

STOCK PRICE PREDICTION USING MACHINE LEARNING

Anurag Pandey^{*1}, AbhinavPratap Singh², Amandeep Singh³

¹B.Tech. Scholar, Department of Information Technology, PSIT Kanpur

E-mail: anuragpandey9950@gmail.com

²B.Tech. Scholar, Department of Information Technology, PSIT Kanpur

E-mail: abhinavps129@gmail.com

³B.Tech. Scholar, Department of Information Technology, PSIT Kanpur

E-mail: amandeepsingh9559@gmail.com

Conflicts of interest: Nil

Corresponding author: Anurag Pandey

Abstract

Forecasting stock exchange prices have always been challenging task for several business analyst and researchers. In fact, stock exchange price prediction is a stimulating area of research for investors. For successful investment, many investors have an interest in knowing about the longer-term situation of the market. Effective prediction systems indirectly help traders by providing supportive information like the longer-term market direction.

This project aims at predicting the stock market by historical data. It proposes a completely unique method for the prediction of the stock exchange price. Many researchers have contributed to the present area of chaotic forecast in their ways. Fundamental and technical analyses are the normal approaches thus far.

Keywords: Stock Price Prediction, Machine Learning, Supervised Learning, Regression.

1. Introduction

Prediction of stock prices is one of the most difficult things to do. The stock market is a platform for trading a company's stocks. In any country stock exchange is one of the foremost emerging sectors. Nowadays, many of us are indirectly or directly associated with this sector. Therefore, it becomes essential to understand market trends. Thus, with the event of the stock exchange, people have an interest in forecasting stock price. Stock markets are mostly a non-parametric, non-linear, deterministic chaotic system. There are several factors involved in the prediction – physical factors vs. physiological, rational and irrational behaviour and also so many other factors involved in this. All these aspects together make share prices volatile and very

difficult to predict with a high degree of accuracy. Systems instead of fundamental analysis for predicting prices of stocks, which helps them to require immediate investment decisions. One of the most aims of a trader is to predict the stock price such he can sell it before its value decline or buy the stock before the price rises. The efficient market hypothesis states that it's impossible to predict stock prices which stock behaves within the stochastic process. It is very difficult to replace the professionalism of an experienced trader in predicting the stock price. But due to the supply of an interesting amount of knowledge and technological advancements we will now formulate an appropriate algorithm for prediction whose results can increase the profits for traders or investment firms.

As the technology is increasing, stock traders are moving towards using Intelligent Trading Systems instead of fundamental analysis for predicting prices of stocks, which helps them to require immediate investment decisions. One of the most aims of a trader is to predict the stock price such he can sell it before its value decline or buy the stock before the price rises. The efficient market hypothesis states that it's impossible to predict stock prices which stock behaves within the stochastic process. It is very difficult to replace the professionalism of an experienced trader in predicting the stock price. But due to the supply of an interesting amount of knowledge and technological advancements we will now formulate an appropriate algorithm for prediction whose results can increase the profits for traders or investment firms.

We can use machine learning as a game-changer during this domain. In machine learning, there are so many techniques are present which can help in predicting the stock prices, machine learning techniques have the potential to unearth patterns and insights we didn't see before, and these are often wont to make unerringly accurate predictions.

In this paper, we will work with historical data on the stock prices of a publicly listed company. We will implement a mixture of machine learning algorithms to predict the longer term stock price of this company, starting with simple algorithms like averaging and rectilinear regression, and then move on to advanced techniques like Auto LSTM.

The main idea behind this article is to show how these algorithms are implemented. Here I will briefly describe the technique and provide relevant links to brush up on the concepts as and when necessary.

2. Related Works

2.1 Stock Closing Price Prediction using Machine Learning Techniques

Accurate prediction of stock exchange returns may be a very challenging task thanks to the volatile and non-linear nature of the financial stock markets. With the introduction of AI and

increased computational capabilities, programmed methods of prediction have proved to be more efficient in predicting stock prices. during this work, Artificial Neural Network and Random Forest techniques are utilized for predicting subsequent day price for five companies belonging to different sectors of operation. The financial data: Open, High, Low and shut prices of stock are used for creating new variables which are used as inputs to the model. The models evaluated using standard strategic indicators: RMSE and MAPE. The low values of those two indicators show that the models are efficient in predicting stock price.[1]

