Stock Market Prediction using Linear Regression in Machine Learning

APEL MAHMUD[1], MD. SHARIAR FAHIM[2], RAZIA SULTANA MISU[3]

ID: 193-15-2984 ID: 193-15-2996 ID: 193-15-2976

[1-3]Department of Computer Science and Engineering, Daffodil International University

Abstract— 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. We predict close values based on certain data. Like high value, low value, open value, etc. This is the stock market prediction. And also we have worked on the linear regression of the paper. And after working on linear regression, our algorithm worked a good prediction. We have worked with linear regression and working with linear regression has given us a very good result. We have analyzed that graph, we have found that the error is less than the main value, and is giving an error of very limited around.

Keywords— Linear Regression, Stock Market, Close Value, High Value, Volume, Average Value.

1. Introduction

A stock market prediction is an attempt to forecast the future value of an individual stock, a particular sector or the market, or the market as a whole. Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision. The output column is then assigned to the target variable in the following step. It is the adjusted relative value of the TESLA Stock in this situation. Furthermore, we pick the features that serve as the independent variable to the target variable (dependent variable). We choose four characteristics to account for training purposes: Open, High, Low, Volume. And we are predicting close value hare. The Adjusted Close Value is the final output value that will be forecasted using the Machine Learning model. An American car manufacturing company Tesla Inc. was founded on 2003.It is one of the companies that developed electric cars to a manufacturing standard and made it available for the consumer market. In the beginning of 2010, the stock price of Tesla started. A huge demand increased in the used car market for Tesla vehicles. But the price for Tesla vehicle depended on the stock price of the vehicle and it impacted largely in the Tesla vehicle market. Tesla vehicles highly depends on Tesla company's post sells services. In order to keep the vehicles usable in the market and also decline of the stock prices the accurate prediction of car prices ware very important for the new and existing customers and the company itself. The prices of a same vehicle were different in a range of three months. This means, the month of recorded price of a certain Tesla vehicle was also very important. The attributes that was found in the datasets were: Date, Open, High, Low, Close, Adj Close, Volume. Since the prices were getting very saturated, two major attributes were focused to predict the price of Tesla vehicles in

order to have an accurate result. These are: Actual, Predicted. We will also discuss the test out of a different machine learning techniques that were primarily chosen to predict the price. Later the Linear Regression was selected as the main algorithm of the price prediction system.

2. Research Questions

Many have worked on stock market prediction before. Some of their work is given below.

Zhen Hu et al. [1] utilized the help of vector machine calculation to anticipate the financial exchange price. By utilizing this calculation, they have anticipated the precision result. SVM is utilized here to fathom the direct compelled quadratic programming issue. Utilizing the SVM the bring one-of-a-kind answers for the issue. Mehak Usmani [2] done the work done in this paper to count on Karachi inventory exchange. This paper offers a Neural device and Support vector gadget calculation is Ob-jective is to foresee the marketplace execution of Karachi inventory exchange (KSE). This paper precept purpose for current is simply to foresee the Karachi inventory exchange. The results of the using Gradient-Descendent to attain to perfect worth. Tejas Mankar [3] done on this paper expectation depending on the social emotions utilizing device learning. The evaluation given with the aid of using humans in standard approximately the company. Based at the optimistic and destructive observation of open approximately the corporation they're foreseeing the economic exchange. Utilizing as-sumption research at the tweets collected utilising the Twitter API and moreover the give up estimations of various stocks, we attempt to bring together a framework that gauges the inventory price improvement of various organizations. El Mehraz [4] paintings carried out in this paper foreseeing the securities alternate expectation utilising Hybrid technique which consolidates Support Vector Regression. 332 Reshma R et al. / Stock Market Prediction Using Machine Learning Techniques Our goal on these exploration paintings is to advocate a half of and half of technique that joins bolster vector relapse (SVR) and Hodrick-Prescott channel (HP), for upgrading the forecast of inventory fee with the aid of using studying the recorded statistics of utilising our proposed model. TESLA with this dataset and linear regression nobody has done well before so we are going to do it.

3. Research Methods

we have worked on the linear regression of the paper. And after working on linear regression, our algorithm worked a good prediction. We have worked with linear regression and working with linear regression has given us a very good result. We have analyzed that graph, we have found that the error is less than the main value, and is giving an error of very limited around. In this section, we present the proposed methods and the design of the proposed solution. Moreover, we also introduce the dataset as well as algorithmic and implementation details.

