Stock Market Prediction

Sharanya Banerjee, Neha Dabeeru, R. Lavanya

Abstract: The Stock Market is a challenging forum for investment and requires immense brainstorming before one shall put their hard earned money to work. This project aims at processing large volumes of data and running comprehensive regression algorithms on the dataset; that will predict the future value of a stock using the regression model with the highest accuracy. The purpose of this paper is to analyze the shortcomings of the current system and building a time-series model that would mitigate most of them by implementing more efficient algorithms. Using this model, anyone can monitor the preferred stock that they want to invest in; and maximize profit by purchasing volume at the lowest price and liquidating the stock when it's at its highest.

Keywords: Stock Market, Forecast, Regression, Time-Series Prediction.

I. INTRODUCTION

Stocks form the corner-stone of any business portfolio and may be purchased privately, or from public forums. Any such transaction must conform to legal norms that have been established by the government; in order to prevent illegal practices. *"A stock (also known as "shares" or "equity") is a type of security that signifies proportionate ownership in the issuing corporation. This entitles the stockholder to that proportion of the corporation's assets and earnings."* Historically, stocks have survived the relentless wrath of time, surpassing all its predecessors. Stocks can be bought at the stock exchange or from many online stock brokers.

Significant profit can be yielded from the successful prediction of a stocks future.[11] *"Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange."* The efficient-market hypothesis indicates that the prices of stock reflect all information, available currently and hence, any changes in price that are not based on information that recently came into light, are consequently unpredictable. The people who oppose this hypothesis therefore, possess myriad methods and technologies which calculatedly allow them to gain information about the future price of stock. Stock market price data is tediously voluminous and extremely volatile. Stock market trading is an extremely complicated and ever-changing system where people will either gain a fortune or lose their entire life savings.[11] In this work, an attempt to build a Time-Series prediction model, to predict stock prices.[11]

This paper takes the current stock values from the data sets gathered. The data gathered is modelled into various sub parts or data sets which is used to train and test the algorithm. We use regression models in python or R to model the data. We run a comprehensive search algorithm on the data sets and create a summary table based on the output. We plot the values on a chart and apply regression and clustering techniques to find out the increase or decrease in price of that stock. Based on the calculation, we extrapolate the current stock prices to generate a prediction after a given time. The model is developed using supervised machine learning algorithms.[11] The output will be in graphical form and will change with change in dataset. We expect up to 66% in-sample accuracy and 35% out-of-sample accuracy using supervised machine learning algorithms on prediction model. This will enable the user to take better decisions while investing.[18]

II. LITERATURE SURVEY

The Stock Market of a country is considered an accurate reflection of its economic prowess. Stock Market prices are ever-changing as they are prominently affected by the ebb and flow of finances throughout various economic domains. Prices of stock are invariably dependent on the demand and supply curve, that is a fundamental rule of economics. An increase in demand of a particular stock, will increase its price whereas a decrease in popularity of another, will reduce its price.[11] Even though this fluctuation in stock prices in necessary to reap profits from investment that have been made, it is of paramount importance that we predict the future value of a stock so that loss on investments is mitigated.[13] This review is done in order to predict the prices of stock, so as to make more informed investments.

Recently occurring trends in a market is studied and different types of machine learning classifiers and regression techniques are applied on them. Various approaches[Fig. 1] and results from past studies are weighted based on various parametrics;[Table 1] and then is displayed in a graphical format. The survey brings to light, different conventional approaches to stock market prediction, that has already been built. In addition to that, it discusses recent application of machine learning techniques along with strengths and weaknesses of each technique for effective prediction of stock prices, in the future.[12]
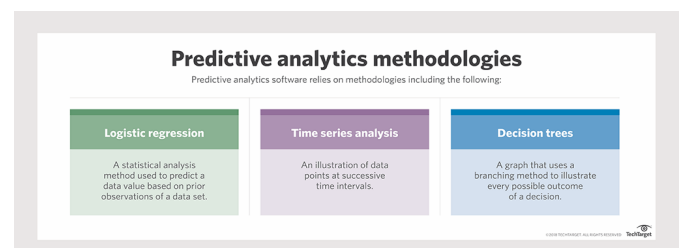

Fig. 1: Predictive Analysis Methodologies[13]

