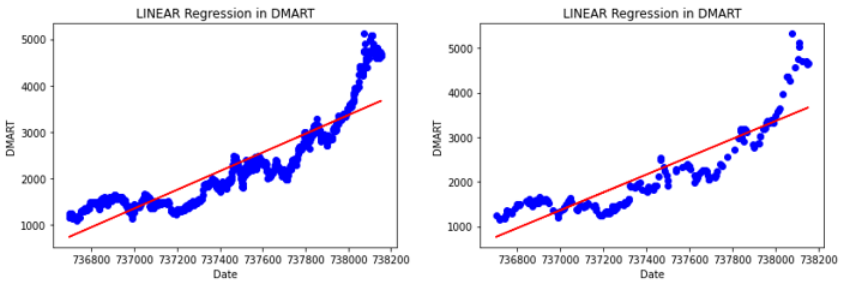

Figure 6.8: Linear regression over closing price of DMART portfolio (b). testing of Linear regression model over closing price of DMART portfolio

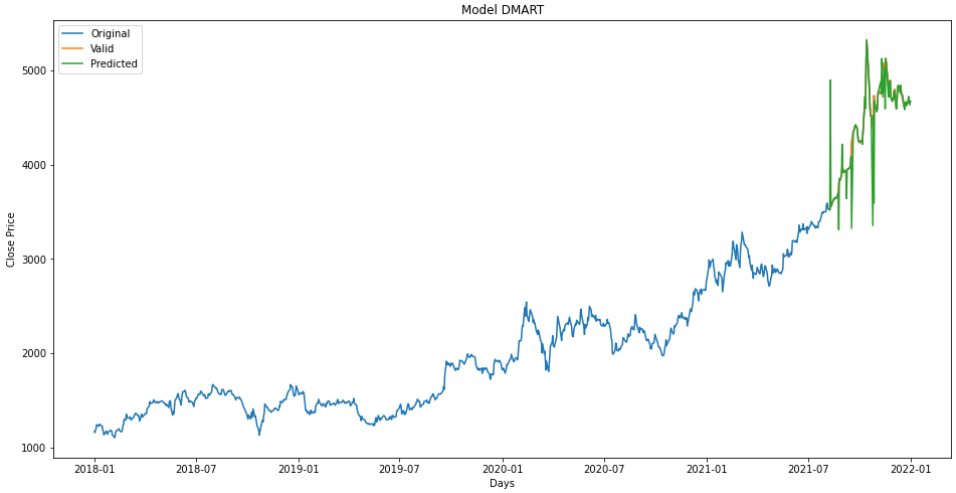

Figure 6.9: Decision tree regression on Avenue Supermart

In fig 6.8.(a) The linear regression model is used to analyse the past trend and volatility of the portfolio stock Avenue Supermart which holds 6.97% of the portfolio of invested amount. Whereas in 6.8.(b) Model is tested on this portfolio by selecting 20% of the data for training from the raw dataset.

In fig 6.9 The closing price value of the portfolio stock Avenue Supermart is predicted using the decision tree regression model. The Avenue Supermart portfolio's actual closing price and anticipated closed price values for the past 100 days are displayed in this model.

6.2.5 MODEL 5 - STOCK OF BAJAJ FINANCE FOR LINEAR REGRESSION AND DECISION TREE REGRESSION

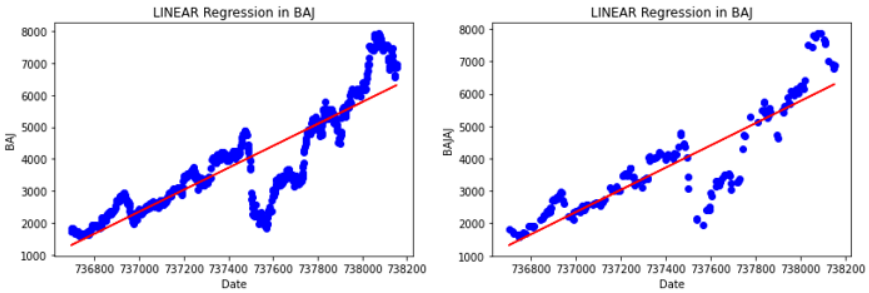

Figure 6.10: (a) Linear regression over closing price of BAJAJ portfolio (b)testing of Linear regression model over closing price of Infosys portfolio

