International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT22855

Stock Prediction Using Machine Learning Algorithms

Siddharth Meghwal[1], Irfan Khan[2]

[1]Research (MTech) Scholar (CSE), Shekhawati Institute of Engineering and Technology, Sikar, Rajasthan, India

[2]Assistant Professor (CSE), Shekhawati Institute of Engineering and Technology, Sikar, Rajasthan, India

Article Info

Publication Issue :
Volume 8, Issue 5
September-October-2022

Page Number : 18-30

Article History
Accepted: 01 Sep 2022
Published: 09 Sep 2022

ABSTRACT

In the recent times, the stock markets have emerged as one of the top investment destinations for individual and retail investors due to the lure of huge profits that are possible with stock investments compared to more traditional and conservative forms of investments such as bank deposits, real estate and gold. The stock markets unlike other forms of investment are highly dynamic due to the various variables involved in stock price determination and are complex to understand for a common investor. Individual and small-time investors have to generate a portfolio of common stocks to reduce the overall risk and generate reasonable returns on their investment. This phenomenon has given way too many individual and retail investors incurring huge losses because their decisions are based on speculation and not on sound technical grounds. While there are financial advisory firms and online tools where individual investors can get professional stock investment advice, the reliability of such investment advice in the recent past has been inconsistent and not meeting the rigor of quantitative and rational stock selection process. Many of such stock analysts and the tools mostly rely on short term technical indicators and are biased by the speculation in the market leading to huge variances in their predictions and leading to huge losses for individual investors. While the use of Artificial Intelligence (AI) and Machine Learning (ML) techniques is widely adopted in the financial domain, integration of AI/ML techniques with fundamental variables and long-term value investing is a lacking in this domain. Some of the stock portfolio tools available in the market use AI/ML techniques but are mostly built using technical indicators which makes them only suitable for general trend predictions, intraday trading and not suitable for long term value investing due to wide variances and reliability issues. The availability of a Financial Decision Support System which can help stock investors with reliable and accurate information for selecting stocks and creating an automated portfolio with detailed quantitative analysis is lacking. A Financial Decision Support System (DSS) that can establish a relationship between the fundamental financial variables and the stock prices that can VII automatically create a portfolio of premium stocks shall be of great utility to the individual investment community. As part of this thesis, the researcher has designed and developed a Financial Decision Support System (DSS) for selecting stocks and automatically creating portfolios with minimal inputs from the individual investors. The Financial DSS is based on a System Architecture combining the advantages of Artificial Intelligence (AI), Machine learning (ML) and Mathematical models. The design and development of the Financial DSS is based on the philosophy to combine various independent models and not rely on a single stock price model to increase the accuracy and reliability of the stock selections and increase the overall Return on Investment (ROI) of the stock portfolio. The Machine learning models are used to establish the relationship between fundamental financial variables and the price of the stock, a mathematical model is developed to calculate the intrinsic value of the stock taking in to account the full lifecycle of the stock which involves various phases and a comprehensive model to analyze the financial health of the stocks. The AI/ML stock models are independently trained using historical financial data and integrated with the overall Financial DSS. Finally, the Financial DSS tool with a graphical user interface is built integrating all the three models which shall be able to run on a general-purpose desktop or laptop. To reliably validate the Financial DSS, it has been subjected to wide variety of stocks in terms of market capitalization and industry segments. The Financial DSS is validated for its short term and long-term Return on Investment (ROI) using both historical and current real-time financial data. The researcher has reported that the accuracy of the AI/ML stock price models is greater than 90% and the overall ROI of the stock portfolios created by the Financial DSS is 61% for long term investments and 11.74% for short term investments. This system has the potential to help millions of individual investors who can make their financial decisions on stocks using this system for a fraction of cost paid to corporate financial consultants and value eventually may contribute to a more efficient financial system.

Keywords: Decision Support Systems (DSS), Stock Prediction, Stock Portfolio, Artificial Intelligence (AI), Machine Learning (ML), Fundamental Analysis, Intrinsic Value, Stock Health Analysis, Artificial Neural Network (ANN), Decision Trees, Ensemble Algorithms

I. INTRODUCTION

In today's financial sector, there are many computerized systems deployed to perform online banking, payment transactions, online trading, forex and many such areas (Goldstein, Jiang, & Karolyi, 2019). Many of these systems are using Artificial Intelligence and Machine learning to detect fraudulent banking transactions, predict the stock markets and we also find many algorithmic based trading systems which are based on complex mathematical models. These systems are having a huge impact on the way the financial sector is operating and gives is some clues on the future of this sector (Gai, Qiu, & Sun, 2018).

Machine Learning is a branch of Artificial Intelligence which is having a great impact on many fields including the financial domain. Machine learning systems use past data to build insights and knowledge to solve complex problems (Goodarzi, 2010). Decision support systems is also a branch of Artificial intelligence where these systems have a specific job of guiding the user on complex decisions and in many cases, completely automate the decision making process (Palma-dos-Reis & Zahedi, 1999). These systems are quite useful in domains such as financial sector where there is voluminous data to analyze quantitatively and arrive at quick decisions related to investments, trading, payments and other related areas.

