Visualization and Forecasting Stocks

Dheeraj Parmar 20190802113 | Akshat Sinha 20190802012

D.Y Patil International University

TITLE

Visualizing and Forecasting Stocks using Python

PREFACE

The report has been made in fulfillment of the requirement for the subject : Stock Prediction and Visualization in August 2021 under the supervision of Dr. Bahubali Shiragapur by Dheeraj Parmar and Akshat Sinha.

For making this project we have studied various concepts related to the stock market and how they can be used. We also studied various Machine Learning Algorithms and tools that can be used to solve the problem easily.

The project aims to apply an algorithm; ARIMA and analyze this algorithm to predict the stock market.

ABSTRACT

The prediction of a stock market direction may serve as an early recommendation system for short term investors and as an early financial distress warning system for long term shareholders. Forecasting accuracy is the most important factor in selecting any forecasting methods. Research efforts in improving the accuracy of forecasting models have been increasing since the last decade. The appropriate stock selections that are suitable for investment is a very difficult task. The key factor for each investor is to earn maximum profits on their investments.

In this research project, **Autoregressive Integrated Moving Average (ARIMA)** is used. We investigate the predictability of financial movement with ARIMA. To evaluate the forecasting ability of ARIMA, we compare its performance with decision trees and LSTM models earlier.

This model is applied on 1 year of data retrieved from Yahoo Finance and the results will be used to analyze the stock prices and their prediction in depth in future research efforts.

INTRODUCTION

OBJECTIVE:

In the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers. The objective of the proposed work is to study and improve the given task with supervised learning algorithms to predict the stock prices.

The general research associated with the stock or share market is highly focused on neither buy nor sell but it fails to address the dimensionality and expectancy of a new investor. The common trend towards the stock market among the society is that it is highly risky for investment or not suitable for trade so most of the people are not even interested. The seasonal variance and steady flow of any index will help both existing and naïve investors to understand and make a decision to invest in the stock/share market.

To solve these types of problems, the time series analysis will be the best tool for forecasting the trend or even future. The trend chart will provide adequate guidance for the investor.

In this project, we will work with historical data about the stock prices of a publicly listed company. We will implement a machine learning algorithm to visualize and predict the future stock price of this company, starting with simple algorithms like Averaging and splitting data, and then move on to advanced techniques like ARIMA.

So let us understand this concept in great detail and use a machine learning technique to forecast stocks.

About Stocks: The stock market is a market that enables the seamless exchange of buying and selling of company stocks. Every Stock Exchange has its own Stock Index value. The index is the average value that is calculated by combining several stocks. This helps in representing the entire stock market and predicting the market's movement over time. The stock market can have a huge impact on people and the

country’s economy as a whole. Therefore, predicting the stock trends in an efficient manner can minimize the risk of loss and maximize profit.

How does the stock market work?

The concept behind how the stock market works is pretty simple. Operating much like an auction house, the stock market enables buyers and sellers to negotiate prices and make trades. The stock market works through a network of exchanges — you may have heard of the New York Stock Exchange, Nasdaq or Sensex. Companies list shares of their stock on an exchange through a process called an initial public offering or IPO. Investors purchase those shares, which allows the company to raise money to grow its business. Investors can then buy and sell these stocks among themselves, and the exchange tracks the supply and demand of each listed stock. That supply and demand help determine the price for each security or the levels at which stock market participants — investors and traders — are willing to buy or sell.

Note: Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction — physical factors vs. physiological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.

Literature Review

Definition of the Problem: Stock market attracts thousands of investors' hearts from all around the world. The risk and profit of it has great charm and every investor wants to book profit from that. People use various methods to predict market volatility, such as

K-line diagram analysis method, Point Data Diagram, Moving Average Convergence Divergence, even coin tossing, fortune telling, and so on. Now, all the financial data is stored digitally and is easily accessible. Availability of this huge amount of financial data in digital media creates appropriate conditions for data mining research. The important problem in this area is to make effective use of the available data.