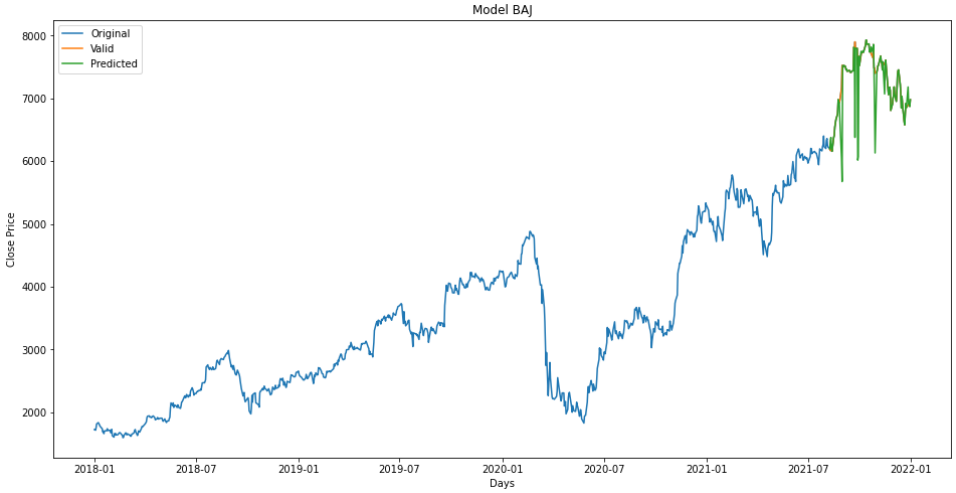

Figure 6.11: Decision tree regression on Bajaj Finance

In fig 6.10.(a) The linear regression model is used to analyse the past trend and volatility of the portfolio stock Bajaj Finance which holds 9.69% of the portfolio of invested amount. Whereas in 6.10.(b) Model is tested on this portfolio by selecting 20% of the data for training from the raw dataset.

In fig 6.11 The decision tree regression model is applied to predict the closing price value of the portfolio stock Bajaj Finance. In this model, last 100 days show the actual closing price and predicted closed price values of the Bajaj Finance portfolio.

6.2.6 MODEL 6 - STOCK OF HDFC BANK FOR LINEAR REGRESSION AND DECISION TREE REGRESSION

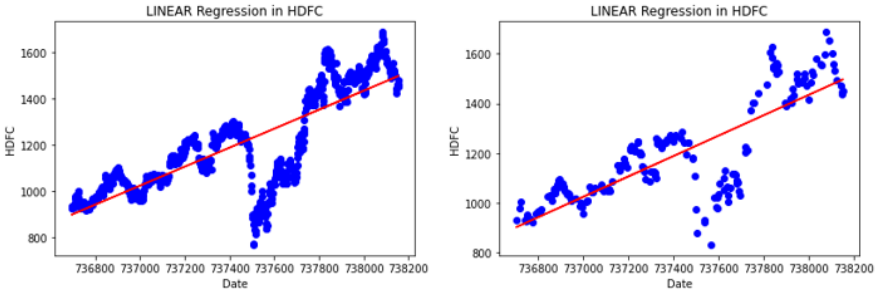

Figure 6.12: (a) Linear regression over closing price of HDFC Bank portfolio (b)testing of Linear regression model over closing price of HDFC Bank portfolio