Based on the literature review we believe that the financial DSS proposed in this study is first of its kind which combines three independent models (i.e. AI//ML stock price model, Intrinsic Value mathematical model and Comprehensive stock health analysis model) to analyze and select top performing stocks and create automated stock portfolios with minimal inputs from stock investors. The Financial DSS in this study was tested and validated on a wide range of stocks ranging across industry segments such as large capitalizations stocks, mid-size capitalization stocks and small capitalization stocks.

1.1 Objectives of Stock Investment

While there are many financial avenues for potential investors to put their money like the gold, commodities, savings accounts, fixed deposits, post office schemes, real estate and other modes of investment, the unique feature of stock investment is to earn higher returns compared to all other forms of investments. Historically, the stock market returns have been higher than any other forms of investment due to the compounding growth of stocks and exponential growth of company profits. The key objectives of potential investments in stock markets are (Ballestero, Bravo, Pérez- Gladish, Arenas-Parra, & Plà-Santamaria, 2012):

- Earning of premium returns compared to other forms of wealth creation, i.e.High reward for people who are willing to take measured risks involved with stock markets.
- Mitigation of Un-systematic risk and profit maximization.
- Risk reduction by qualitative and Quantitative research.
- Higher liquidity than other forms of financial instruments.
- Beating the inflation by safe margin which is difficult with financial instruments like savings and fixed deposits.
- Partial ownership of a firm which is not possible by other financial instruments and potential for long term financial security via the assurance of dividends from the stocks owned by the investors.
- Effective trade-off and balancing between risk and reward which is not possible using other financial instruments.

1.2 Stock Market Return

The returns on the stock investments can be categorized in to four forms:

- ROI via the stock price changes – In this mode of ROI, the profits or losses can be calculated based on the stock price increase or decrease. This is direct way of wealth accumulation based on the performance of the company.
- ROI via the dividends distribute by the companies – In this mode of ROI, the return calculations take it to account the dividends distributed by companies. Dividend distribution is a unique policy of each company and depends on the business cycle that the company operates (Hayes, 2021f).

- ROI via stock splits and bonus stocks – Occassionally companies announce stock splits based on the performance of the company and also for introducing liquidity in their stocks. In this mode, for every share held by the investor, they get an additional share and the price of the split share decided by the collective market forces (Will Kenton, 2020).

- ROI via stock buybacks – When companies have large cash deposits with them and the management decides to pass this cash holding back to the stock holders, they often choose stock buyback policy as it is efficient way to return the profits accumulated by the company back to their stock holders. In this mode, the buy back stock price is fixed at a higher price compared to the open market prevalent prices and thus existing stock holders get the benefit of higher stock prices (Cory Janssen, 2021).

II. Valuation Methods Chosen For The Study

There are some important hypothetical equity prediction models that can be used for valuation. These models require:

a. Estimation of the future payoff- These methods are modelled, ignoring the anticipated future payoffs.
b. No Estimation of the future payoff – The probable problems in predicting the future payout can be avoided.
c. Risk Based Models- The risks caused by the variables are considered in these models.

According to the Drakopoulou et al. (2015), three valuation models are found suitable for the prediction of Sensex stocks. Following are the valuation models used by the proposed approach:

a. Relative Valuation Models

Relative valuation model -P/E Model: Market price divided by earnings and share gives the price to earnings ratio. Price to earnings coefficient is a function, company's beta, growth rate, payout ratio and earnings
per share.

$$\boldsymbol{P/E\ Ratio} = (\boldsymbol{Market\ Price}(\boldsymbol{MP}))/(\boldsymbol{Earnings\ Per\ Share}(\boldsymbol{EPS})) \quad (3.1)$$

Estimating Intrinsic Value using Price to Earnings Model

Regressing the variables such as expected growth in EPS, beta coefficient, and dividend payout ratio studied the relationship between the variables to predict the yearly price earnings ratio. P/E model looks at the bond of ratio of the P/E of a stock to the P/E of Sensex while predicting. In essence, the predicted value of a stock is calculated based on how comparable Sensex stocks are valued, based on P/E value.

Using the above model, P/E of a Stock is calculated by (3.2) .The variables used in the models are defined in section 1.8.2

$$\mathrm{P/E} = (\ \boldsymbol{Market\ Price}\ )/\boldsymbol{EPS} \quad (3.2)$$

b. Risk Based Models

Many stock valuing models are developed based on P/B value for Sensex stocks. The relationship of the P/B of a stock to the P/B of the Sensex stocks is studied in the P/B model to predict the intrinsic value. Based on how stocks are valued using the P/B ratio, the intrinsic value is calculated.

Estimating intrinsic value using P/B model

Regression analysis studied the relationship between expected growths in EPS, beta co efficient, return on equity and P/B ratio. The intrinsic value, is then calculated by (3.3) and (3.4)

$$\boldsymbol{Intrinsic\ value} = \boldsymbol{Predicted\ P/B} \times \boldsymbol{BVPS} \quad (3.3)$$

where,

$$\boldsymbol{Predicted\ P\ B} = \boldsymbol{\beta 0} + \boldsymbol{\beta 1\ BETA} + \boldsymbol{\beta 2\ EPSg} + \boldsymbol{\beta 4\ DPR} + \boldsymbol{\varepsilon i} \quad (3.4)$$

III. Descriptive Statistics for variables used in the study

Descriptive statistics, mean, maximum, minimum and standard deviation of the variables, market price, EPS Growth, beta, EPS, DPR, BVPS, ROE, stock index returns, of each Sensex stock for the chosen three fundamental models are calculated. Descriptive statistics of Rf, Rm, Repo rate and the RMSE for the equity models are also calculated.