2.2 A Bayesian regularized artificial neural network for stock market forecasting

In this paper, a Bayesian regularized artificial neural network is proposed as a completely unique method to forecast financial market behaviour. Daily market prices and financial technical indicators are utilized as inputs to predict the at some point future price of individual stocks. The prediction of stock price movement is generally considered to be an important and challenging task for financial time series analysis. The accurate prediction of stock price movements could play a crucial role in helping investors improve stock returns. The complexity of predicting these trends lies within the inherent noise and volatility in daily stock price movement. The Bayesian regularized networks assign a probabilistic nature to the network weights, allowing the network to automatically and optimally penalize excessively complex models. The proposed technique reduces the potential for overfitting and overtraining, improvement of the prediction quality and generalization of the network. Experiments were performed with Microsoft Corp. and Goldman Sachs Group Inc. stock to determine the effectiveness of the model. The results indicate that the proposed model performs also because the more advanced models without the necessity for pre-processing of knowledge, seasonality testing, or cycle analysis.[2]

2.3 Stock Price Prediction Using Convolutional Neural Networks

Prediction of future movement of stock prices has been a topic matter of the many research work. In this proposed work, we have a hybrid approach for stock price prediction using machine learning and deep learning-based methods. We select the NIFTY 50 index values of the National stock market of India, over a period of 4 years, from January 2015 till December 2019. Based on the NIFTY data during the said period, we build various predictive models using machine learning approaches then use those models to predict the Close value of NIFTY 50 for the year 2019, with a forecast horizon of 1 week. For predicting the NIFTY index movement patterns, we use a number of classification methods, while for forecasting the actual Close values of the NIFTY index, various regression models are built. We, then, augment our predictive power of the models by building a deep learning-based regression model using a Convolutional Neural Network with a walk-forward validation. The CNN model is fine-tuned for its parameters so that the validation loss stabilizes with an increasing number of iterations, and the training and validation accuracies converge. We exploit the power of CNN in forecasting the future NIFTY index values using three approaches which differ in the number of variables used in forecasting, the number of sub-models used in the overall models and, the size of the input data for training the models. Extensive results are basically presented on various metrics for all classification and regression models. The results clearly indicate that the CNN-based multivariate forecasting model is the most effective and accurate in predicting the movement of NIFTY index values with a weekly forecast horizon.[3]

3. Machine Learning

Machine Learning is that the science of getting computers to find out and act like humans do, and improve their learning over time in an autonomous fashion, by feeding them data and knowledge within the sort of observations and real-world interactions. within the increasingly competitive corporate world, ML is changing and reworking automation industries. There are numerous apps that are build by using machine learning technologies. during this article, we'll

dive deeper into what machine learning is, the basics of ML, kinds of machine learning algorithms, and a few samples of machine learning in action.

Types of Machine Learning:

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

3.1 Supervised Learning

In supervised learning, as the name indicates, it works as a supervisor. Basically supervised learning is once we teach or train the machine using data that's well labelled. Which means some data is already tagged with the proper answer. After that, the machine is given a replacement set of examples(data) in order that the supervised learning algorithm analyses the training data(set of coaching examples) and produces an accurate outcome from labelled data.

Types of Supervised Learning

- a) Regression
- b) Logistic Regression
- c) Classification
- d) Naive Bayes Classifiers
- e) K-NN (k nearest neighbors)
- f) Decision Trees
- g) Support Vector Machine

3.2 Unsupervised Learning

Unsupervised learning is that the training of a machine using information that's neither classified nor labelled and allowing the algorithm to act thereon information without guidance. Here the task of the machine is to group unsorted information consistent with similarities, patterns, and differences with none prior training of knowledge .

Unlike supervised learning, no teacher is as long as means no training are going to be given to the machine. Therefore the machine is restricted to seek out the hidden structure in unlabelled data by itself.

Types of Unsupervised learning

- a) Clustering

b) Association

3.3 Reinforcement learning

Reinforcement learning is a neighbourhood of Machine Learning. it's about taking suitable action to maximise reward during a particular situation. it's employed by various software and machines to seek out the simplest possible behaviour or path it should absorb a selected situation. Reinforcement learning differs from supervised learning during a way that in supervised learning the training data has the solution key with it therefore the model is trained with the right answer itself whereas in reinforcement learning, there's no answer but the reinforcement agent decides what to try to to perform the given task. Within the absence of a

training dataset, it's sure to learn from its experience.

