

# **Stock Price Prediction Using Machine Learning**

By

20BCE10976 : Divyanshu Jain

Student at VIT Bhopal

Guided by

Moses K Sir

Indra Sir

A Project Report Submitted to

“SmartInternz”

in Partial Fulfilment of the Requirements for  
the “Applied Data Science” Internship Programme

July 2023



**VIT<sup>®</sup>**  
**B H O P A L**  
[www.vitbhopal.ac.in](http://www.vitbhopal.ac.in)



## Abstract

Researchers have been studying different methods to effectively predict the stock market price.

Useful prediction systems allow traders to get better insights about data such as: future trends.

Also, investors have a major benefit since the analysis give future conditions of the market.

One such method is to use machine learning algorithms for forecasting. This project's objective is to improve the quality of output of stock market predicted by using stock value.

A number of researchers have come up with various ways to solve this problem, mainly there are traditional methods so far, such as artificial neural network is a way to get hidden patterns and classify the data which is used in predicting stock market.

This project proposes a different method for prognosting stock market prices. It does not fit the data to a specific model; rather we are identifying the latent dynamics existing in the data using machine learning architectures.

We attempt to implement machine learning approach to predict stock prices. Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades. There are two types of stocks. You may know of intraday trading by the commonly used term "day trading." Interday traders hold securities positions from at least one day to the next and often for several days to weeks or months. LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down.

## **Chapter 1. Introduction**

### **1.1 Stock Price Prediction**

Due to the high profit of the stock market, it is one of the most popular investments. People investigated for methods and tools that would increase their gains while minimizing the risk, as the level of trading and investing grew. Two stock exchanges namely- the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE), which are the most of the trading in Indian Stock Market takes place. Sensex and Nifty are the two prominent Indian Market Indexes. Since the prices in the stock market are dynamic, the stock market prediction is complicated.

From gradually the very past years some forecasting models are developed for this kind of purpose and they had been applied to money market prediction. Generally, this classification is done by:

1. Time series analysis
2. Fundamental analysis
3. Technical analysis

### **Time Series Analysis**

The definition of forecasting can be like this the valuation of some upcoming result or results by analysing the past data. It extents different areas like industry and business, economics and finance, environmental science. Forecasting problems can be classified as follows:

- ☐ Long term forecasting (estimation beyond 2 years)
- ☐ Medium-term forecasting (estimation for 1 to 2 years)
- ☐ Short term forecasting (estimation for weeks or months, days, minutes, few seconds)

The analysis [1] of time consist of several forecasting problems. The designation of a time series is a linear classification of observations for a selected variable. The variable of the stock price in our case. Which can weather multivariate or univariate? Only particular stock is included in the univariate data while more than one company for various instances of time is added in multivariate. For investigating trends, patterns and cycle or periods the analysis of time series advantages in the present data. In spending money wisely an early data of the bullish or bearish in the case of the stock market. Also, for categorizing the best-performing companies the analysis of patterns plays its role for a specific period. This makes forecasting as well as time series analysis an important research area.

## **Fundamental analysis**

Fundamental Analysts are concerned with the business that reasons the stock itself. They assess a company's historical performance as well as the reliability of its accounts. Different performance shares are created that aid the fundamental forecaster with calculating the validity of a stock, such as the P/E ratio. Warren Buffett is probably the foremost renowned of all Fundamental Analysts.

What fundamental analysis within the stock market is making an attempt to reach, is organizing the true value of a stock, that then will be matched with the worth it is being listed on stock markets and so finding out whether or not the stock on the market is undervalued or not. Find out the correct value will be completed by numerous strategies with primarily a similar principle. The principle is that an organization is price all of its future profits. Those future profits has to be discounted to their current value. This principle goes on the theory that a business is all about profits and nothing else. Differing to technical analysis, the fundamental analysis is assumed as further as a long approach.

Fundamental analysis is created on conviction that hominoid society desires capital to make progress and if the company works well, than it should be rewarded with an additional capital and outcome in a surge in stock price. Fundamental analysis is usually used by the fund managers as it is the maximum sensible, objective and prepared from openly existing data like financial statement analysis.

