IJARIIT
INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 2 - V7I2-1525)
Available online at: https://www.ijariit.com

Stock price prediction using machine learning models in python

Kathika Sai Krishna
saikathika011@gmail.com
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu

Mettupalli Hari Naveen Reddy
mettupallinaveenreddy@gmail.com
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu

Appidi Koushik Reddy
appidikoushik2126@gmail.com
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu

Mukkamalla Naveen Reddy
mukkamallanaveen360@gmail.com
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu

ABSTRACT

The project aims to provide retail investors with Machine Learning models to navigate through the stock market. This is achieved through the use of machine learning and Python. Several stock price prediction approaches and models are developed including Recurrent neural networks i.e. LSTM, simple linear regressions, and Decision Tree Models. By taking the past stock prices the models were trained and tested on that data to predict the stock prices.

Keywords: *Stock Price, Machine Learning, Algorithms.*

1. INTRODUCTION

Nowadays,Stocks are possibly the most popular financial instrument invented for building wealth and are the centerpiece of any investment portfolio. The advances in commerce technology has opened the markets in order that these days nearly anybody will own stocks. From past few decades, there seen drastic increase in the average person's interest for stock market. In a eplosive finance market as the stock market, it is important to have a very accurate prediction of a future trend. Because of the financial crisis and recording profits, it is compulsory to have a secure prediction of the values of the stocks.In our research, we are going to use Machine Learning Algorithm specially focus on Linear Regression (LR),Decision Tree and Long Term Short Term Memory(LSTM).Three month Moving Average(3MMA),we used matlab library as best statistical tool for graphical representation of prediction results.

We did took the stocks data from Yahoo Finance for TATA CONSULTANCY SERVICES(TCS) stocks, RELIANCE stocks after implementing Linear Regression,decision tree and long term short term memory(LSTM) we successfully predicted stock market prices for next month and also measured accuracy of prediction according to measurements.

2. LITERATURE SURVEY

From the research paper "Stock Price Prediction Using LSTM"[8] done by Pramod B.S, and Mallikarjuna Shastry P.M. from REVA University, Bengaluru. They used the recurrent neural network(RNN) technique called Long Short Term Memory(LSTM)[7]. This approach considers available historic data of share and it predicts on a particular feature. That feature of shares are opening price, day high, day low, previous day opening price, close price. This model which is done by mallikarjuna shastry[8] and pramod use the time series analysis in order to predict stock price for a particular time span. So, after going through that model one doubt raised and it beacame as our project that why did they used only lstm rather than using other techniques. So, we want to check with other technique and we want to know how they work comparing to lstm. That's why we used linear regression model and decision tree model in addition to lstm. In this feature to predict the stocks we took stocks data, opening and closing price into consideration. We divided the data into two parts in the ratio 8:2 means our models takes 80% data for train the model and 20% of stock data for testing. By this our model predicts the price of future and will represents graphically.

3. EXISTING SYSTEM

From the vast study, the existing system[8] used Long Short Term Memory-Recurring Neural Network for predicting the prices of the stocks. Because of that the outcome is only up to a certain limit. The existing system using the market data to predict the share price using machine learning techniques like recurrent neural network named as Long Short Term Memory[7], in that process weights are corrected for each data points using stochastic gradient descent. This system will provide accurate outcomes in comparison to currently available stock price predictor algorithms. In this the network is trained and evaluated with various sizes of input data to urge the graphical outcomes.

4. PROPOSED SYSTEM

First we need to extract the past data from market i.e. yahoo website using pandas[4] in-built function data reader. Then we need to scale the data and convert them to training and testing datasets. Then train the model with training data and test it with the testing data set and the final step is to visualize the predictions using mat plot library. In the figure 1. System Architecture was shown.

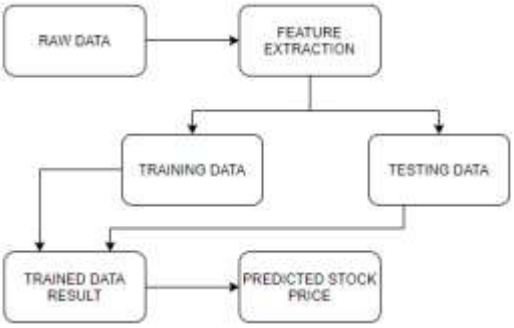


Fig.1: System Architecture

Here we have used three algorithms

Linear Regression: It's a supervised learning technique, The aim of this algorithmic rule is to identify how the input variable influences the output variable. A set of data points inputs and responses are given as x and y. This model tries to fit a straight line that passes through maximum number of points.

Decision Tree: Decision tree builds regression or classification models in the form of a tree structure. It breaks an existing dataset into smaller sub-trees while at the same time an associated decision tree is incrementally developed. The final result is a tree having decision and leaf nodes.