Types of Reinforcement learning

- a) Positive
- b) Negative

4. Methodology

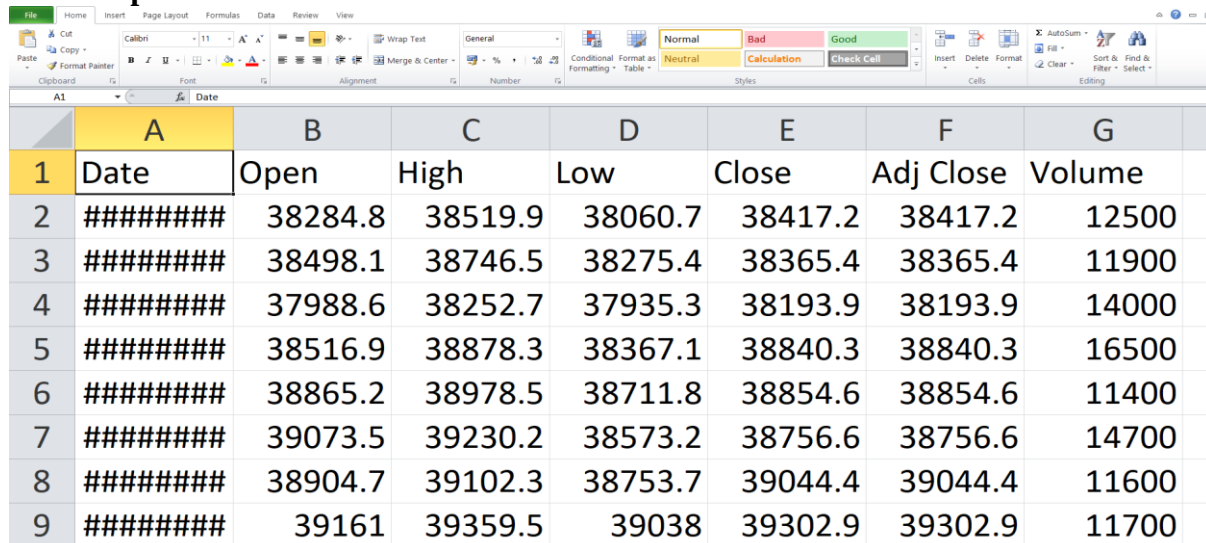
4.1 Description of data

The historical data for the various companies are collected from Yahoo Finance. the info set includes several years of knowledge from date so far of various companies. the info contains information about the stocks like High, Low, Open, Close, Adjacent close and volume. Only the day-wise price of the stock has been extracted. The below table shows statistics of the dataset that's used for training and testing.

Table 1: Statistics of the dataset

	Dataset	Training dataset	Testing dataset
Time Interval	02/06/2015-02/06/2020	02/06/2002-02/06/2014	02/06/2020-02/06/2021

Dataset Example:



	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	#####	38284.8	38519.9	38060.7	38417.2	38417.2	12500
3	#####	38498.1	38746.5	38275.4	38365.4	38365.4	11900
4	#####	37988.6	38252.7	37935.3	38193.9	38193.9	14000
5	#####	38516.9	38878.3	38367.1	38840.3	38840.3	16500
6	#####	38865.2	38978.5	38711.8	38854.6	38854.6	11400
7	#####	39073.5	39230.2	38573.2	38756.6	38756.6	14700
8	#####	38904.7	39102.3	38753.7	39044.4	39044.4	11600
9	#####	39161	39359.5	39038	39302.9	39302.9	11700

4.2 Regression Analysis

Regression analysis may be a statistical procedure to model the connection between a dependent (target) and independent (predictor) variables with one or more independent variables. More specifically, multivariate analysis helps us to know how the worth of the variable is changing

like an experimental variable when other independent variables are held fixed. It can predicts continuous/real values like temperature, age, salary, price, etc.

As mentioned above, multivariate analysis helps within the prediction of endless variable. There are various scenarios within the world where we'd like some future predictions like weather , sales

prediction, marketing trends, etc., for such case we'd like some technology which may make predictions more accurately. So for such case we'd like multivariate analysis which may be a statistical procedure and utilized in machine learning and data science. Below are another reasons for using multivariate analysis:

- Regression estimates relationship between the target and the independent variable.
- It is used to find the trends in data.
- It helps to predict real/continuous values.
- By performing regression, we can confidently determine the most important factor, the least

important factor, and how each factor is affecting the other factors.