Theoretical Background of the Problem: Stock market is highly volatile. At the most fundamental level, it is said that supply and demand in the market determines stock price. But, it does not follow any fixed pattern and is also affected by a large number of highly varying factors. The investors on Wall Street are split into two largest factions of adherents; those who believe the market cannot be predicted and those who believe the market can be beaten.

Related Research to solve the Problem: Recently, a lot of interesting work has been done in the area of applying Machine Learning Algorithms for analyzing price patterns and predicting stock price. Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions. Recent research uses input data from various sources and multiple forms. Some systems use historical stock data, some use financial news articles, some use expert reviews while some use a hybrid system which takes multiple inputs to predict the market. Also, a wide range of machine learning algorithms are available that can be used to design the system. These systems have different approaches to solve the problem. Some systems perform mathematical analysis on historic data for prediction while some perform sentiment analysis on financial news articles and expert reviews for

prediction. However, because of the volatility of the stock market, no system has a perfect or accurate prediction.

Advantages: The research helps a lot of new investors in deciding when to buy or sell a particular stock. It also helps in understanding the sentiments of experienced financial analysts and financial news data more quickly than doing the same manually.

Our Solution to solve this Problem: We will implement the system using a machine learning technique. We have used the ARIMA . We will train the system using 75% of 1 year of historic data and then test our model to check the system yields better output using the remaining 25% of historic data.

Algorithm

Machine learning in stock market:

Stock and financial markets tend to be unpredictable and even illogical. Due to these characteristics, financial data should be necessarily possessing a rather turbulent structure which often makes it hard to find reliable patterns. Modeling turbulent structures requires machine learning algorithms capable of finding hidden structures within the data and predicting how they will affect them in the future. The most efficient methodology to achieve this is Machine Learning and Deep Learning. Deep learning can deal with complex structures easily and extract relationships that further increase the accuracy of the generated results.

Machine learning has the potential to ease the whole process by analyzing large chunks of data, spotting significant patterns and generating a single output that navigates traders towards a particular decision based on predicted asset prices. Stock prices are not randomly generated values; instead they can be treated as a **discrete-time series model** which is based on a set of well-defined numerical data items collected at successive points at regular intervals of time. Since it is essential to identify a model to analyze trends of stock prices with adequate information for decision making, we felt that transforming the time series using **ARIMA is a better algorithmic approach** than forecasting directly, as it gives more authentic and reliable results.

The Autoregressive Integrated Moving Average (ARIMA) Model converts **non-stationary data to stationary data** before working on it. It is one of the most popular models to predict linear time series data.

ARIMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short-term share market prediction.

It is specified by three ordered parameters (p,d,q), where:

p is the order of the autoregressive model(number of time lags)

d is the degree of differencing (number of times the data have had past values subtracted)

q is the order of moving average.

Before building an ARIMA model, we have to make sure our data is stationary.

So, For a data to be stationarize:

The mean of the series should not be a function of time.

The variance of the series should not be a function of time.

the covariance of the i th term and the (i + m) th term should not be a function of time.

Because when running a linear regression the assumption is that all of the observations are all independent of each other. In a time series, however, we know that observations are time dependent. It turns out that a lot of nice results that hold for independent random variables hold for stationary random variables. So by making the data stationary, we can actually apply regression techniques to this time dependent variable.

There are two methods to check the stationarity of a time series. The first is by looking at the data. By visualizing the data it should be easy to identify a changing mean or variation in the data. For a more accurate assessment there is the Dickey-Fuller test. We have implemented the graphs for the mean and all others to visualize and check the stationarity of the data plus also implemented the dickey-fuller test. You will see all test in the next section.

Discussion on implementation

Implementing stock price forecasting: The dataset consists of stock market data of Tata Motors Limited. We have taken the data from yahoo finance for a particular period of time i.e. 1 year of data is used here.