Table 3.2 shows the descriptive statistics of market price (MP) of Sensex stocks from 2008-09 to 2017-18. The price has an extreme value of Rs.8863.15 for M&M and least value of Rs.33.65 for Tata Motors. Standard deviation of market price at a lowest cost of Rs. 27.64 for NTPC directs the low risk in buying the stock. Standard Deviation, for most of the large market capitalized Sensex stocks in Indian Stock Market is very high. Bajaj Finance and M&M shows high standard deviation which is mainly due to the fact that stock prices fluctuate according to the interest rates, inflation rates and other market information.

Standard Deviation is very high for most of the large market capitalized Sensex stocks in Indian Stock Market. Bajaj Finance and M&M shows high standard deviation which is mainly due to the fact that stock prices fluctuate according to the interest rates, inflation rates and other market information.

Table 3.1

Market Price of Sensex Stocks in India from 2008-09 to 2017-18

Descriptive Statistics (in Rs.)

Stocks	Mean	Max	Min	SD	N
Asian Paints	1792.36	4914.40	546.50	1406.15	10
Axis Bank	890.00	1460.45	414.50	439.59	10
Bajaj Auto	1962.03	2807.50	618.45	642.59	10
Bajaj Finance	1881.50	6929.20	68.60	2100.25	10
Bharathi Airtel	373.63	625.80	291.80	94.77	10
Coal India	251.80	362.40	283.50	135.54	10
HCL Tech	722.69	1390.50	101.75	373.19	10
HDFC	1294.97	2712.85	673.60	624.85	10
HDFC Bank	1256.41	2342.95	520.05	618.89	10
Hero Motors	2282.42	3545.50	1070.15	797.07	10
Hindustan Unilever	622.92	1335.90	238.20	364.29	10
ICICI Bank	668.35	1245.45	236.55	412.36	10
IndusInd	674.88	1795.60	32.30	576.29	10
ITC	270.82	352.95	181.45	59.78	10
Kotak Bank	737.56	1313.15	282.95	294.78	10
LT	1372.14	1719.00	672.65	305.07	10
Maruti	858.74	1284.70	383.20	301.41	10
MM	3032.08	8863.15	775.10	2622.02	10
NTPC	161.67	207.00	119.95	27.64	10
ONGC	395.03	1098.50	177.80	300.84	10
Power Grid	129.89	197.20	95.65	38.14	10
Reliance	1016.91	1523.20	748.25	246.46	10
SBI	1300.35	2767.90	194.30	990.75	10
Sun Pharma	832.98	1789.60	441.80	401.98	10
Tata DVR	485.65	1247.50	180.30	313.49	10
Tata Motors	184.06	328.14	33.65	92.99	10
Tata Steel	432.56	632.65	206.00	146.02	10
TCS	1771.47	2849.30	540.00	825.23	10
Vedanta	223.25	471.10	89.90	111.66	10
Yes	535.91	1549.10	49.90	433.06	10

* Results computed using R.4.0.2

IV. Multiple Linear Regression For Refining The Equity Valuation Models

The revised P/E and P/B regression model is widely used to predict the intrinsic value of the stocks based on the variables EPS, EPS Growth, repo, BVPS and DPR. Linear regression is an algorithm in statistics. It is an approach to model the relationship between a response variable and one or more predictor variables. A linear relationship exists between independent variables and a single dependent or output variable.

Therefore, output variable, is the linear combination of the input variables. There exist two types of linear regression:

1. Simple linear regression, when there is only one independent variable It is a model for understanding relationships between a single input and output variable.
2. Multiple linear regression, when there are multiple independent variables.

However, multiple linear regression goes one step ahead further and adds more than one independent variable. For example, if additional variables like DPR, EPS influence the market price, then the equation (4.7) becomes,

*market price= β0 + β1 * repo rate +β2*DPR+β3*EPS+ Ɛ.* ***(4.1)***

There are mainly three domains where the Regression-based machine learning algorithms Pai et al. (2021) can be applied.

- Quantifying the relations between the variables: Correlation identifies the strength and direction of the association between two variables. These relationships are quantified using multiple regression.
- Prediction: Regressions' ability to quantify the relationships/association can be turned into solving prediction problems.
- Forecasting: Estimation of some variables of interest at some specified future period is another domain where multiple regressions display outstanding performance. Prediction is similar to forecasting, but the estimations occur for a prolonged period when compared to the ones done for a specific date in forecasting Multiple linear regression works by changing the parameter values to reduce the cost, which is the degree of error between the predicted value and the values of the training set.

A calculating metric known as cost function determines how well a hyper plane represents a training set. The cost function represents the degree of error between the hyper plane's values and those of the training set. Different formulas can calculate the cost. The one that the regression model uses is called the **mean squared error (MSE).** This cost function determines how well a hyper plane represents the data. The regression model can fine-tune the parameters' values to find the best fit hyper plane that defines the data the gradient descent function is used.

