

Open Price Prediction of Stock Market using Regression Analysis

Mr. Pramod Mali¹, Hemangi Karchalkar², Aditya Jain³, Ashu Singh⁴, Vikash Kumar⁵

Professor, Computer Department, SKNCOE, Pune, India¹

Student, Computer Department, SKNCOE, Pune, India^{2,3,4,5}

Abstract: The development of a dynamic application for analysing and predicting stock market prices is a basic tool aimed at accelerating the rate of investors' interest in stock market. The development and implementation of a stock price prediction is explained in this project and for this purpose regression algorithm and object oriented approach of software development is used. In most of the economies trading in shares is big business. From the information gathered from their website, it appears that the stock brokers do not have any intelligent tool which can help them advice their clients on which stocks are proper for them to buy or sell. The prevailing methods show a trend on future movement of stocks and not the likely price for any stock in the future. It is therefore preferable to have a tool that does not just point a direction towards price movement, but also indicates the most likely price of the stock itself.

Keywords: Data mining, Regression, Prediction, Coefficient of correlation, Training set, Testing set.

I. INTRODUCTION

In most of the economies trading in shares is a big business. From the information gathered from their website, it appears that the stock brokers do not have any intelligent tool which can help them advice their clients on which stocks are proper for them to buy or sell. The prevailing methods show a trend on future movement of stocks and not the likely price for any stock in the future. It is therefore preferable to have a tool that does not just point a direction towards price movement, but also indicates the most likely price of the stock itself to a numerical value that is predicted by regression. Where the target values have been defined already, regression performs operations on a data set. By adding new information the result can be extended.

A pattern can be made between predictor and target values by establishing relations through regressions. By using the historical data we can prepare the input data. The input data is grouped into 2 sets as training data set and testing data set in this model. To estimate the unknown coefficients of the regression equation and to train a model the training data set is used. To predict the future price of a stock, the estimated coefficients are used. The comparison is done between actual price and predicted price by using the coefficients to test the training data set and finally the price of the desired stock is predicted.

II. LITERATURE SURVEY

Dr. P. K. Sahoo, Mr. Krishna charlapally in [1] have predicted the future stock values using auto regression. If there is a linear relationship between input and output values, then it is call auto regression. The current value of output variable is predicted using the past values of the output variable and present value of input variables. They have used Moore and Penrose technique to estimate the coefficients of regression equation and these coefficients are further used to predict the future stock values.

Regression analysis is a statistical tool for investigating the relationship between a single dependent variable and more than one independent variable. Input dataset is prepared using historical values and categorized as training dataset and testing dataset. This training dataset is used to train the model and to estimate the unknown coefficients of the auto regression equation. Now, the future stock values are predicted using these coefficients. [1]

Farhad Soleimanian Gharehchopogh, Tahmineh Haddadi Bonab and Seyyed Reza Khaze implemented linear regression method for predicting the behavior of S&P 500 index and showed that their proposed method yields a good performance in comparison to real volumes and their method can be used by stockholders, stockbrokers to invest in stocks. This paper states 61.35% similarity between the actual and the predicted values. The regression equation is formulated by using trading volume as the dependent variable and the average price per share as the independent variable. Average price is calculated as the mean of the prices of Open, Low, High and close of that particular stock. [2]



SACHIN KAMLEY, SHAILESH JALOREE and R. S. THAKUR applied the multiple regression technique based on three variables to predict the stock market price from stock market data. In this approach monthly data of past four and half years for Infosys Company is obtained from yahoo finance and used for building the prediction model and the accuracy of the model is checked by using same data of past one and half years. This paper predicts the closing price of the stock using open and close price as dependent variables. The least square method is used to calculate the slope of the regression equation associated with the corresponding dependent variable (X_i). The advantage of this method is that it minimizes the total error. This approach of multiple regression using 3 variables the accuracy of system is stated to be 89%, thereby indicating better accuracy as compared to linear regression method used for prediction. [3]

K. V. Sujatha, S. Meenakshi Sundaram presented a comparative survey of different non-parametric linear regression and artificial neural network models and compared their accuracy of prediction of closing price of stock index data where the dataset is not normalized. This paper further hints that predictive models built using neural networks yield much better results. [4]