Stock Market Prediction

TABLE 1: Comparative study of the Literature Survey

S.NO	TECHNIQUES USED	POSITIVES	LIMITATIONS
1.	**Stock Market Prediction using Simple Linear Regression.** [1]International Conference on Electronics, Communication and Aerospace Technology (ICECA 2017)	1. By using Linear Regression, the project forecasts the TCS datasets behaviour and the net result is compared and evaluated against the outcome of other approaches. [1] 2. The model incorporates techniques for real world machine learning applications including acquiring and analysing a large dataset; using a variety of techniques to train the model and predict potential outcomes. [1]	1. The prediction model was run on one stock set only and not on the entire market. This creates a certain degree of short sightedness on the entire evaluation process. 2. The method of prediction using Linear Regression is comparatively less accurate, when compared to other methods. 3. Does not consider Random Forest prediction model for prediction, which theoretically gives a higher accuracy when run on a small dataset.
2.	**Stock Market Prediction using SVM and PCA.** [2]2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania	1. This project makes use of intricate Machine Learning algorithms like SVM and PCA and explains them mathematically. 2. It takes care of boundary conditions in SVM by solving the Karush -Kuhn-Tucker (KKT) problem or duality problem mathematically. 3. Comparing the simple methods from SVM and evolving to GASVM and PCASVM, this project develops a new prediction model that is more accurate for real world prediction.	1. The complexity of the undertaken project is extremely high and requires intensive prior knowledge in the field. 2. This paper does not provide a benchmark comparison or study, to evaluate the developed model against.
3.	**Stock market Prediction using Artificial Neural Networks.** [3]2017 International Conference on Technical Advancements in Computers and Communications.	1. This project employs Sentiment Analysis of the social media platform which gives an accurate prediction or evaluation of human behavioural tendencies. 2. The model has multifaceted use; in the stock market, finance, auditing, investment patterns and business strategies. 3. The usage of MLR and SVM in an ANN using deep learning provides the most accurate output in a large dataset.	1. This paper aims at developing an algorithm to predict stock values but does not talk about the accuracy of prediction. It gives a qualitative approach, not a quantitative one. 2. It provides a semantic figure and not a visual outcome to analyse the prediction. No graphs are provided to demonstrate patterns in investment, in the market.
4.	**Stock Market prediction using Hidden Markov Model.** [4]2013 IEEE International Conference on Granular Computing (GrC).	1. This paper employs a new technique: Hidden Markov Model in order to forecast stock prices, which is experimentally shown to give a pretty accurate result. 2. Unlike regression analysis, and other prediction algorithms, this model tries to avoid the factor of selection, combination and transformation of parameters. This allows the avoidance of unnecessary problems. 3. This model chooses pattern matching sequences by historical matching to generate a probabilistic and statistical approach.	1. The Hidden Markov Model is only accurate in a short period prediction. If the time period increases, then the model's performance deteriorates considerably. 2. Only an empirical comparison of prediction models is done. The models are not tested on an out-of-sample dataset. So, we cannot talk about the real-time performance of the model from the given literature.
5.	**Stock Market Prediction using Sentiment Analysis: Hybrid Approach.** [5]International Conference on Computing, Communication and Automation (ICCCA2016)	1. This model employs both sentiment analysis and clustering techniques to give a more accurate result. 2. Hybrid model is superior to individual approaches since it merges the positive factors of both methods. 3. This model can be extended to predict the degree of change. This may be achieved by predicting to which extent specific stock price will move, such as 'very high', 'little high', 'little low', 'very low'. [5]	1. No visual representation of the prediction model is provided. 2. Does not implement regression. Only clustering techniques is used to predict stock values. 3. Does not provide a dynamic model for comparison or prediction, rather uses only fixed values. Hence, it's applicability in the real world is questioned, without any valid evidence to prove otherwise.
6.	**Stock Market Prediction Using Long Short Term Memory.** [6]2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)	1. The more the system is utilized, the greater the accuracy for a large dataset is obtained. The upgraded LSTM approach is more accurate than the regression-based models.[6] 2. This model provides graphical data to visualize data. [6]	1. Improvement in the accuracy of the system can be attained by utilizing a larger dataset.[6] 2. Other emerging models of ML should also be tested for more accuracy of prediction. 3. Sentiment analysis through ML should be done because it is an important factor in stock price fluctuation.
7.	**Stock Transaction Prediction Modelling and Analysis Based on LSTM.** [7]2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)	1. This is the first model to solely use, Long Short-Term Memory (LSTM) as a prediction technique which gives a more accurate result. 2. This paper provides a good explanation of the LSTM model and how it is trained. This provides both the graphical and mathematical implementations.	1. The LSTM model requires a large number of layers to be stacked in order to provide good accuracy. So, it is a tedious process. 2. It is necessary for the LSTM network to be combined with existing clustering techniques to gain large speed ups in training and testing times at the cost of a small drop in performance.[7]