The gradient descent function helps the machine learning models to refine the parameters so that the optimal values for the coefficient of the hyper planes can be identified. The algorithms use gradient descent in order to converge upon a parameter value that produces the lowest error. The models can later use these values for future predictions with a new set of data; in this study, multivariate gradient descent function to be used as we are dealing with a multilinear regression model.

After applying the gradient descent, the model will converge on the coefficients b0, b1, b2….bn and finalize the parameter values and create the equation of the optimal hyper plane. Finally, plugging in the actual values to this optimized hyper plane equation will give the estimated value of the model.

The empirical analysis of the three equity valuation models used in the study had errors. So, in order to revise and refine the model, the multiple linear regression machine learning algorithm is used. The various steps used involved in creating the machine learning model are:

1. Data Preprocessing
2. Creating Train and Test Datasets
3. Creating and training the linear regressor, Boosted Decision Tree Regression, Decision Forest tree regression.
4. Testing the Linear regressor, Boosted Decision Tree Regression, Decision Forest tree regression.

Data Preprocessing: All the variables used in the data set are numerical. Identifying, eliminating inexact records from the dataset, recognizing incomplete, undependable, imprecise, non-relevant parts of the data is performed to clean the data. The inconsistencies identified or removed may have been caused by user entry mistakes,
errors in the storage.

Creating the Model: The multiple linear regressors is created for revised P/E, P/B and CAPM. The array of the independent variables used to create the model for the predictor or the dependent variables. For example, for the dependent variable, market price, the independent variables to the regressor are:

Preparing the Training and the Test data: The whole dataset is divided into two training sets and the test set. Then, ratio of 80:20 is selected, which divides 80% of the data into the train data and the rest 20% to the test set.

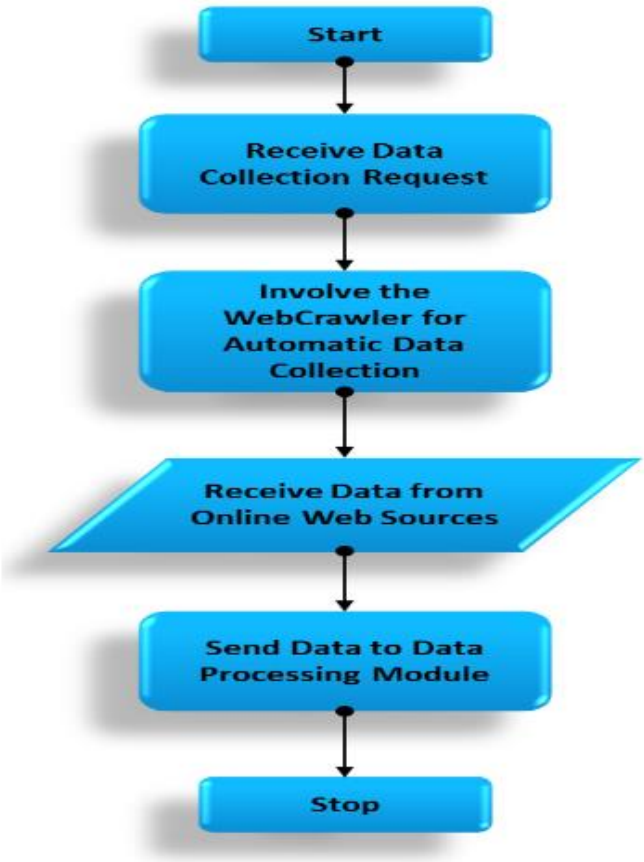
Training the Linear Regression: A linear regression model essentially adds a coefficient to each input variable, determining its importance. The value of input variables is then multiplied with the equivalent coefficient, and the bias (intercept) term is added to summation. That is essentially the predicted value.

Testing the Linear regression: For testing the regressor, use model to predict the test data. The performance measures adapted for the testing of the models will give the accuracy of the model. RMSE, MSE, R-squared metric are the commonly used metric used to evaluate the performance of a model.

4.1 Select Dataset

The Data Collection module is the key module of the Financial DSS that is responsible for all the data collection requirements both from external world (Online Web Sources) and locally on the desktop / laptop environment. The major functionalities of this software module are:

- Receiving data collection and initiation request from the data processing software module
- Implementation of Web Crawler functionality for automatic data collection from online web sources
- Management of data sources and formats
- Transfer of online data received via Web Crawler to the Data Processing software module