One more meaning of fundamental analysis is on the far side bottom-up business analysis, it discusses the top-down analysis since initial analysing the world economy, followed by country analysis and also sector analysis, and last the company level analysis.

## **Technical analysis**

Chartists or the technical analysts are not involved with any other of the fundamentals of the company. The long run price of a stock based generally exclusively on the trends of the past value (a form of time series analysis) that is set by them. The head and shoulders or cup and saucer are various numerous patterns that are employed. Also the techniques, patterns are used just like the oscillators, exponential moving average (EMA), support and momentum and volume indicators. Candlestick patterns, believed to have been initial developed by Japanese rice merchants, are nowadays widely used by technical analysts. For the short-term approaches, the technical analysis is used compare to long-run ones. So, in commodities and forex markets it is more predominant wherever traders target short-term price movements. There are basic rules are used in this analysis, first all significant about a company is already priced into the stock, another being that the value changes in trends and finally that history (of prices) tends to repeat itself that is especially due to the market science.

## **1.2 Applications**

- Business
- Companies
- Insurance company
- Government Agency
- This application is helpful for stock investors, sellers, buyers, brokers.

### **1.3 Objectives**

A stock market prediction is described as an action of attempting to classify the future value of the company stock or other financial investment traded on the stock exchange. The forthcoming price of a stock of the successful estimation is called the Yield significant profit. This helps you to invest wisely for making good profits.

### **1.4 Motivation**

The future price of a stock is the main motivation behind the stock price prediction. In various cases like business and industry, environmental science, finance and economics motivation can be useful. The future value of the company's stock can be determining.

## **Chapter 2. Tool & Technologies**

### **2.1 PYTHON**

The language of select for this project was Python. This was a straightforward call for many reasons.

1. Python [19] as a language has a vast community behind it. Any problems which may be faced is simply resolved with visit to Stack Overflow. Python is the foremost standard language on the positioning that makes it is very straight answer to any question.
2. Python [19] is an abundance of powerful tools ready for scientific computing Packages. The packages like NumPy, Pandas and SciPy area unit freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast.
3. Python is a language as [19] forgiving and permits for the program that appear as if pseudo code. This can be helpful once pseudo code give in tutorial

papers should be required and verified. Using python this step is sometimes fairly trivial.

However, Python is [19] not without its errors. The python is dynamically written language and packages are area unit infamous for Duck writing. This may be frustrating once a package technique returns one thing that, for instance, looks like an array instead of being an actual array. Plus the standard Python documentation did not clearly state the return type of a method, this can't lead without a lot of trials and error testing otherwise happen in a powerfully written language. This is a problem that produces learning to use a replacement Python package or library more difficult than it otherwise may be.

## **2.2 NUMPY**

Numpy is python package which provide scientific and higher level mathematical abstractions wrapped in python. It is [20] the core library for scientific computing, that contains a provide tools for integrating C, strong n-dimensional array object, C++ etc. It is also useful in random number capability, linear algebra etc.

Numpy's array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy additionally provides basic numerical routines, like tools for locating Eigenvectors.

## **2.3 SCIKIT LEARN**

Scikit-learn [21] could be a free machine learning library for Python. It features numerous classification, clustering and regression algorithms like random forests, k-neighbours, support vector machine, and it furthermore supports Python scientific and numerical libraries like SciPy and NumPy.

In Python Scikit-learn is specifically written, with the core algorithms written in Cython to get the performance. Support vector machines are enforced by a

Cython wrapper around LIBSVM .i.e., linear support vector machines and logistic regression by a similar wrapper around LIBLINEAR.

## **2.4 TENSORFLOW**

In the TensorFlow [22] has an open source software library for numerical computation using data flow graphs. Inside the graph nodes represent mathematical formulae, the edges of graph represent the multidimensional knowledge arrays (tensors) communicated between them. The versatile architecture permits to deploy the computation to at least one or many GPUs or CPUs in a desktop, mobile device, servers with a single API. TensorFlow was firstly developing by engineers and researchers acting on the Google Brain Team at intervals Google's Machine Intelligence analysis organization for the needs of conducting deep neural networks research and machine learning, but, the system is generally enough to be appropriate in a wide range of alternate domains as well.