Han Lock Siew, Md Jan Nordin have used the regression techniques for predicting the stock value by transforming the data set in ordinal data format. The data retrieved contains values of heterogeneous data types used for handling currency values and financial ratios. This is transformed to a standardized ordinal data type which is used to rank the stock values for comparison purpose. WEKA machine learning software is used as a reference to design the regression analysis model. The initial data sets are corporate annual reports (balance sheet, income statement and cash flow statement). Fundamental analysis was used to select the variables and then the classifiers in WEKA were used to produce the outcomes. This study shows the importance of standardized dataset in prediction using regression techniques. [5]

Azadeh Nikfarjam, Ehsan Emadzadeh, Saravanan Muthaiyah have proposed a text mining approach to predict the behaviour of stocks. They have made use of the fact that news has a very crucial impact on the behavior of stocks and hence analyzed the news to predict stock's behavior. This news based market prediction method operates in two phases, the training phase and the operational phase. First, the data is trained/prepared using a set of training dataset and then in operational phase, the incoming news is classified into one of the classes prepared in training phase. [6] has suggested 3 components in this predictive systems, which are News labelling, Classifier input generation, Classification. After the data is classified in operational phase (news labelling), it is given to the classifier which trains itself and then classifies the future data whose output is unknown. [6].

III. SYSTEM ARCHITECTURE

The architecture comprises of four layers namely Presentation Layer, Persistence Layer, Business Layer and Resource Layer. Presentation Layer is the layer visible to the end user. Business Layer contains the logic/algorithm for prediction of Open Price of a particular stock. Persistence Layer selects the historical and updated stock data for prediction and forwards it to the Business Layer for processing. Resource Layer contains databases namely user database for storing the credentials of registered users and stock database for storing the historical data of stocks.

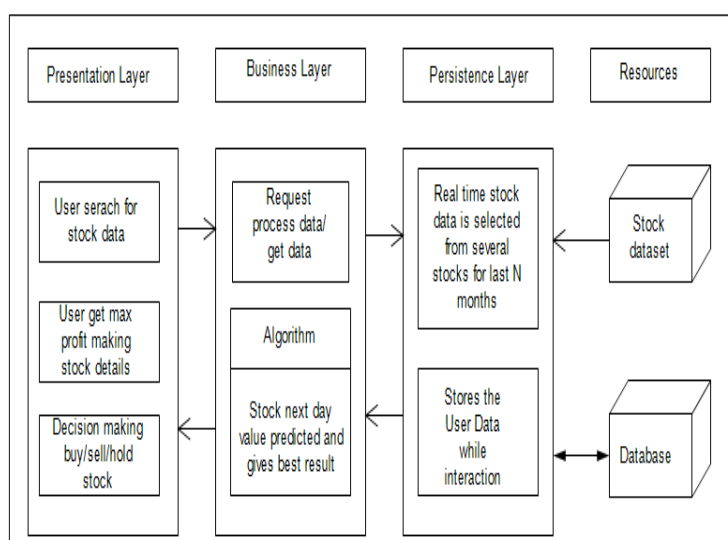


Fig. 1. Architecture of the proposed system

IV. COMPARATIVE STUDY

TABLE I COMPARISON OF TECHNIQUES

Criteria	[1]	[2]	[3]	[4]	[5]	[6]
Neural Network	No	No	No	Yes	No	No
Auto Regression	Yes	No	No	No	No	No
Linear Regression	Yes	Yes	Yes	Yes	Yes	No
Text Mining	No	No	No	No	No	Yes
Multiple Regression	Yes	No	Yes	No	No	No

V. ALGORITHM

The dictionary meaning of the word Regression is ‘Stepping back’ or ‘Going back’. Regression measures the relationship between two or more variables in terms of the original units of the data.

It attempts to establish the nature of the relationship between variables that is to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting.

Regression analysis is a statistical technique that explains the variation of one variable w.r.t. change in some other variable(s). The variable, of which the variation is explained, is called dependent variable, while the variable(s) which are used to explain the variation are called independent variables.

This technique has been widely used for forecasting, time series modelling and finding the cause and effect relationship between the variables.

Different types of regressions are Linear regression, Multiple regression, logistic regression, Linear Regression, Logistic Regression, Polynomial Regression, Stepwise Regression, Ridge Regression, Lasso Regression, Elastic Net Regression but the prime focus of this paper is on multiple regression.

Multiple regression is an extension of simple linear regression in which more than one independent variable (X) is used to predict a single dependent variable (Y).