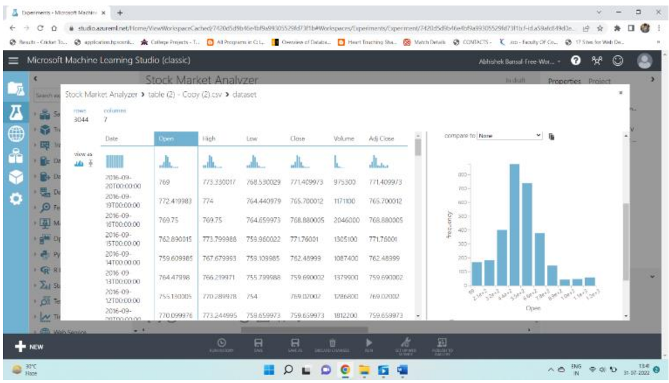


Figure 4.1 Dataset in .csv format implemented on Azure

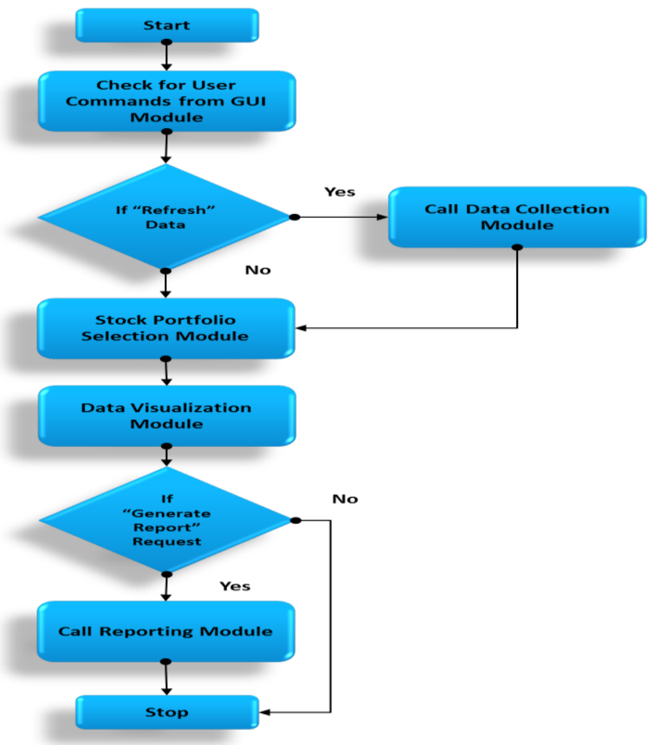
4.2 Information Preprocessing

It is a cycle of evacuating all the loud and missing information from the informational collection.
At the point when information is given as information, it is important to preprocess the data. Text preprocessing is the cycle of getting ready and cleaning the information of dataset for characterization. It assists with decreasing the clamor in the content, improve the exhibition of the classifier and accelerate the characterization cycle. Preprocessing information has following 3 stages

•Tokenization: It is a sort of pre-preparing where running content is divided into words or sentences. Before any genuine content preparation is to be finished, the text should be sectioned into phonetic units, for example, words, accentuation, numbers, alpha-numeric, and so on this cycle is called tokenization. Tokenization, when applied to records, is the cycle of subbing a touchy information component with a non-sensitive same alluded as a token that has no outward or exploitable significance or worth. A record is considered as a string, and afterward parceled into a rundown of tokens. Stop words, for example, "the", "a", "and", and so forth are every now and again happening; in this manner the inconsequential words should be taken out.

•Stop word evacuation: In figuring, stop words are words which are sifted through previously or after preparing of normal language information (text). Stop words normally allude to the most widely recognized words in a language. The most widely recognized words are in text reports are relational words, articles, and favorable to things and so forth, that doesn't give the significance of the records. These words as treated as stop words. Model for stop words: the, in, a, an, with, and so forth [7] Hence it is vital to eliminate those words which show up excessively often that give no data to the errand. Stop words are eliminated to save both existences. Stop words are a fundamental piece of data recovery measure. The expulsion of stop words increments execution and indexed lists. The stop words need to be eliminated for an explanation since they give no particular data for order reason.

• Stemming: It is the cycle for diminishing inferred words to their stem or root structure for example is primarily eliminates different additions therefore in the decrease of a number of words. For Example, the words client, clients, utilized, utilizing all can be decreased to "USE". This will diminish the necessary time-space



4.3. Train and assemble the AI model

In this progression, the dataset is isolated into two sections: preparing dataset and testing dataset. Preparing dataset contains 70% and the testing dataset contains 30% which are chosen haphazardly.

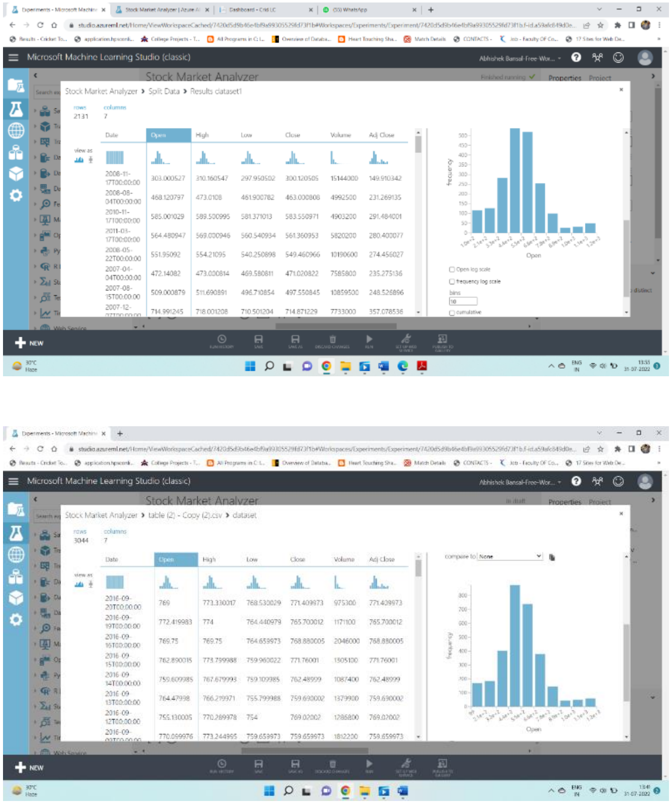


Figure 4.2 Splitting of Dataset with 30 :70 for training and testing purpose in .csv format implemented on Azure

4.4 Highlight Selection

In the wake of preprocessing and Transformation the significant advance of text order is highlight determination. The fundamental thought of highlight choice is to choose a subset of highlights from the first record. The information contains numerous highlights, yet all the highlights may not be important so the element choice is utilized in order to kill the unessential highlights from the information absent a lot of loss of the data. Highlight the choice is otherwise called ascribes determination or variable selection[13]. It is performed by keeping the words with most elevated score according to the foreordained proportion of the significance of the word.