3.1 Data Collections

DATASET:

I took a dataset; this is a CSV type dataset. There are some heads is this dataset like, Date, Open, High, Low, Close, Adj Close, Volume. We have the total number of columns and rows, 7 columns and 2193 rows. I took the dataset from kaggle. This is a TESLA dataset. Depending on the TESLA company depends on their stock market. we'll be utilizing scikit-learn, csv, NumPy and matplotlib bundles to actualize and picture direct relapse toward the mean. The Table 1 describes the sample dataset.

	Date	Open	High	Low	Close	Adj Close	Volume
0	29-06-2010	19.000000	25.00	17.540001	23.889999	23.889999	18766300
1	30-06-2010	25.790001	30.42	23.299999	23.830000	23.830000	17187100
2	01-07-2010	25.000000	25.92	20.270000	21.959999	21.959999	8218800
3	02-07-2010	23.000000	23.10	18.709999	19.200001	19.200001	5139800
4	06-07-2010	20.000000	20.00	15.830000	16.110001	16.110001	6866900

Table 1 Sample data set

3.2 Data Analysis [Description of algorithm used]

We have got this dataset, we have given this dataset as High value, Low value, Volume, Open value. will predict a close value. I used linear regression to do this. I have applied linear regression in the sklearn of python.

Linear Regression:

The most basic machine learning algorithmic rule that may be enforced on this information is regression. The regression model returns associate degree equation that determines the link between the independent variables and therefore the variable.

The equation for regression may be written as:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Here, x1, x2,…. xn represent the freelance variables whereas the coefficients θ1, θ2, …. θn represent the weights.

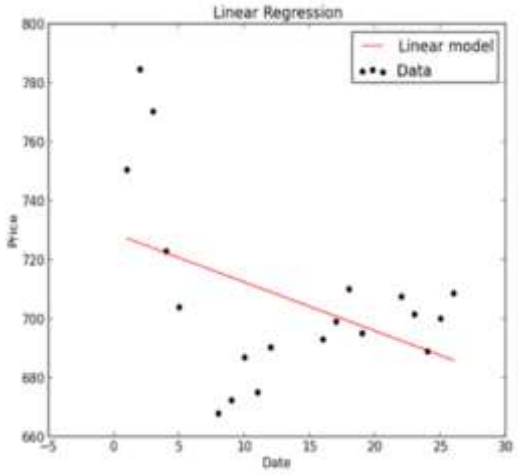

Fig 1 represents the Linear regression model for price and date of the stock market.

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

4. Results

We have predicted values, in some cases some less and some more, but we have got a fairly close value. Didn't get too gross way. This is a regression model problem. In this we got Mean Absolute Error = 1.3782280626252839, Mean Squared Error = 3.9090508840931393, Root Mean Squared Error = 1.9771319844899429. We will try to reduce this error in the future. In the venture Linear Regression have been utilized to predicted the stock qualities from the dataset. To plot the diagram and foresee the qualities we have utilized Pandas, NumPy, Sklearn with direct model, matplotlib, tesla.CSV. Here is the plotted diagram.

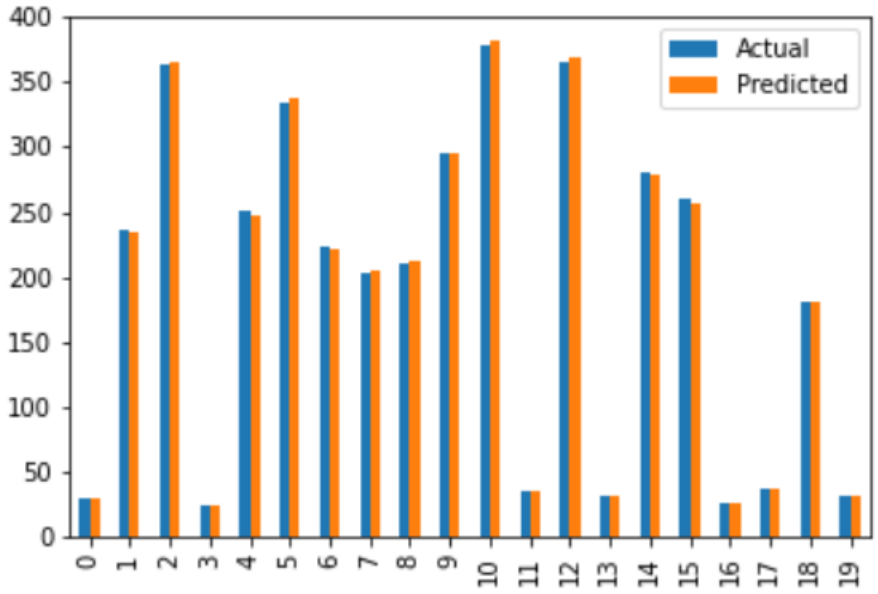

Fig 1: Actual & Predicted plot for the dataset

The yellow specks in the above plot demonstrate the information focuses plotted at each Predicted and Actual The blue line is the condition shaped by the fit technique for the direct model.