LSTM-RNN: It is a kind of RNN 'Recurrent Neural Network', It uses three types of gates; Input gate: It is used the update the cell information;

Forget gate: It is used to remove the information which is not required from the cells; Output gate: It is used to give the final outputs of LSTM model.

5. IMPLEMENTATION AND RESULTS

5.1. Linear Regression and Decision Tree

All Machine Learning related code are written in Python. In this project we have used four types of machine learning models. First the implementation starts by importing required modules for the project. Required modules were Linear regression, Decision Tree regression and matplot[6] library for creating graphs. After that the dataset is uploaded to the Jupyter environment which was downloaded from the yahoo finance. Then read the data using pandas[4] and store it in a data frame. Test if the data is readable.

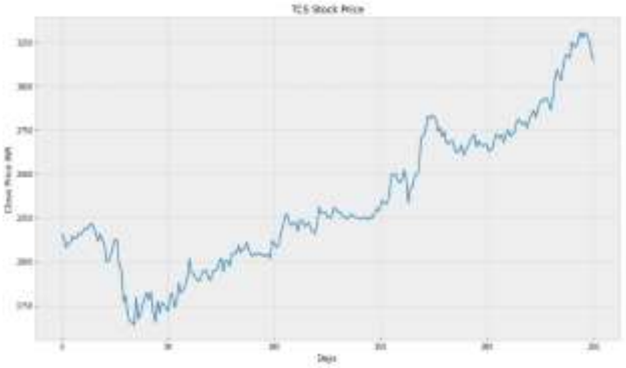


Fig. 2: Close price of historical data

create a variable called 'predicted_days' i.e. future days. It is used to test the model. And the remaining data is to train the model. By using NumPy[3] create an array named 'x' containing the data which is subtracted by the predicted_days. And create another array named 'y' containing the predicted data. Divide the data in the ratio of 8:2. And Train the model using 80% of the data and with remaining 20% of the data is to test the model. Create Linear regression model and Decision tree model and fit it with training and testing data. Create x_future as a variable that is used to store the predicted_days data. Linear regression prediction is done by using x_future variable.

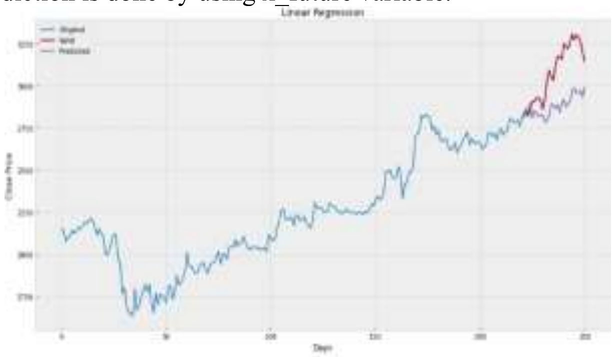


Fig. 3: Predictions of Linear Regression

The above figure i.e. figure 3 contains the values predicted by the Linear-Regression model.

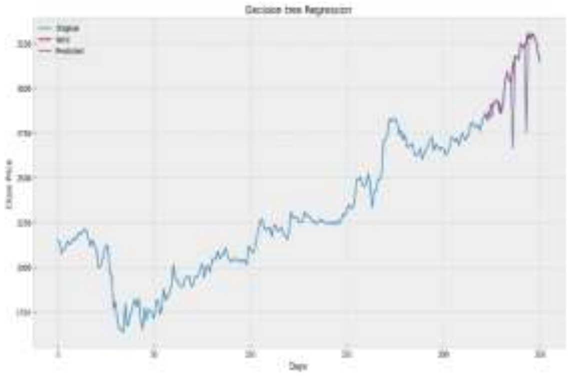


Fig. 4: Predictions of Decision Tree

The above figure i.e. figure 4 contains the values predicted by the Decision Tree model. Based on the outcomes of both the models, Decision tree predictions were better than the Linear regression model predictions.

5.2. LSTM-RNN

The LSTM model implementation starts with by importing required modules for the project. Required modules were MinMaxScaler, LSTM, Sequential, pandas[4] for reading and manipulating the data and matplot[6] library for making plots.

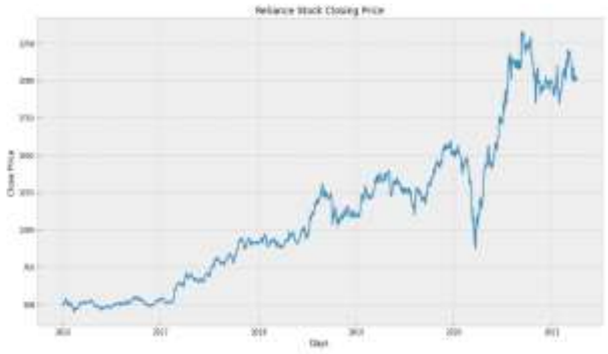


Fig. 5: Actual Close price of historical data

Create a numpy array named dataset with the data as values and create a variable called train_data_length to calculate number of rows and assign it with the 80% of the data. That 80% of data is used to train the LSTM model. Now preprocess the data i.e. scaling the input data before it is added to the RNN, it is a good practice. By using MinMaxScaler method scale the data with feature range of 0, 1. and fit transform the data and store it to the scaled_data. split the data into x_train and y_train lists. By using the for loop append the first 60 days worth of data to the x_train and append the 61st data to y_train. Reshape the x_train data to 3 Dimensional dataset because LSTM expects the data