4.5 Arrangement

The archives can be arranged by administered what're more, solo strategies. At the point when the class mark of each the report is realized that is managed when the class name of the archive is not known that is called unaided.

4.6 Execution Measure

This is the last advance of information text characterization. This is tentatively done, as opposed to systematically. In this progression measures the exhibition. Numerous measures have been utilized like exactness and review.

4.7 Point of the Proposed System

This module also processes the request for data visualization for the users to clearly evaluate the stock selection process. The major functionalities of this module are:

- Receives stock selection data from the data processing unit after analysis by the Stock selection software module.
- Provides the reasons for the stock selection.
- Processes the data visualization requests and provide appropriate data for GUI software module.

The following reasoning features are provided for each selected stock:

- Name of the stock.
- Current real-time stock price.
- Intrinsic Value of the Stock.
- Type of stock labelled as “Under Valued”, “Over Valued” or “Fairly” priced.
- Growth rate of the company.
- Profitability criteria of the stock in terms of “Low Margin Business” or “High Margin Business”.
- Cash flow and short term liabilities situation of the company.
- Stock Price growth trajectory in the last one year.

- Net profit growth trajectory in the last one year.
- Revenue growth trajectory in the last one year.
- Comparison analysis of Industry P/E vs Stock's P/E.
- Competitive Advantage of the stock.

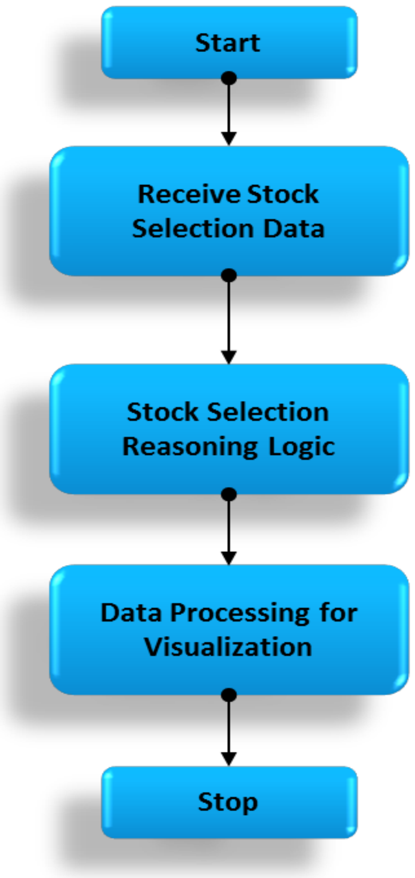


Figure 4.3 System Implementation Flow

In the above fig 4.3, crude information is gathered from this present reality. The gathered information is being prepared and the missing dataset is cleaned. At that point, the information model is prepared and broke down. After the investigation cycle, calculations are applied to the prepared model. After the preparation cycle, a report will be pictured and information items will be shaped. Through the representation, report choices will be made and the legitimate forecast will be examined however score esteems.

The import information alludes to the information kind of every segment dependent on the qualities it contains and places the information into Azure Machine Learning Studio workspace. The yield of import information is a dataset that can be utilized in any trial. Information procurement in AI comprises of two things: information and model. When gathering the information it must have enough highlights with the goal that it can assist with foreseeing the infection and effectively train the learning model. In a pre-preparing step, the information is to be cleaned and disentangled. By pre-preparing of the information, we can get all the more effectively make important highlights from information. After pre-preparing, chose calculations are applied on the AI model, used to gauge the exactness of the expectations. Scoring is the way toward producing esteems or scores dependent on a prepared AI model. The qualities or scores that are delivered can speak to ailment forecasts of future qualities.

4.8 Implemented Algorithms

4.8.1 Decision Forest Regression:

Two-Class Decision Regression is utilized in Azure Machine Learning Studio to build up an AI model that depends on a directed gathering learning calculation called choice wilderness. This calculation turned around to an undeveloped more tasteful. At that point a specific model is prepared by utilizing a Train model or Tune Model Hyper Parameters on a marked preparing datasets. Later the prepared model is utilized to make out forecasts. The Two-Class Decision regression module restores an undeveloped classifier. You at that point train this model on a marked preparing informational collection, by utilizing Train Model or Tune Model Hyperparameters. The prepared model would then be able to be utilized to make forecasts.