5. Discussion

This report was made for Stock market predictions based on Tesla Company. We have used Linear Regression Algorithm. But in future work, we will do better algorithm system for this type of report so that we can get more knowledge about the report we have done. We can also do better in our data set by giving different types of data, so that the information of calculation improves in the future. In the future market, the prediction will be easier than now, because of using powerful algorithm systems and heavy datasets

6. Conclusion

In the project, we proposed the use of the data collected from Tesla Company with machine learning algorithms in order to predict the stock index movements. linear regression algorithm works on the large dataset value which is collected from Tesla Company. Also, SVM does not give a problem of over fitting. Various machine learning based models are proposed for predicting the daily trend of Market stocks. Numerical results suggest the high efficiency. The practical trading models built upon our well-trained predictor. The model generates higher profit compared to the selected benchmarks.

References

1. Kaya. M.I.Y., Karsligil. M.E. in 2ND IEEE International Conference on Information and FinancialEngineering (ICIFE), pages 478-482, 17th to 19th September 2010.
2. Podding, T and Rehkugler, H, “A world of integrated financial markets using artificial neural networks”,Nerocomputing, pages 251-273, 1996.
3. Murphy J.J., “Technical Analysis of Financial Markets – A Comprehensive Guide to Trading Methods andApplications”, New York Institute of Finance, 1999.
4. Trippi, R.R., DeSieno, D., “Trading Equity Index Futures with a Neural Network”, the Journal of PortfolioManagement, pp. 27-33, fall 1992.
5. Kryzanowski, L., Galler, M., Wright, D.W., “Using Artificial Networks to Pick Stocks”, Financial Analysts Journal, pages 21-27, August 1993.
6. Lee, Jae Won, Jonghun Park, Jangmin O, and Jongwoo Lee. 2007. ”A Multiagent Approach to Q-Learning for Daily Stock Trading.” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS 864-877
7. C. Tsai, and S. Wang 2009. Stock Price Forecasting by Hybrid Machine Learning Techniques. Proceedings of the International MultiConference of Engineers and Computer Scientists, 20-26, 2009.
8. Yan, Robert, and Charles Ling. 2007. ”Machine Learning for Stock Selection.” Industrial andGovernment Track Short Paper Collection 1038-1042.
9. Hurwitz, E, and T Marwala. 2009. ”Common Mistakes when Applying Computational Intelligence and Machine Learning to Stock Market modelling.” University of Johannesburg Press.
10. Nikfarjam, A., Emadzadeh, E., &Muthaiyah, S. (2010). Text mining approaches for stock market prediction. In Computer and Automation Engineering (ICCAE), 2010 the 2nd International Conference on (Vol. 4, pp. 256- 260). IEEE.
11. Zhang, X., Fuehres, H., &Gloor, P. A. (2011). Predicting stock market indicators through twitter “I hope it is not as bad as I fear”. Procedia-Social and Behavioral Sciences, 26, 55-62.
12. Karabulut, Y. (2011). Can Facebook predict stock market activity? Available at http://bus.miami.edu/umbfc/_common/files/papers/Karab ulut.pdf [last Available 02.07.2013].
13. Demirguc-Kunt, A., & Levine, R. (Eds.). (2004). financial structure and economic growth: A cross-country comparison of banks, markets, and development. MIT press.
14. Rouwenhorst, K. G. (1999). Local return factors and turnover in emerging stock markets. The Journal of Finance, 54(4), 1439-1464.
15. Pesaran, M. H., &Timmermann, A. (1994). Forecasting stock returns an examination of stock market trading in the presence of transaction costs. Journal of Forecasting, 13(4), 335-367.

16. Gharehchopogh, F. S., &Khalifehlou, Z. A. (2012). A New Approach in Software Cost Estimation Using Regression Based Classifier. AWERProcedia Information Technology and Computer Science, Vol: 2, pp. 252-256.
17. Draper, N. R., Smith, H., &Pownell, E. (1966). Applied regression analysis (Vol. 3). New York: Wiley.
18. Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: data mining in financial application. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 34(4), 513-522.