Google Brain's second-generation system is TensorFlow. Whereas the reference implementation runs on single devices, TensorFlow can run on multiple GPUs and CPUs. TensorFlow is offered on Windows, macOS, 64-bit Linux and mobile computing platforms together with iOS and Android

## **2.5 KERAS**

Keras is [23] a high-level neural networks API, it is written in Python and also capable of running on top of the Theano, CNTK, or. TensorFlow. It was developed with attention on enabling quick experimentation. having the ability to travel from plan to result with the smallest amount doable delay is key to doing great research. Keras permits for straightforward and quick prototyping (through user-friendliness, modularity, and extensibility). Supports each recurrent networks and convolutional networks, also as combinations of the 2. Runs seamlessly on GPU and CPU. The library contains numerous implementations of generally used neural network building blocks like optimizers, activation functions, layers, objectives and a number of tools to create operating with text and image data



easier. The code is hosted on GitHub, and community support forums embody the GitHub issues page, a Gitter channel and a Slack channel.

## **2.6 COMPILER OPTION**

Anaconda is [19] free premium open-source distribution of the R and Python programming languages for scientific computing, predictive analytics, and large-scale process that aim is to modify package managing and deployment. Package versions unit managed by the package management system conda.

## **2.7 JUPITER NOTEBOOK**

The Jupyter Notebook is an open-source web application that enables to making and sharing documents that contain visualizations, narrative text, live code and equations. Uses include: data , data visualization, data transformation, statistical modelling, machine learning, numerical simulation, data cleaning and much more [24].

## **Chapter 3. Website Page and explanation**

First of all, we have a simple text input field where the user has to enter the stock ticker. Stock ticker is nothing but the symbol that is assigned to that stock in the market. For Tesla we have TSLA.

The next thing we are having the data description which is printed using `df.describe` function.

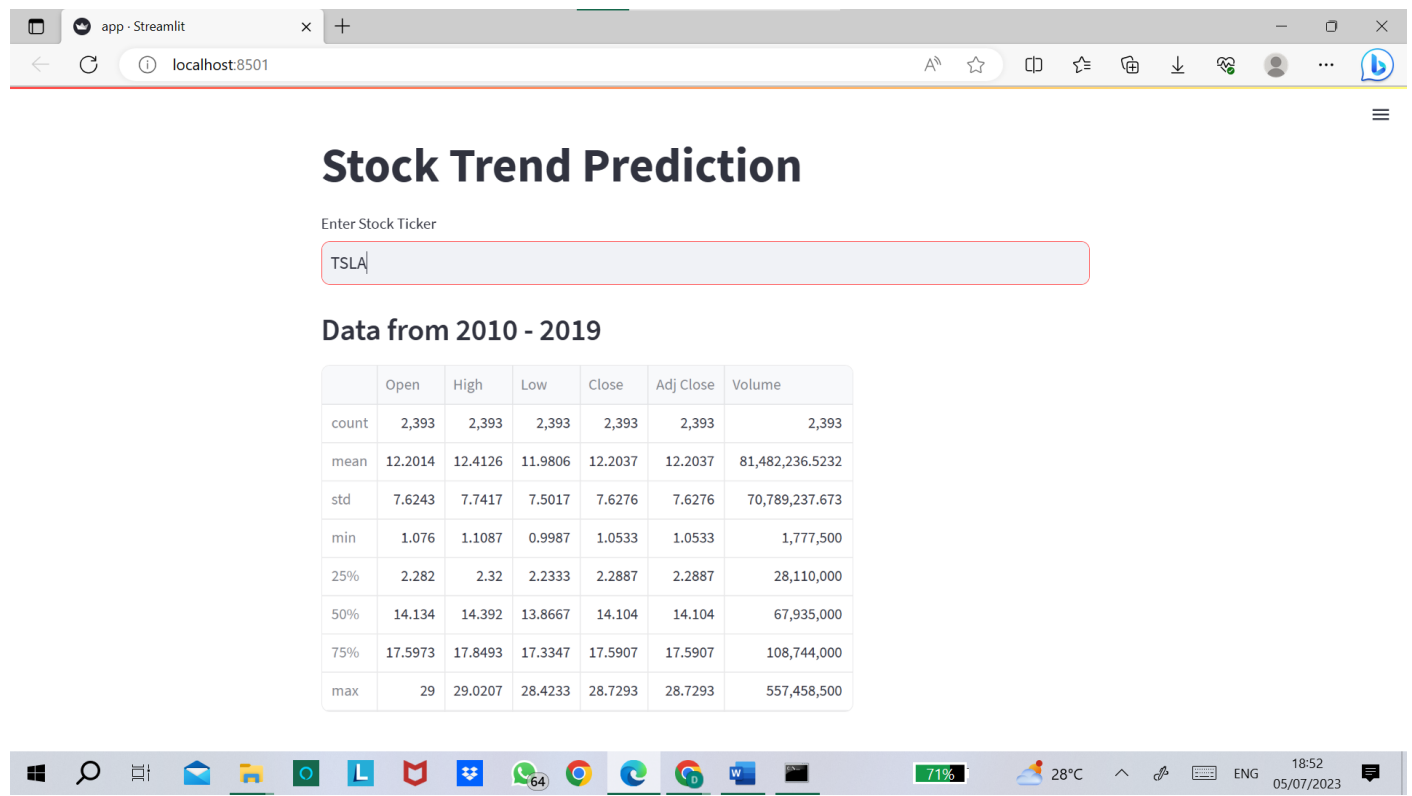
I have chosen the data from 2010 to 2019 and I have divided this complete dataset into two parts that is training and testing.