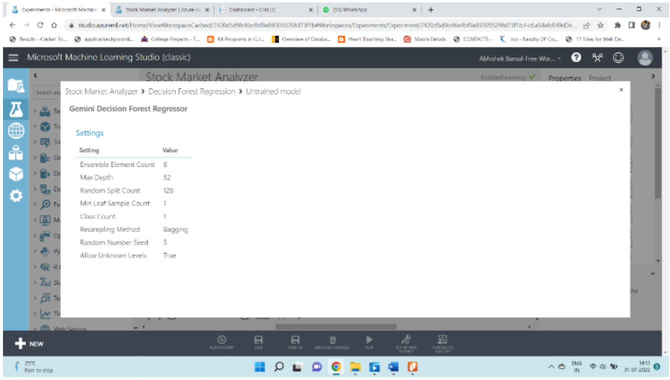


Fig.4.4 Two Class Decision Regression in Stock Analysis untrained model

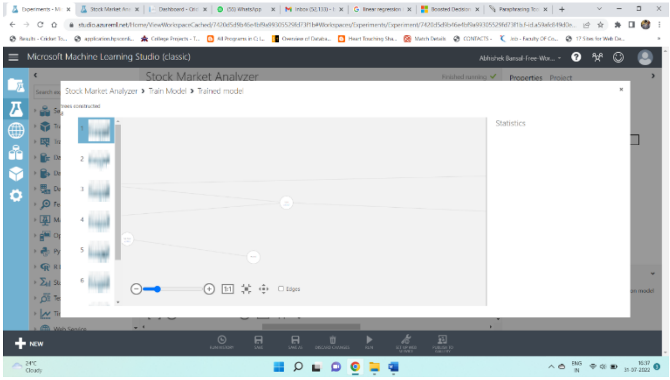


Fig.4.5 Two Class Decision Regression in Stock Analysis trained model

4.8.2 Linear Regression:

A standard statistical approach, linear regression has been implemented in machine learning and improved with many new ways for fitting the line and assessing error. Linear regression is a method that has been used for a long time. In its most basic form, regression is defined as the prediction of a numerical goal. When you need a straightforward model for a fundamentally predictive endeavour, linear regression is still a solid option to look at. Additionally, linear regression has a propensity to function well on data sets that are high-dimensional, sparse, and lack complexity.

In addition to linear regression, Azure Machine Learning is capable of supporting a wide number of alternative regression models. On the other hand, "regression" is a phrase that can have a broad meaning, and not all forms of regression that are offered by other tools are supported here. A single independent variable and a single dependent variable are required to solve the traditional regression issue. This phenomenon is known as simple regression. This component provides support for straightforward regression. The concept of multiple linear regression refers to a statistical technique in which two or more independent variables are used to analyse a single dependent variable. One name for the kind of problems known as multivariate linear regression is "problems with numerous inputs used to predict a single numerical output." The Linear Regression component, along with the majority of the other regression components, is capable of providing a solution to these difficulties. The process of making predictions about several dependent variables using just one model is referred to as multi-label regression. One use of multi-label logistic regression, for instance, allows for a sample to be allocated to numerous distinct labels. (The challenge of predicting many levels inside a single class variable is distinct from this one.)

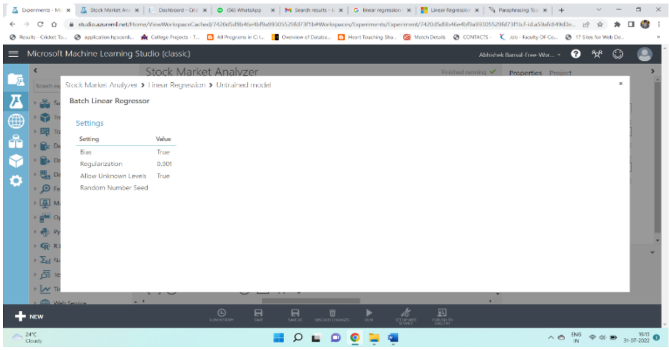


Fig.4.7 Linear Regression in Stock Analysis untrained model

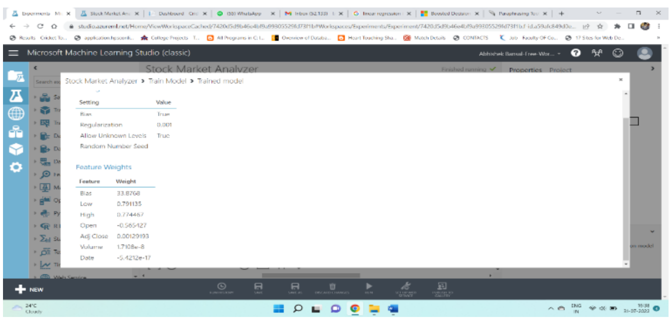


Fig.4.8 Linear Regression in Stock Analysis trained model

4.8.3 Boosted Decision Tree Regression:

Creating ensemble models may be accomplished using a variety of tried-and-true strategies, such as bagging, random forests, and boosting, to name just a few. Boosted decision trees in Azure Machine Learning make advantage of an effective version of the MART gradient boosting technique. A strategy for solving regression issues using machine learning is called gradient boosting. It develops each regression tree in a step-by-step manner, utilising a specified loss function to quantify the error in each step and then correcting for it in the subsequent step as it grows the tree. As a result, the prediction model is really an ensemble of a number of less accurate prediction models.

Boosting is a method that is used to solve regression issues by first building a succession of trees in a step-by-step manner and then selecting the best tree by the application of an arbitrary differentiable loss function.