The goal is to train an ARIMA model with optimal parameters that will forecast the closing price of the stocks on the test data.

The profit or loss calculation is usually determined by the closing price of a stock for the day, hence we will consider the closing price as the target variable.

So start with

1. Loading all the required libraries.
2. Load the dataset.
3. Visualize the closing price of the stock.
4. Let us plot the scatter plot.

Also, a given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise.

First, we need to check if a series is stationary or not because we have already seen that time series analysis only works with stationary data.

First we will check through the **ADF (Augmented Dickey-Fuller) Test:**

The Dickey-Fuller test is one of the most popular statistical tests. It can be used to determine the presence of unit root in the series, and hence help us understand if the series is stationary or not.

If we fail to reject the null hypothesis of this test, we can say that the series is non-stationary. This means that the series can be linear or different stationary.

Also, If both **mean and standard deviation are flat lines** (constant mean and constant variance), the **series becomes stationary**.

So now let's check for stationarity.

Through the graph that we have implemented in the code, we saw the increasing mean and standard deviation and hence our series is not stationary.

Now we are going to create an ARIMA model and will train it with the closing price of the stock on the train data. So let us split the data into training and test sets and visualize it.

It's time to choose parameters p,q,d for the ARIMA model, so we are going to use Auto ARIMA to get the best parameters.

Okay, now let's **start forecasting the stock prices:**

Start forecasting the stock prices on the test dataset keeping 95% confidence level.

As you can see in the graph that we have implemented in the code, our model did quite handsomely. Let us also check the commonly used accuracy metrics to judge forecast results.

MSE: 0.03330921053066402

MAE: 0.13801238336759786

RMSE: 0.18250811086267923

MAPE: 0.035328833278944705

Here are the above output results.

Output Generation: The application reads the input as an input data file and applies the prediction algorithm to it to generate the output. The output consists of the respective graphs, i.e. Closing Price, Scatter Plot, Mean and Standard Deviation, Moving Average, Test and Train of closing price and finally the Prediction of the Stock. All the output is generated for the input stock company's data.

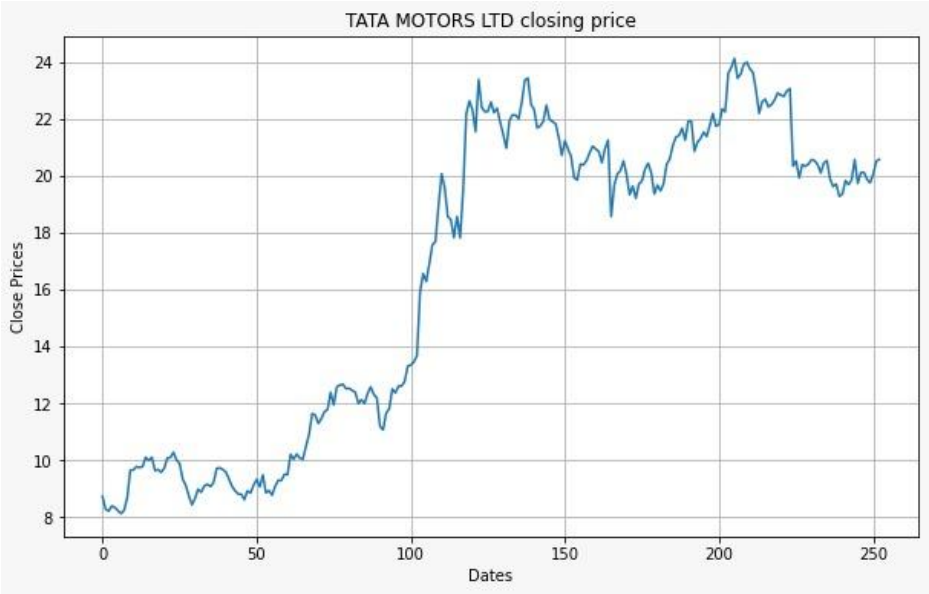
Many indicator functions and their permutations were tested while training and testing the system. Of all the indicator functions instead, the ones which gave the best prediction result were selected. The system performs very well for the prediction of the selected stocks.