The predicted value of Y is a linear transformation of the X variables.

Multiple Regression Equation having “K” independent variables is given by:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

For three independent variables, the prediction of Y is expressed by the following equation:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

In our case,

X1= High value of a stock of previous day,

X2=Low value of a stock of previous day,

X3=Close value of a stock of previous day.

VI. CONCLUSION

In this paper we studied some well-known prediction algorithms concerned with regression. A comparative study has been made using tabular format. The accuracy yielded by them and the various parameters used for prediction have been stated. It is observed that prediction using multiple regression yields better results than linear regression. Further the use of neural networks for prediction sounds to be a promising field in the future and can be used for real time trading in stock market.

VII. FUTURE SCOPE

The efficiency of this model can be improved by incorporating the concept of Artificial Neural Network to account for the dynamic factors like present day happenings around the world.

ACKNOWLEDGMENT

We would like to give our sincere thanks to our guide **Prof. Pramod Mali** and co-guide **Prof. Raghbir Nasri** for their constant support and guidance. We are also grateful to **Dr. P. N. Mahalle**, Head of Computer Engineering Department, STES’ Smt. Kashibai Navale College of Engineering for his indispensable support, suggestions.



We would also like to extend our sincere thanks to each and every individual who was even remotely associated in the successful completion of this project

REFERENCES

- [1] Dr. P.K.Sahoo, Mr. Krishna charlapally , "Stock Price Prediction Using Regression Analysis", International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015 ISSN 2229-5518
- [2] Farhad soleimani gharehchopogh, tahmineh haddadi bonab and seyed reza khaze , "a linear regression approach to prediction of stock market trading volume: a case study", international journal of managing value and supply chains (ijmvsc) vol.4, no. 3, september 2013
- [3] Sachin kamley1 , shailesh jaloree2 & r. S. Thakur3 , "multiple regression: a data mining approach for predicting the stock market trends based on open, close and high price of the month", international journal of computer science engineering and information technology research (ijcseitr) issn 2249-6831 vol. 3, issue 4, oct 2013
- [4] K. V. Sujatha , S. Meenakshi Sundaram , "Stock Index Prediction Using Regression and Neural Network Models under Non Normal Conditions", Emerging Trends in Robotics and Communication Technologies (INTERACT) IEEE, 31 January 2011
- [5] Han Lock Siew , Md Jan Nordin , "Regression Techniques for the Prediction of Stock Price Trend", Statistics in Science, Business, and Engineering (ICSSBE)IEEE, 31 December 2012
- [6] Azadeh Nikfarjam , Ehsan Emadzadeh , Saravanan Muthaiyah , "Text mining approaches for stock market prediction", Computer and Automation Engineering (ICCAE), IEEE, Volume: 4, March 2010
- [7] Arora, P.N. and Arora, S. (2009). Statistics for Management. New Delhi: S. Chand & Company Limited. ISBN: 81-219-2285-2.
- [8] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From Data Mining to Knowledge in Databases. Menlo Park, Calif: American Association for Artificial Intelligence.
- [9] Ravindranath, B. (2003). Decision Support Systems and Data Warehouses. New Delhi: New Age International Limited. ISBN: 81-224-1454-0.
- [10] Kannan, K.S., Sekar, P.S., Sathik, M.M. and Arumugan, P. (2010). Financial Stock Market Forecast using Data Mining Techniques. Hong Kong: Proceedings of International MultiConference of Engineers and Computer Scientist 2010 Vol. I. IMECS 2010. ISBN: 978-988-17012-8-2.
- [11] Han, J. and Kamber, M. (2006). Data Mining: Concepts and Techniques, (2nd ed.). Morgan Kaufmann.
- [12] Larose, D.T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. New York: John Wiley and Sons, Inc. ISBN: 0-471-66657-2
- [13] Portia, A.C.(n.d). Data Mining and Neural Networks from a Commercial Perspective. Australia: University of Technology Sydney, Australia.
- [14] Two Crows Corporation (2005). Introduction to Data Mining and Knowledge Discovery, (3rd ed.). Potomac, USA: Two Crows Corporation. ISBN: 1-892095-02-5
- [15] Weiss, G.M., and Davison, B.D. (2010). Data mining. Handbook of Technology Management, H.Bidgoli (Ed.), John Wiley and Sons.