8.	**Real-Time Stock Prediction using Neural Network.** [8]2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)	1. This paper incorporates a culmination of concepts like traded share volume in real time, number of transaction and fluctuations in price analysis, in order to calculate stock prices using a feed forward neural network.[8] 2. The hidden layer neurons were tuned individually for separate stock models. The most accurate outcome was found for ADBL, with an accuracy of 86.12%, modelled by a 3-20-10 network.[8]	1. Requires in-depth knowledge about deep learning and neural networks. There is a large skill gap in this field. 2. The implementation of this model requires heavy computational power to execute.
9.	**Stock Market Prediction Model using ANN and Back Propagation.** [9]ICCCNT'12 26th _28th July 2012, Coimbatore, India	1. This prediction model uses artificial neural networks which gives an overall prediction rate of 63% and market direction accuracy of 81%, theoretically. 2. This paper uses back propagation algorithm for training the data which is not very commonly used. 3. This model gives extensive graphical output to visualize the result.	1. Requires in-depth knowledge of deep learning and neural networks, to understand and implement. 2. As a vast difference in the empirical and actual accuracy as represented by the models.
10.	**Multiple Linear Regression Model to predict Audit Opinions.** [10]2009 ISECS International Colloquium on Computing, Communication, Control, and Management	1. This is an extension of stock prediction model. 2. This paper has selected those financial factors that have a crucial impact on audit opinion; and hence can be used to analyse the relevance and the risk involved in auditing.[10] 3. This paper uses the method of mean squares and best fit which is the most commonly used mathematical method for forecasting.	1. This paper only studies the Shanghai and Shenzhen Stock Markets of 2007 for a year and hence is considerably short sighted.[10] 2. This study does not clearly classify clean audit opinions from the unclean ones.

III. INFERENCE FROM RELATED WORK

As discussed in Table 1, the present system only focuses on a singularity of regression models like Simple Linear Regression, which is not as accurate as their combination. This is because, while performing Linear Regression, we assume that the dependent and independent variables are linearly correlated. However, that is not the case, every time. Therefore, not only may it be infeasible in certain circumstances, the dataset needs to be tailored to the specific algorithm to get an accurate outcome. The present system is seemingly short-sighted. This is because it focuses on only a single stock value but ignores the various other non-linear parameters that may exist and my affect the accuracy of the model. It does not talk about the relative accuracy of similar prediction models. Therefore, it is safe to say that there may exist a more accurate outcome.

IV. PROPOSED WORK

In this paper, an attempt is made to predict the stock market trend using a culmination of predictive modelling and regression algorithms. As shown in Fig. 2, this model takes the current stock values from the data sets gathered. The data gathered is modelled into various sub parts or data sets which is used to train and test the algorithm. Then, regression models in Python or R is used to model the data. This model runs a comprehensive search algorithm on the data sets and create a summary table based on the output. The values are plot on a chart and regression and clustering techniques are applied to find out the increase or decrease in price of that stock. Based on the calculation, this model extrapolates the current stock prices to generate a prediction after a given time. The models are built and trained using supervised machine learning algorithms. The output will be in graphical form and will change with change in dataset. This model expects up to 69% of in-sample accuracy and 37% or out-of-sample accuracy using supervised ML algorithms.

Benefits of the Proposed system

The time-series prediction model will implement various regression algorithms on the same dataset, finding out the relative accuracies of the different models. The most accurate model will be deployed to predict the stock prices, working on the given dataset. It will also take into consideration those parameters, that are not linearly correlated; therefore, expanding the scope of the system and eliminating the issues of short-sightedness.

The range of applicability of the system will also be expanded to such an extent that it can run over various datasets and provide an equally accurate outcome. The efficiency of the system is expected to be second to none. Graphical outputs will provide excellent visualization of the outcome, which will make it easy to understand, even for novice users.

V. IMPLEMENTATION

"*Predictive modelling is a process that uses data mining and probability to forecast outcomes in a given situation.*"[17] Every model is made up of a culmination of variables that are likely to influence future stock prices. Once data has been collected for relevant predictors, a statistical model is formulated by application of various mathematical techniques.[14]

"*Regression is a statistical measurement used in finance, investing and other disciplines that attempts to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables).*"[15] In our endeavour, we are focussing on two main parameters: **Time** and **Price**.

Regression takes two variables: an independent variable X which is invariable of probabilistic parameters; and a dependent variable Y, which is contingent on X in one way or another. The correlation may be linear or non-linear. Certain constants are also incorporated within the equations to generate situations.