which is in the 3 Dimensional shape, reshape it by using the numpy array[3] inbuilt function reshape. create LSTM model with 50 neurons and set return_sequences as true because we need another layer and create another layer with 50 neurons and set return_sequences as false and create dense layer with 25 neurons. And compile the model with 'adam' optimizer and add loss function 'mean_squared_error' to measure the loss. Train the model with epochs as 1 i.e. we need our model to train one time through all the data, and batch_size as 1. Then create a dataset for testing the model with 60 days worth of data by creating x_test and y_test lists. Convert the x_test list to numpy array so that we can use it smoothly. And reshape the dataset to 3 Dimensional. Get the Predicted Values and inverse_transform the predicted values so that we can easily understand the data.

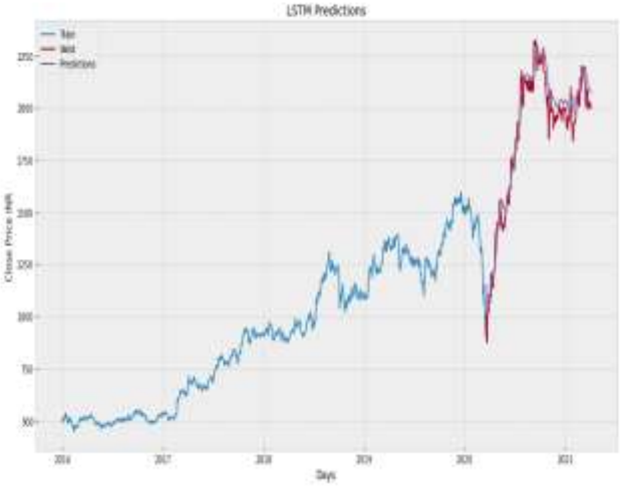
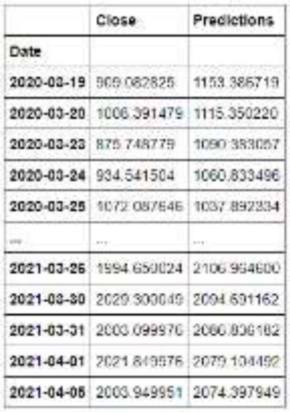


Fig. 6: Visualization of LSTM predictions

The above figure i.e. figure 6 contains the values predicted by the Long Short-term Memory-RNN.



	Close	Predictions
Date		
2020-03-19	909.082825	1153.386719
2020-03-20	1006.391479	1115.350220
2020-03-23	875.748779	1090.393057
2020-03-24	934.541504	1060.833496
2020-03-25	1072.087646	1037.892334
...	...	...
2021-03-26	1994.650024	2106.954600
2021-03-30	2029.300049	2094.691162
2021-03-31	2003.099976	2080.836182
2021-04-01	2021.849976	2079.104492
2021-04-05	2003.949951	2074.397949

Fig. 7: Comparing Actual Prices with the Predicted prices.

The above figure i.e. figure 7 contains the comparison between the Actual Close price and Predicted values.

6. CONCLUSION

Price predictions done by machine learning models, can be used by the stock investors though it is not 100% accurate but it can be used as a reference. Based on the outcomes of all the models LSTM-RNN model preforms better than the Linear regression and Decision tree. The prediction can more reliable if we train the models with large data sets.

In future enhancement the inclusion of various other Time series models such as ARIMA, Autoregression, VARMA etc. To increase the accuracy of predictions and by comparing various model outcomes.

7. REFERENCES

[1] Stock Price Prediction App using Machine Learning Models Optimized by Evolution By CHAUTsunMan. https://www.cse.ust.hk/~rossiter/fyp/103_RO4_Final_201819.pdf

[2] Documentation of scikit-learn: http://scikit-learn.org/stable/documentation.html

[3] Numpy manual: https://numpy.org/doc/stable/user/

[4] pandas: powerful Python data analysis toolkit : https://pandas.pydata.org/docs/

[5] Colaboratory-Google: https://colab.research.google.com./

[6] matplotlib-documentation: https://matplotlib.org/stable/contents.html

[7] Long-Short-term-Memory-RNN: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[8] Stock Price Prediction Using LSTM By Pramod, Mallikarjuna Shastry Pm, Reva university: https://www.researchgate.net/publication/348390803_Stock_Price_Prediction_Using_LSTM