Training set I have decided to be 70% and testing part to be 30%.

So, my final prediction is going to be on testing part only.

And we are going to plot a simple graph which will be showing me the predicted values and the actual values.

### 3.1

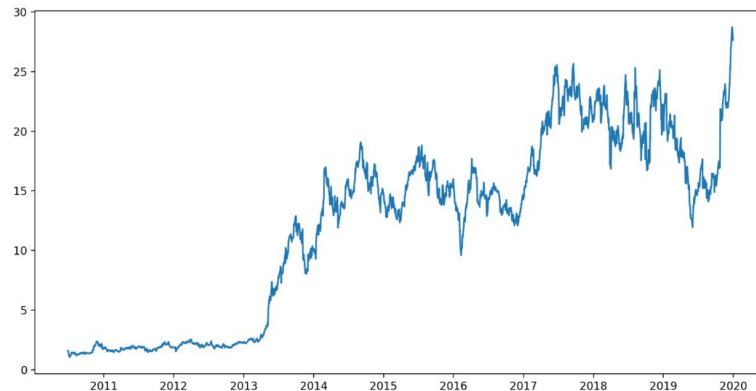


In this table we are going to see the description of the data that is high, low, open and close for a particular day. In the high we can see that the maximum high in those around 10 years, we are having 87 dollars. In the low we can see the minimum low as 2.99 dollars. Similarly, we can have a simple description of the data between 2010 and 2019.

## 3.2



Closing Price vs Time Chart



Closing Price vs Time Chart with 100 MA



This is the simple visualisation of the closing price. Here we will be working on this closing column because we are interested in finding or predicting the closing price for a particular day. Therefore, this is the Closing Price versus Time Chart.