Around 3.5% MAPE(Mean Absolute Percentage Error) implies the model is about 96.5% accurate in predicting the test set observations.

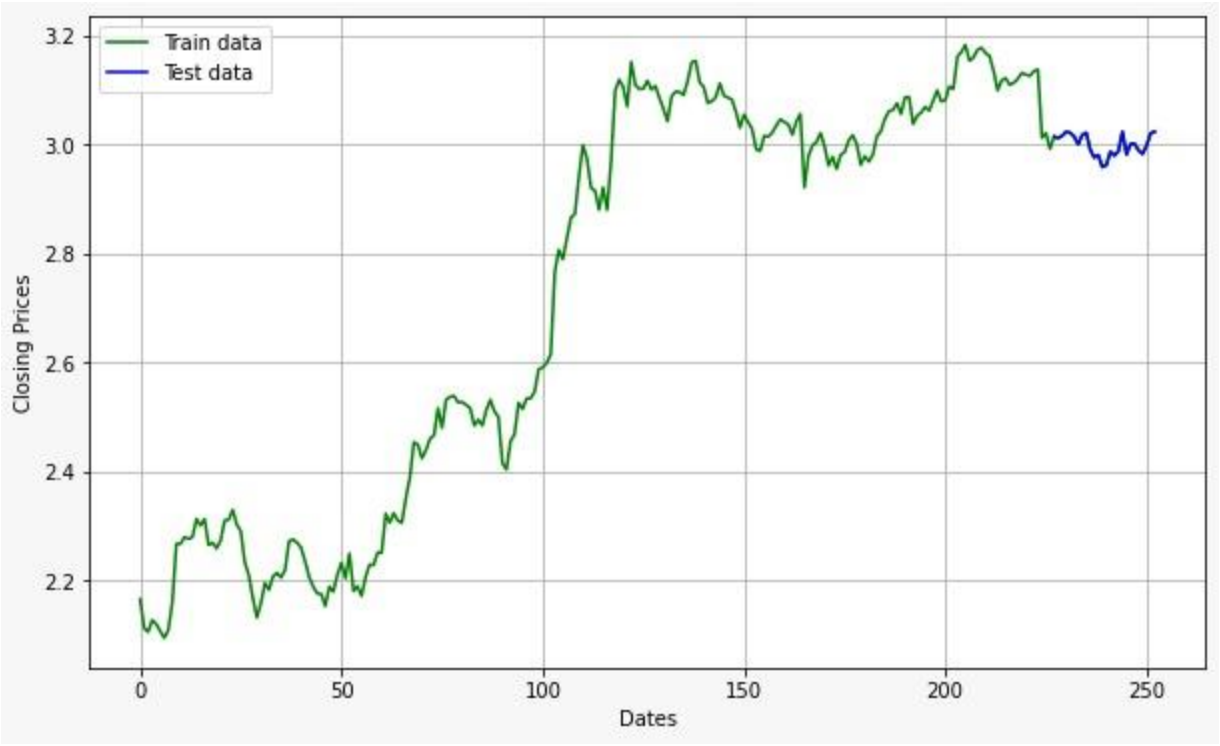
Testing and Error Debugging

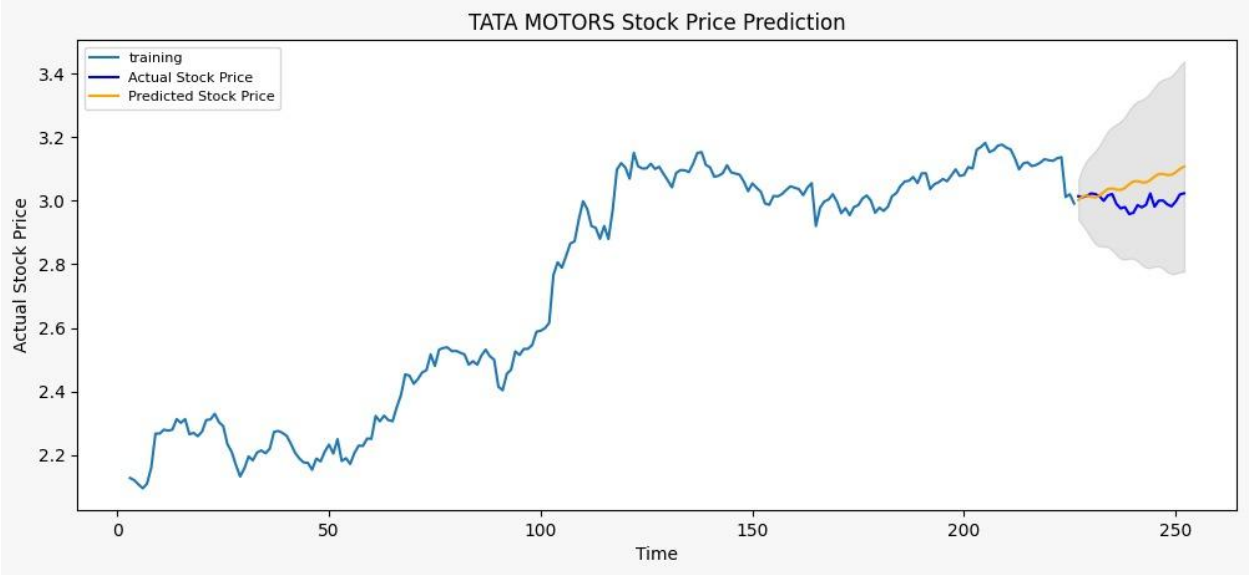
We faced too many errors in implementing this model but finally we have achieved our milestone. Earlier we used the LSTM Stacked Model also but there also we faced the error in prediction model as well so that we dropped that model and decided to use the ARIMA Model.

I am attaching some screenshots of the output that we get from our implementation.




Scatter plot of closing price
24
22
20
18
16
14
12
10
8
0
50
100
150
200
250


Train data
Test data
Closing Prices
Dates
3.2
3.0
2.8
2.6
2.4
2.2
0
50
100
150
200
250


TATA MOTORS Stock Price Prediction
training
Actual Stock Price
Predicted Stock Price
Actual Stock Price
Time
3.4
3.2
3.0
2.8
2.6
2.4
2.2
0
50
100
150
200
250

Conclusion

In this paper, we study the use of the ARIMA model to predict the financial movement direction and we saw that this model gave us better results. The data has been collected from Yahoo Finance. The historical data of the past 1 year is taken into account for analysis. The ARIMA model's methodology is used to train and predict the stock prices on the test dataset. We can say that it is a promising tool for financial forecasting and superior to the other individual classification methods in forecasting the movement directions. However, each method has its own strengths and weaknesses. We also saw that the choice of the indicator can dramatically improve or reduce the accuracy of the prediction system. Also a particular Machine Learning Algorithm might be better suited to a particular type of stock, say first_stock whereas the same algorithm might give lower accuracies while predicting some other types of stocks, say second_stock.

Future Scope

The Arima model is used in this study and only one dataset from Yahoo Finance was applied to train and test the models. The system can only predict the direction(up/down) for the next particular days. In the future, this model will be used in order to predict the price moment for the future and the results will be compared with other data mining techniques by applying different dataset from different stock indexes.

References

Youtube

Blog Posts : analyticsvidhya.com

Blog Posts : devpost.com

Blog Posts : datacamp.com

Kaggle

Other Data Science materials.

THANK YOU !!