Procedure of Work

Step1: Login

Step2: Import the file

Step3: Data Pre-Processing

Step4: Normalization

Step5: Perform operations

Step 6: Result



Figure 1: First Page

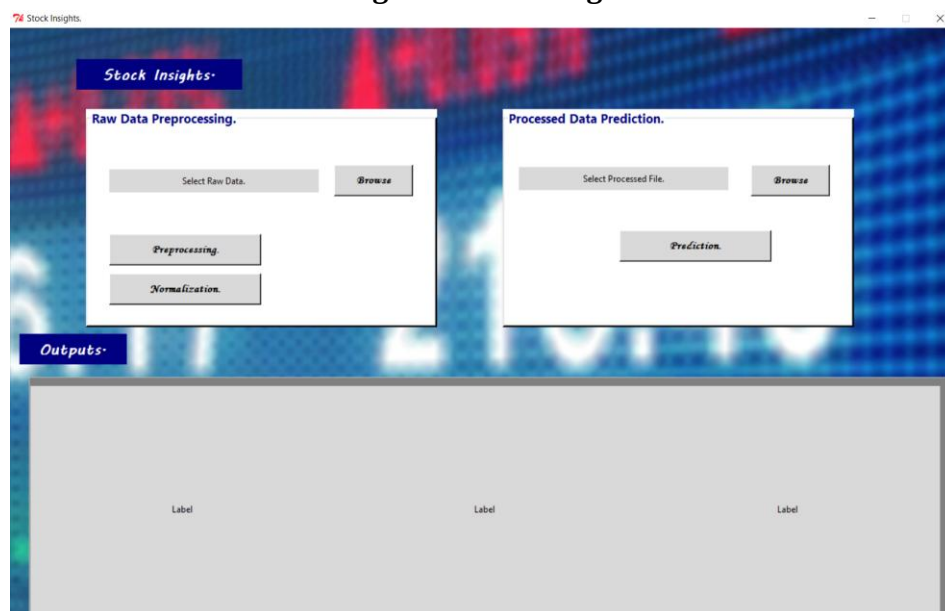


Figure 2: Second page

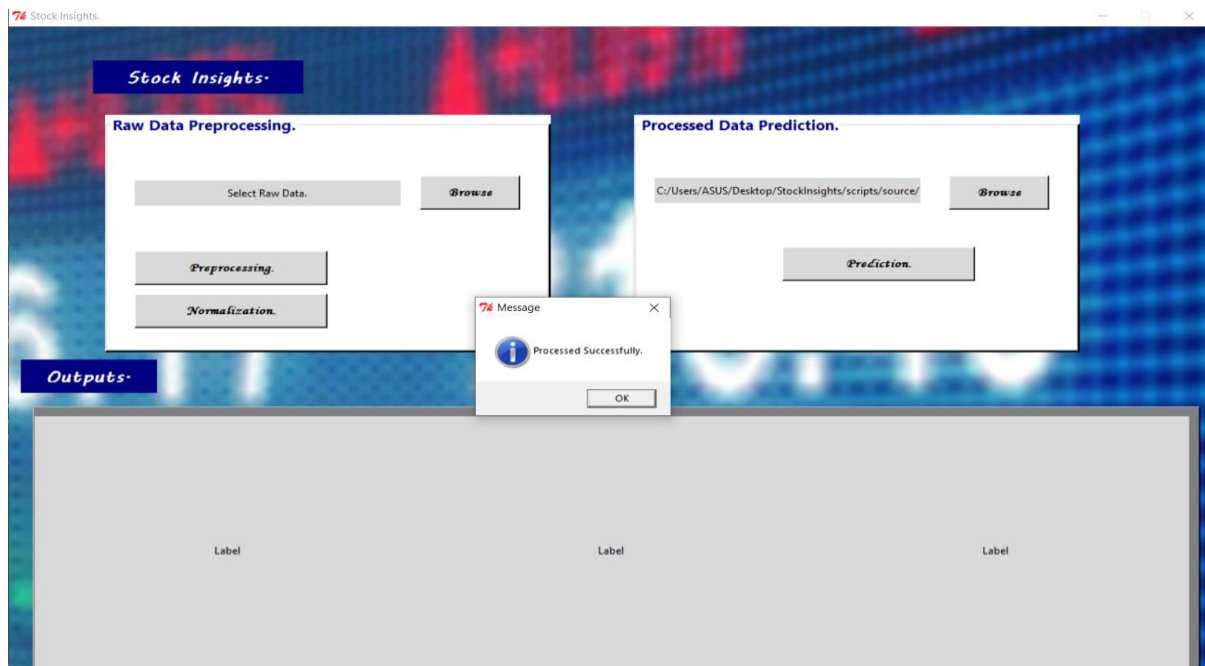


Figure 3: Browsing File

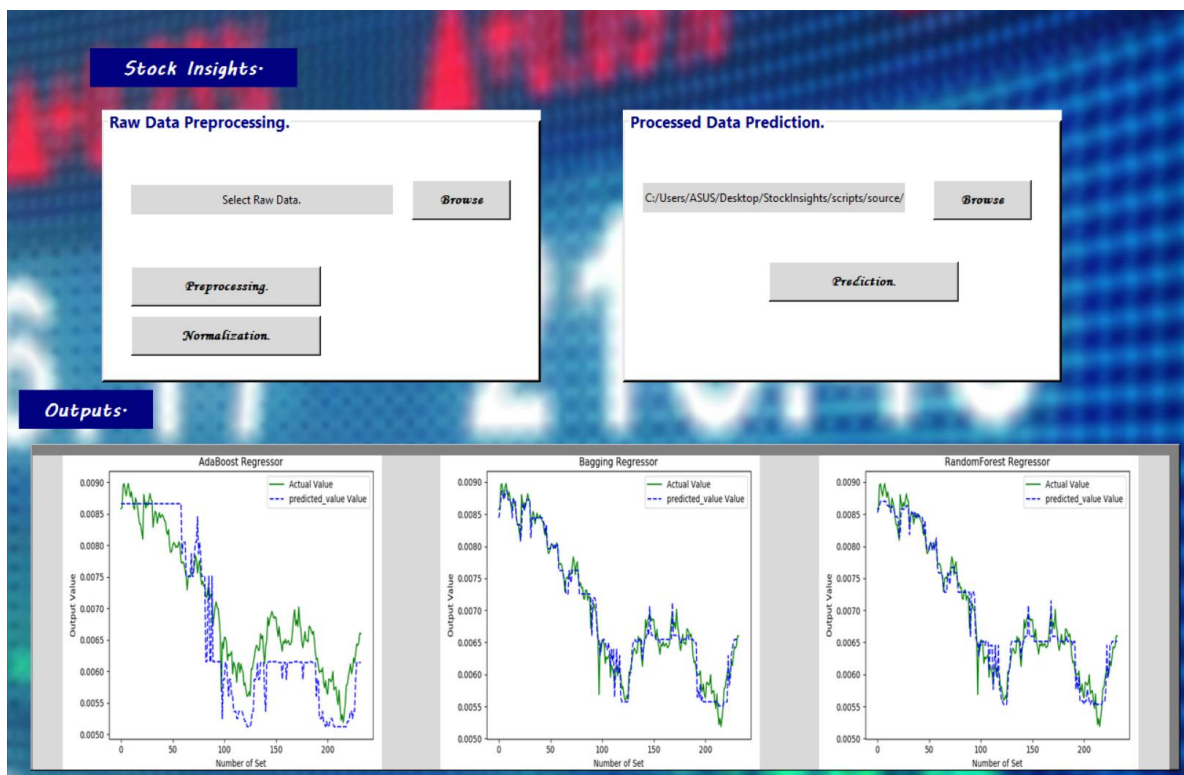


Figure 4: Results

5. Conclusions

Techniques have been utilized during this paper: Regression, on the Yahoo finance dataset. Both the techniques have shown an improvement within the accuracy of predictions, thereby

yielding positive results. Use of recently introduced machine learning techniques within the prediction of stocks have yielded promising results and thereby marked the utilization of them in profitable exchange schemes. it's led to the

conclusion that it's possible to predict stock exchange with more accuracy and efficiency using machine learning techniques.

Acknowledgment

We are really grateful to all the people who directly or indirectly helped us in this research work. We are also thankful to Mr. Arunendra Singh (Assistant Professor, Department of Information Technology, PSIT Kanpur) under whose guidance this research paper has been a success.

References

1. Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, 167, 599-606.
2. Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert systems with applications*, 40(14), 5501-5506.
3. Mehtab, S., & Sen, J. (2020). Stock price prediction using convolutional neural networks on a multivariate timeseries. *arXiv preprint arXiv:2001.09769*.
4. Leung, C. K. S., MacKinnon, R. K., & Wang, Y. (2014, July). A machine learning approach for stock price prediction. In *Proceedings of the 18th International Database Engineering & Applications Symposium* (pp. 274-277).