

EFFICIENCY ANALYSIS OF STOCK PRICE PREDICTION USING MACHINE LEARNING

Computer Science

Vishnu U

Department of Computer Science and Engineering, Dayananda Sagar University, Hosur Road, Bangalore, Karnataka, India.

ABSTRACT

We have used a stock price dataset to train a Machine Learning model to learn the relation. Then this model is used to predict the stock price during a situation when the economy was unstable. This helps us understand the efficiency of this technique to predict stock prices. A linear model is trained to learn the relationship between the opening and closing price of the data. The same model is then tested on a situation where the closing price of the stock deviates from its normal trend. The performance of the model is evaluated at this time. This experimental study is aimed at proving that stock prices depend on economic and organizational conditions as well and not just on the opening price of that day. This project also aims to convey that Machine Learning is only a partially efficient solution to the stock price prediction problem.

KEYWORDS

Machine Learning, Stock prices, Efficiency Analysis

1. INTRODUCTION

Machine Learning techniques have been used to analyze and predict stock prices of a given company. These systems are quite common nowadays. But how reliable are these predictions? It is a known fact that stock prices fluctuate on many factors like Economic factors, Company trends etc. In this paper, we will evaluate the reliability of these predictions by training a model on a stock price dataset and then trying to predict when the price fluctuated due to some Economic conditions. This will help us analyze the reliability of such systems when situation deviates from the normal situation.

2. Literature Survey

The Stock Price prediction systems consists of a linear or non-linear model (Artificial Neural Network). Majority of the models use Opening price as the feature and Closing price as the target variable while some models include date as a feature well. The historical data of the stock prices are used to train and test the model.

Since Stock prices are Time series data, LSTMs are used majorly for a better prediction. Apart from LSTMs, ANN and Linear Models (for linear data if present) are also used.

Opening price of the stocks are the commonly used feature for the model and the closing price is the target variable. In few systems dates are also used as a feature, although date as a feature is also rarely used.

3. Problem Definition

An ML Model is used to train on the historical data of the stock prices of a chosen stock price dataset (Nifty50). This is then used to predict the stock prices during a recent unstable economic condition (COVID-19 lockdown) to evaluate its reliability.

The dataset used will be Nifty50 Stock Price dataset and the data is taken from January 2018 – February 2020. Hence, we learn the data till the lockdown period and predict the prices during the lockdown period to evaluate how well it copes up with the unstable economic conditions compared to the trend in the previous years.

In this experiment, we will be evaluating the model prediction in both quantitative (absolute error) and qualitative way (From the perspective of an investor) to conclude its reliability.

4. Equations

For the Nifty50 dataset majority of the data follows a linear relationship i.e., Opening vs Closing price follows a linear relationship (shown in a later section). Hence, we will be using a linear model here.

For a linear model, the following equation holds :

$$Y' = WX + \beta$$

Where,

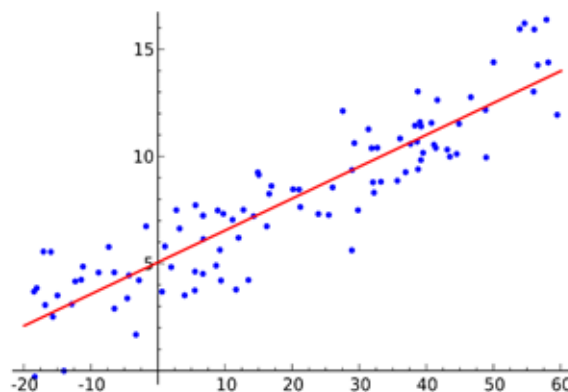
Y' is the predicted value.

W is the weight value which is to be estimated

X is the date item or feature (Opening price here)

β is the bias and is to be estimated

The below graph shows how a linear model will look like in visualization. The red line shows the optimum fit for the data points.