Simple Linear Regression: $Y = \alpha + \beta X + u$ [11] [Fig. 3]

Multiple Linear Regression: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_t X_t + u$ [11] [Fig. 4] (α and β are constants)

Multiple Linear Regression with interactions: With added modifications for Support Vector Machines and Principle Component Analysis. [11] [Fig. 5]

Principle Component Analysis: "A method of analysis which involves finding the linear combination of a set of variables that has maximum variance and removing its effect, repeating this successively."

Support Vector Machines: "A support vector machine (SVM) is machine learning algorithm that analyses data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of two categories. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible. SVMs are used in text categorization, image classification, handwriting recognition and in the sciences."
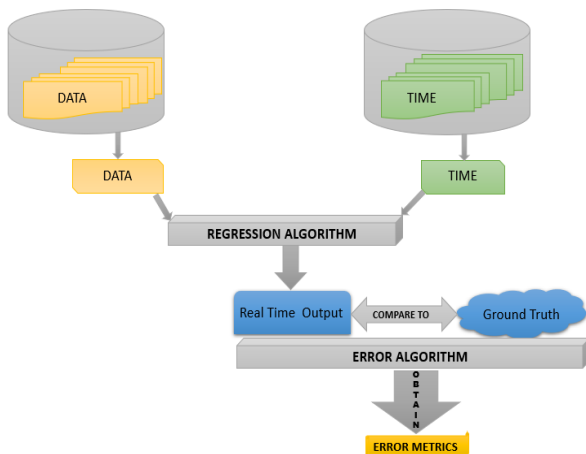

Fig. 2: Block Diagram of the proposed system.

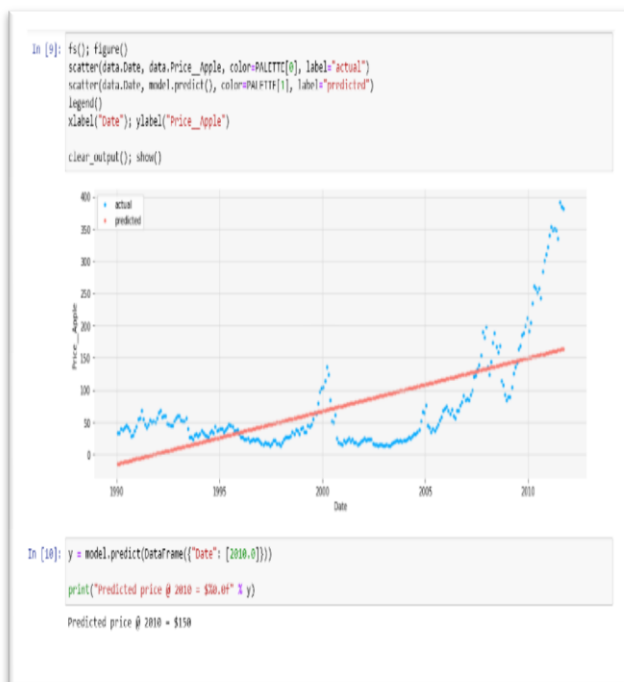## VI. RESULTS AND DISCUSSION

A. Simple Linear Regression:

Fig. 3: Output Graph for Simple Linear Regression predictive modelling technique.

Applying *Simple Linear Regression*, the predicted price is $150 on the selected date in 2010, whereas the actual price is $213. Therefore, the error margin is recorded at $63 or 42%. Accuracy of the prediction is 58%. [Fig. 3]

B. Multiple Linear Regression:

Fig. 4: Output Graph for Multiple Linear Regression predictive modelling technique.

Applying *Multiple Linear Regression*, the predicted price is $162 on the selected date in 2010, whereas the actual price is $213. Therefore, the error margin is recorded at $51 or 31%. Accuracy of the prediction is 69%. [Fig. 4]

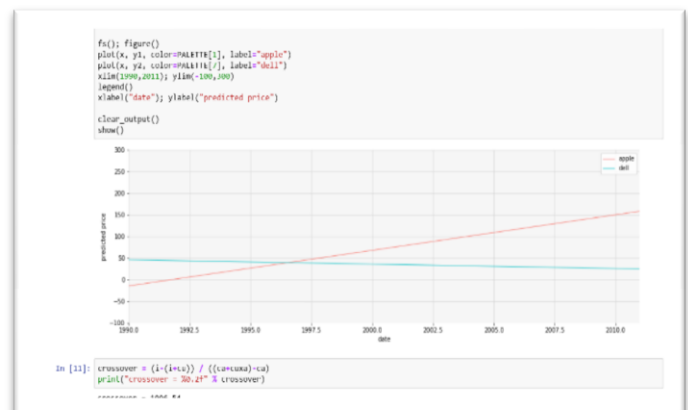*C.* **Multiple Linear Regression with Interactions:**

Fig. 5: Output Graph for Multiple Linear Regression predictive modelling technique, with Interactions.