## 3.3



Closing Price vs Time Chart with 100 MA



Closing Price vs Time Chart with 100 MA and 200 MA

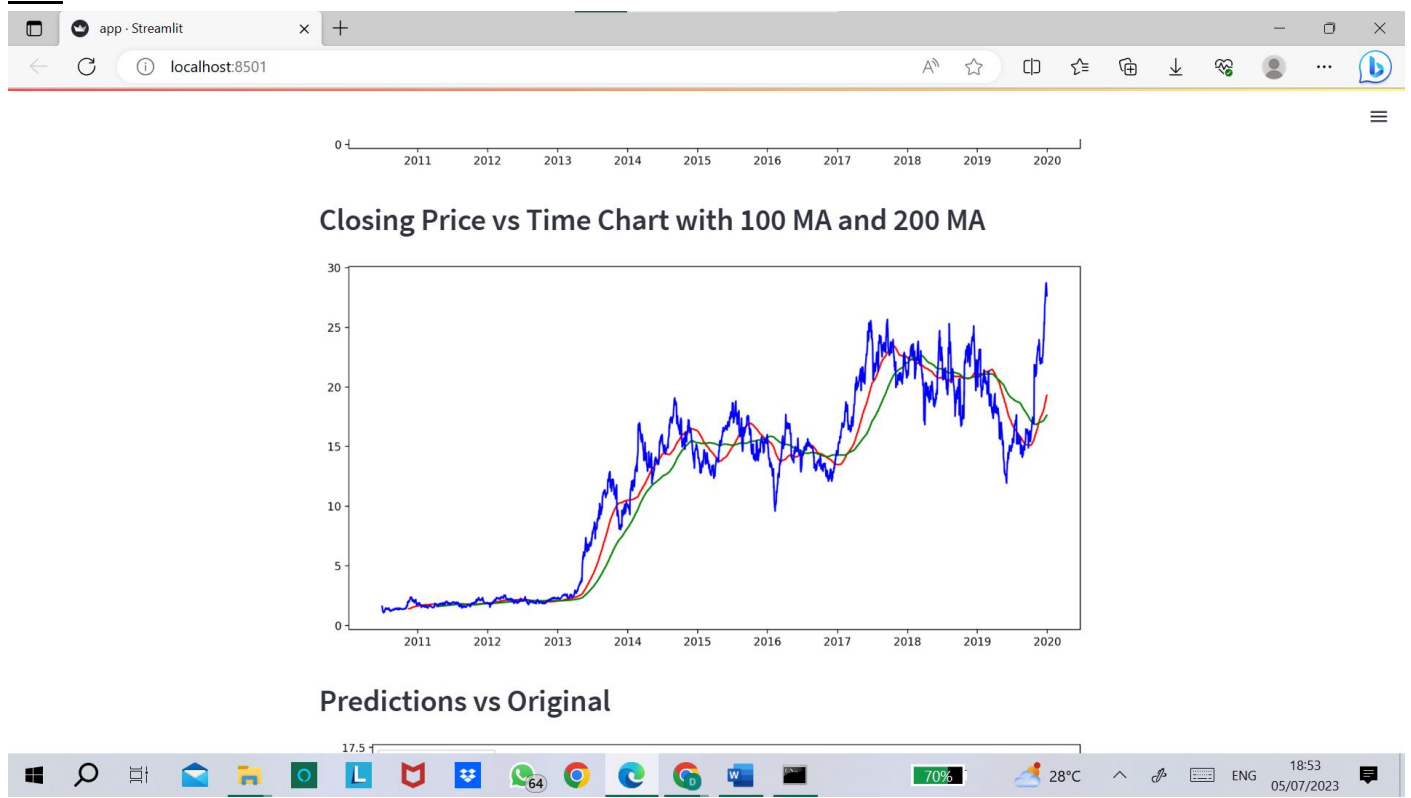


Let me also give you a quick introduction about the moving averages. This red line which you can see is a moving average that is 100 day moving average.

A 100 day moving average is nothing but it is plotted a way such that it will take the average of the previous 100 days closing price, and it will get a value, and it will plot that on the graph.

As you can see over here that, it is not started from the start because it will start after getting the 100 values.

### 3.4



Technical analysts in the markets follow this strategy that - If 100 day moving average crosses above the 200 day moving average, then there is an uptrend, and if 100 day moving average crosses below the 200 day moving average, then there is a downtrend.