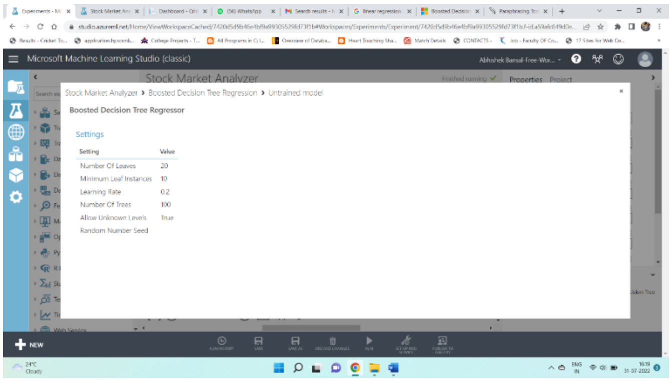


Fig.4.9 Boosted Decision Tree Regression in Stock Analysis untrained model

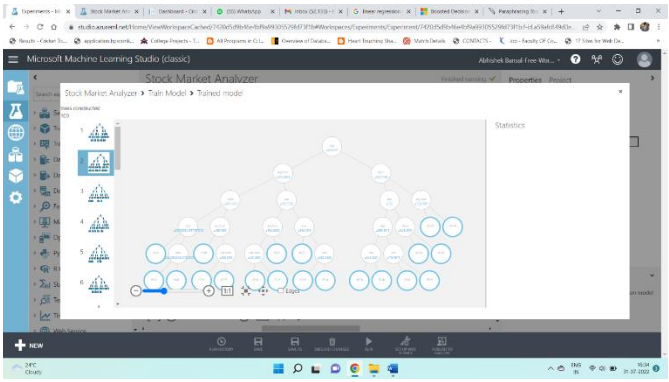


Fig.4.10 Boosted Decision Tree Regression in Stock Analysis trained model

V. CONCLUSION

- Linear Regression is the base of all regression models and provides results that are near to satisfaction but with a greater deviation from original data.
- Boosted Decision Tree Regression gives a distorted line which is not suitable for our case in predicting Stock Price which has to be very near to the actual solution.
- Decision Forest Tree Regression model stands out to be the best model for Stock Price with the least error.

VI. REFERENCES

[1]. Abada, C., Thoreb, S. A., & Laffargaa, J. (2004). Fundamental Analysis of Stocks by. MANAGERIAL AND DECISION ECONOMICS, 241(June), 231–241. https://doi.org/10.1002/mde.1145

[2]. Abbas, G., & Wang, S. (2020). Does macroeconomic uncertainty really matter in predicting stock market behavior? A comparative study on China and USA. China Finance Review International, 10(4), 393–427. https://doi.org/10.1108/CFRI-06-2019-0077

[3]. Adair, M. O. (2018). Behavioral Finance: Understanding How Biases Impact Decisions. Retrieved from https://www.cnr.com/insights/article/white-paper behavioral-finance-2018.ht

[4]. Aditya. (2020). A Simple Explanation of K-Means Clustering. Retrieved from https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k means-clustering/

[5]. Shi, S., Liu, W., & Jin, M. (2012, November). Stock price forecasting using a hybrid ARMA and BP neural network and Markov model. In 2012 IEEE 14th International Conference on

Communication Technology (pp. 981-985). IEEE.

[6]. Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation (pp. 106-112). IEEE.

[7]. Frimpong, J. M., & Oteng-Abayie, E. F. (2006). Modelling and forecasting volatility of returns on the Ghana stock exchange using GARCH models

[8]. Guresen, E., Kayakutlu, G., &Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 38(8), 10389-10397.

[9]. Perwej, Y., & Perwej, A. (2012). Prediction of the Bombay Stock Exchange (BSE) market returns using artificial neural network and genetic algorithm.

[10]. Dong, G., Fataliyev, K., & Wang, L. (2013, December). One-step and multi-step ahead stock prediction using backpropagation neural networks. In 2013 9th International Conference on Information, Communications & Signal Processing (pp. 1-5). IEEE.

[11]. Naeini, M. P., Taremian, H., & Hashemi, H. B. (2010, October). Stock market value prediction using neural networks. In 2010 international conference on Computer Information Systems and Industrial Management Applications (CISIM) (pp. 132-136). IEEE.

[12]. Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. Journal of Economics, Finance and Administrative Science, 21(41), 89-93

[13]. Wei, L. Y., Chen, T. L., & Ho, T. H. (2011). A hybrid model based on adaptive network-based fuzzy inference system to forecast Taiwan stock market. Expert Systems with Applications, 38(11), 13625-13631.

[14]. Dash, R., & Dash, P. (2016). Efficient stock price prediction using a self evolving recurrent neuro-fuzzy inference system optimized through a modified differential harmony search technique. Expert Systems with Applications, 52, 75-90.

[15]. Tang, L. B., Tang, L. X., & Sheng, H. Y. (2009). Forecasting volatility based on wavelet support vector machine. Expert Systems with Applications, 36(2), 2901- 2909.

Cite this article as :

Siddharth Meghwal, Irfan Khan, "Stock Prediction Using Machine Learning Algorithms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 18-30, September-October 2022. Available at doi : https://doi.org/10.32628/CSEIT22855
Journal URL : https://ijsrcseit.com/CSEIT22855