Applying *Multiple Linear Regression with Interactions*, we can simultaneously predict the trend in stock prices for multiple companies and markets over a given interval of time. In the above graph, we can see that in approximately March of 1996, the stock prices of Apple and Dell crossed over, after which the prices of Apple stock rose steadily whereas the prices of Dell stock fell, over the same interval of time. This information can be particularly useful in diversification of funds, for any investor in the market, or to predict the time of liquidation of funds in order to maximise gains. [Fig. 5]

VII. CONCLUSION

This paper is aimed at helping stock brokers and investors to invest their money wisely and knowingly, in the stock market. The applicability of prediction in stock market business is immense,

which is a very complex process owing to ever-changing facade of the stock market. This project aims at finding the best prediction model among a plethora of those existing today, and implementing the one with the highest empirical and/or real accuracy in order to predict stock prices. The purpose of this paper is to analyse the shortcomings of the current system and building a time-series prediction model that would mitigate most of them, by implementing more efficient algorithms. This would indeed make investment in the stock market, a safer bet.

FUTURE WORK

While solving the problem of redundancies, this model is avoiding the usage of intangible parameters that may influence the stock market; like Human sentiments, social media influence, reputation of the firm, etc. There may exist ways that implement these parameters to get a more accurate outcome. Therefore, to avoid unnecessary issues of redundancy in implementation, it has been excluded of the scope of this paper

REFERENCES

1. Stock Market Prediction using a Linear Regression; International Conference on Electronics, Communication and Aerospace Technology (ICECA 2017).
2. Stock Market Prediction; 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania.
3. Developing a Prediction Model for Stock Analysis; 2017 International Conference on Technical Advancements in Computers and Communications.
4. The Analysis and Prediction of Stock Price; 2013 IEEE International Conference on Granular Computing (GrC).
5. Stock Market Prediction using a Hybrid Approach; International Conference on Computing, Communication and Automation (ICCCA2016).
6. Stock Market Prediction Using Machine Learning; 2018 First International Conference on Secure Cyber Computing and Communication(ICSCCC).
7. Stock Transaction Prediction Modelling and Analysis Based on LSTM; 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA).
8. Real-Time Stock Prediction using Neural Network; 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence).
9. A Stock Market Prediction Model using Artificial Neural Network; ICCCNT'12 26th _28th July 2012, Coimbatore, India.
10. Apply Multiple Linear Regression Model to Predict the Audit Opinion; 2009 ISECS International Colloquium on Computing, Communication, Control, and Management.
11. Internet Source: yourpriceinfo.blogspot.com
12. Internet Source: www.mafiadoc.com
13. Internet Source: https://en.wikipedia.org/wiki/Regression.
14. Internet Source: https://en.wikipedia.org/wiki/Stock_market.
15. Student Paper: Submitted to Wakefield College.
16. Student Paper: Submitted to University of Southampton.
17. Personalized Market Basket Prediction with Temporal Annotated Recurring Sequences; IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 31, NO. 11, NOVEMBER 2019.
18. A Prediction Approach for Stock Market Volatility Based on Time Series Data; IEEE ACCESS (Received January 13, 2019, accepted January 17, 2019, date of publication January 25, 2019, date of current version February 14, 2019.)

AUTHOR PROFILES

Sharanya Banerjee: **Sharanya** completed his Bachelor of Technology from SRM Institute of Science and Technology, Chennai in Computer Science and Engineering; and will embark on a Masters in Computer Science course at Arizona State University, Tempe. His research area is Data Science and Big Data systems. This is his first paper being published. He is a member of IET.

NEHA DABEERU: Neha completed her Bachelor of Technology from SRM Institute of Science and Technology, Chennai in Computer Science and Engineering. This is her first paper being published. She is a member of IET. She is current working as a Business Analyst in Bank of America.

R. LAVANYA: R. Lavanya completed her Master of Technology in Pondicherry University and currently pursuing Ph.D. Her research area is Data science and recommender systems. She has published more than 15 papers in reputed journals. She is a member of ISTE, IET etc. She is currently working as Assistant professor at SRM Institute of Science & Technology, Chennai.