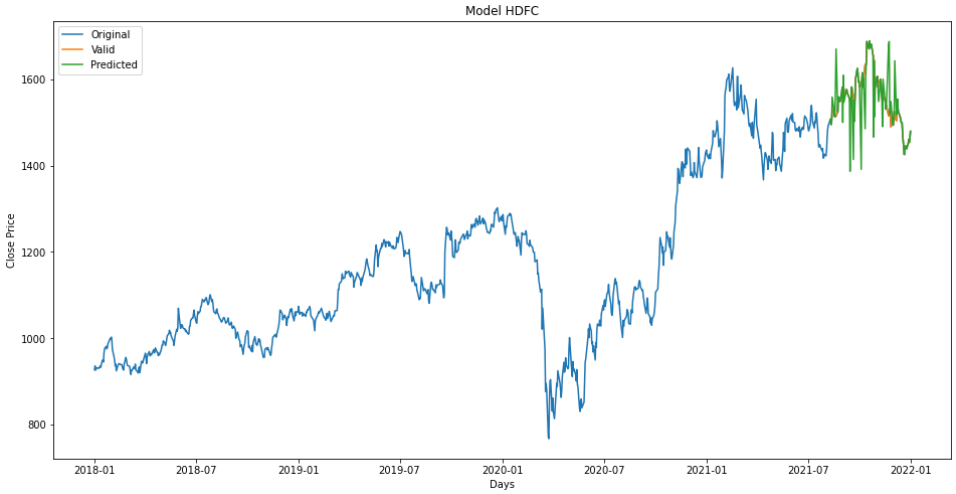

Figure 6.13: Decision tree regression on HDFC Bank

In fig 6.12.(a) The linear regression model is used to analyse the past trend and volatility of the portfolio stock HDFC BANK which holds 7.46% of the portfolio of invested amount. Whereas in 6.12.(b) Model is tested on this portfolio by selecting 20% of the data for training from the raw dataset.

In fig 6.13 The decision tree regression model is applied to predict the closing price value of the portfolio stock HDFC BANK. In this model, last 100 days show the actual closing price and predicted closed price values of

the HDFC BANK portfolio.

6.3 MULTIVARIATE LR TABLE FOR NAV

In the multivariate regression portfolio holding stocks are taken as independent variable whereas NAV net asset value is taken as dependent variable. In this table the result is shown in which random is selected to analyze the predicted nav and actual nav. R square or coefficient of determination is taken to analyze the model.

Table 6.1: Detailed portfolio overview according to its percentage of portfolios holding

Date	Actual NAV	Predicted NAV	R^2 (coefficient of determination)
12-12-2018	16.98	16.53	0.8662
26-06-2018	17.12	17.98	0.8662
12-11-2018	16.20	15.93	0.8662
13-08-2019	17.64	18.03	0.8662
23-09-2019	18.36	17.90	0.8662
01-01-2020	18.87	17.60	0.8662
15-06-2020	14.72	15.03	0.8662
16-06-2021	21.53	22.12	0.8662
31-12-2021	25.01	23.17	0.8662

6.4 FINAL RESULT

We have got 0.8662 coefficient of determination, which represents that our dataset fits with the accuracy of 86.62% for the model.

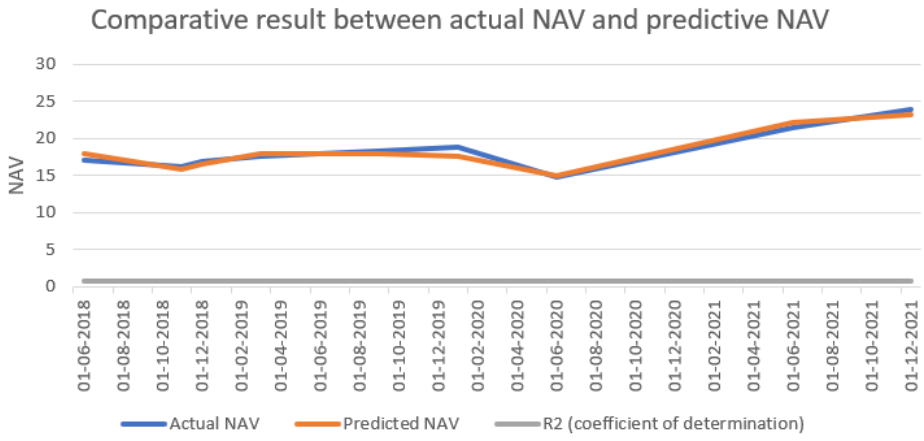

Figure 6.14: comparative result between actual NAV and predictive NAV

Chapter 7

CONCLUSION AND FUTURE SCOPE

Creating models of linear regression and decision tree regression is a notable and critical hard optimization to predict portfolio value. In this project, it presents the benefits of analysing the NAV value from internal portfolio holding using multivariate regression. for example, accomplishing a good return of investment, analysing the investment risk, and expanding the accuracy with a new approach of finding the net asset value of the mutual fund. The testing was done on the portfolio stocks dataset using both linear regression and decision tree regression which gives us good understanding of the stock prices volatility.