Since it is Linear Regression, Mean Squared Error would be an optimal loss function for this problem. The following expression defines MSE function :

$$E = (Y' - Y)^2 / N$$

Where,

E is the error value

Y' is the predicted

Y is the actual value also known as ground truth

N is the total samples in the dataset

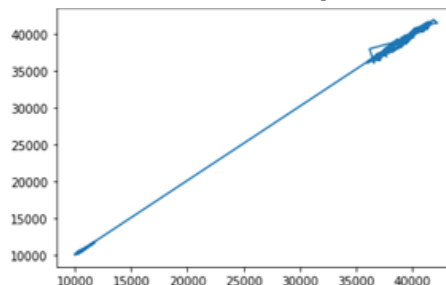
5. Methodology / Approach

The training data is taken from finance websites. Nifty50 data range from January 2018 – February 2020

is taken. This data is then split into train and validation parts.

Firstly, the trend of the opening and closing price is observed which help us determine the kind of model to be used for this purpose.

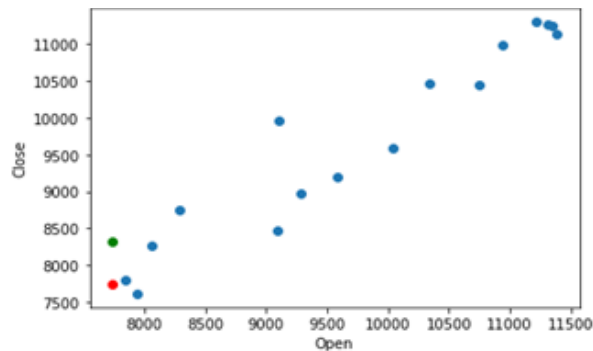
For non-linear fit, Non-linear models like ANN or LSTMs have to be used while for a linear fit, linear model will be apt.



From the graph, it is clearly visible that the data follows a linear fit with very few outliers at the end. Hence, a linear model would be suitable for this problem. Then the model is trained and evaluated using K-fold cross validation.

Now predict a stock price during the COVID-19 Lockdown time to check how the model handles a scenario which is outside of its scope.

The lowest opening price of the Nifty50 during lockdown was 7735.15. The actual closing price for that day was 8317.85 which deviates from its linear trend as we observed before for other data points. But the model predicts the closing price of that day to be 7735.156 which is obtained by following the linear equation of the model.



The green dot refers to the actual closing price and the red dot refers to the predicted closing price. The distance between both is clearly visible and shows how the model becomes unreliable during an unstable economic condition.

6. Results

The train model performs well with a train loss of 0.00005 and validation loss of 0.0001. From the loss graph, it is clear that the train and validation curves have merged completely which shows the model has attained a good fit. But when a prediction is done for an outlier situation it is clear that the model performs poorly. This shows that stock prices depend on many other factors and not only on its opening price. The plummet in opening price was caused by the COVID-19 lockdown situation while the unusual hike in closing price was because of the big traders buying the stock at lower price because of the low price and because they saw potential for growth after the COVID-19 impact.

Performing both qualitative and quantitative analysis, we can confirm that the model performs poorly. The 1000 unit difference in prediction is a huge gap for a predicted stock price.

7. Conclusion

Thus, it can be concluded that stock prediction requires you to consider both industry and economic factors as well along with the opening stock price. Machine Learning projects out certain restrictions when it comes to predicting volatile values like stock prices. Hence, Machine Learning models can be considered as partially efficient for stock price prediction.

REFERENCES

- [1] Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). *Stock Market Prediction Using Machine Learning*. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). R. Caves, Multinational Enterprise and Economic Analysis, Cambridge University Press, Cambridge, 1982. (book style)
- [2] Werawithayaset, P., & Tritilanunt, S. (2019). *Stock Closing Price Prediction Using Machine Learning*. 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE). H.H. Crockell, "Specialization and International Competitiveness," in Managing the Multinational Subsidiary, H. Etemad and L. S. Sulude (eds.), Croom-Helm, London, 1986. (book chapter style)
- [3] Ravikumar, S., & Saraf, P. (2020). *Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms*. 2020 International Conference for Emerging Technology (INCET)