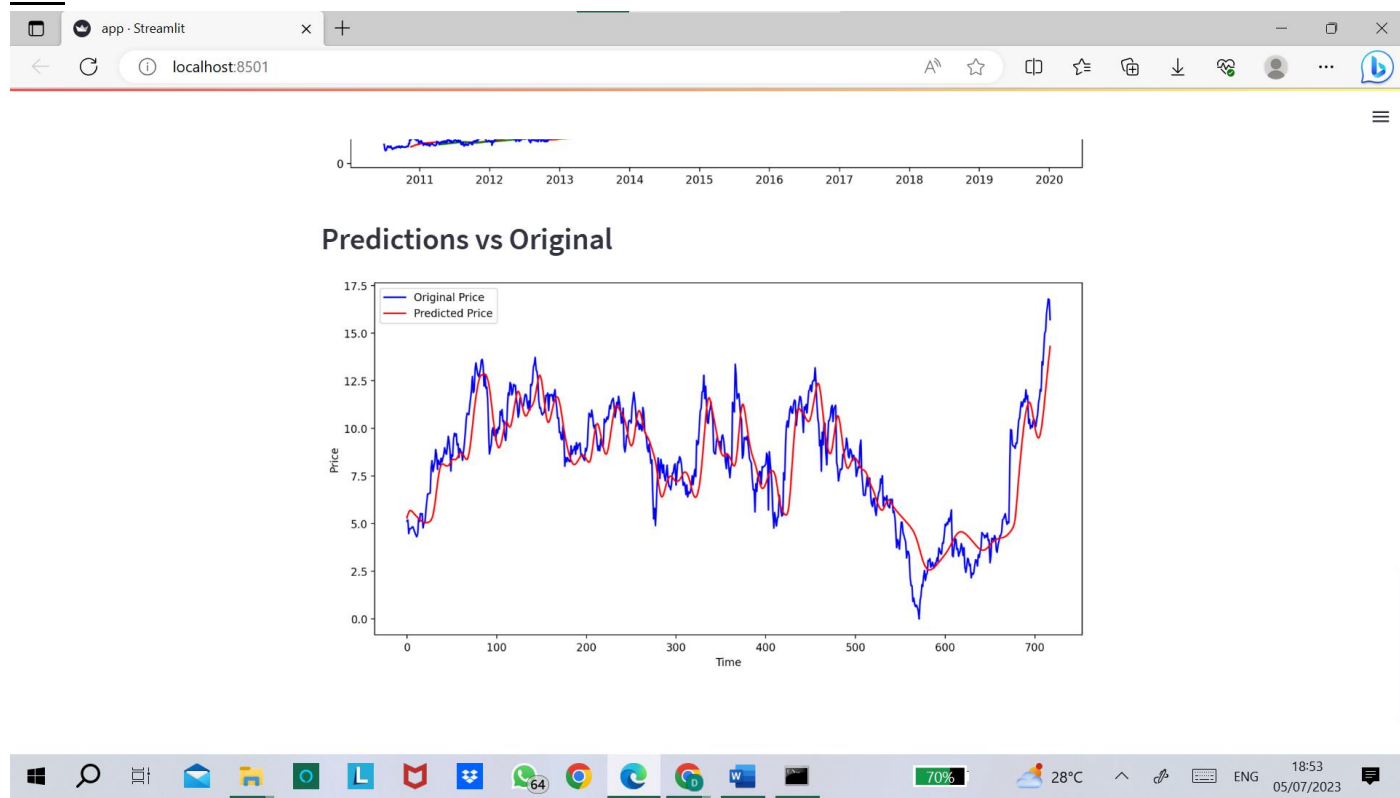
As we can see here, this red line is my 100 day moving average and this green line is my 200 day moving average.

Whenever this red line will cross above the green line, it is the starting of an uptrend.

And whenever, the red line will cross below the green line, it is the starting of a downtrend.

This is a very simple strategy that technical analysts follow in the market to predict or we can say to analyse a stock.

### 3.5



This is the Predictions versus Actual Chart

In the legend, the red line is my original price and the green line is the predicted price.

So, we can have a very good general idea about how my stock is performing in these days.

Here, we can see a simple offset. Offset is nothing but a simple difference between the original and the predicted price, or the gap, and we are having a very less offset, and we are getting very good results, so its totally fine.

Similarly we can see the predictions for any stock in this manner.

## **CHAPTER 4. CONCLUSION AND FUTURE WORK**

### **4.1 Conclusion**

In this project, we are predicting closing stock price of any given organization, we developed a web application for predicting close stock price using LSTM algorithms for prediction.

### **4.2 Future work**

- We want to extend this application for predicting cryptocurrency trading.
- We want to add sentiment analysis for better analysis.