Science this proposed work is based on NAV control and prediction of the mutual fund performance. So, according to the analysis of the mutual fund performance the agent which controls the portfolio can change for high value on return on investment and can make out more profit.

There are various models and prediction methods out there though which portfolio price can be predicted with more accuracy and can take large portfolio dataset to improve the result of the portfolio.

Bibliography

[1] CM Anish and Babita Majhi. Prediction of mutual fund net asset value using low complexity feedback neural network. In *2016 IEEE international conference on current trends in advanced computing (ICCTAC)*, pages 1–5. IEEE, 2016.

[2] Xiaofei Chen, Shujun Ye, and Chao Huang. Cluster-based mutual fund classification and price prediction using machine learning for robo-advisors. *Computational Intelligence and Neuroscience*, 2021, 2021.

[3] Joseph F Hair. Multivariate data analysis: An overview. *International encyclopedia of statistical science*, pages 904–907, 2011.

[4] Sreenivasa Rao Jammalamadaka. Introduction to linear regression analysis, 2003.

[5] S Kavitha, S Varuna, and R Ramya. A comparative analysis on linear regression and support vector regression. In *2016 online international conference on green engineering and technologies (IC-GET)*, pages 1–5. IEEE, 2016.

[6] V Krishna. Nse stock market prediction using deep-learning models. *Procedia Comput. Sci.*, 132(Iccids):1351–1362, 2018.

[7] Kenneth D Lawrence, Gary Kleinman, and Sheila M Lawrence. Time series models to predict the net asset value (nav) of an asset allocation mutual fund vwelx. In *Handbook of Financial Econometrics and Statistics*, pages 2445–2460. Springer, 2015.

[8] Dastan Maulud and Adnan M Abdulazeez. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(4):140–147, 2020.

[9] Adil Moghar and Mhamed Hamiche. Stock market prediction using lstm recurrent neural network. *Procedia Computer Science*, 170:1168–1173, 2020.

[10] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[11] Aparna Nayak, MM Manohara Pai, and Radhika M Pai. Prediction models for indian stock market. *Procedia Computer Science*, 89:441–449, 2016.

[12] Engin Pekel. Estimation of soil moisture using decision tree regression. *Theoretical and Applied Climatology*, 139(3):1111–1119, 2020.

[13] Nuttrachai Promptsook and Kitsana Waiyamai. Thai equity mutual fund net asset value return prediction using internal factors. In *2021 2nd International Conference on Big Data Analytics and Practices (IBDAP)*, pages 119–123. IEEE, 2021.

[14] EL Houssainy A Rady, Haitham Fawzy, and Amal Mohamed Abdel Fattah. Time series forecasting using tree based methods.

[15] Santosh Singh Rathore and Sandeep Kumar. A decision tree regression based approach for the number of software faults prediction. *ACM SIGSOFT Software Engineering Notes*, 41(1):1–6, 2016.

[16] Murtaza Roondiwala, Harshal Patel, and Shraddha Varma. Predicting stock prices using lstm. *International Journal of Science and Research (IJSR)*, 6(4):1754–1756, 2017.

[17] George AF Seber and Alan J Lee. *Linear regression analysis*. John Wiley & Sons, 2012.

[18] Ya Su, Xinbo Gao, Xuelong Li, and Dacheng Tao. Multivariate multilinear regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(6):1560–1573, 2012.

[19] Geoffrey KF Tso and Kelvin KW Yau. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768, 2007.

[20] Gülden Kaya Uyanık and Neşe Güler. A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, 106:234–240, 2013.

[21] Min Xu, Pakorn Watanachaturaporn, Pramod K Varshney, and Manoj K Arora. Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97(3):322–336, 2005.

[22] Kelly H Zou, Kemal Tuncali, and Stuart G Silverman. Correlation and simple linear regression. *Radiology*, 227(3):617–628, 2003.