RAYAT SHIKSHAN SANSTHA’S

# YASHAVANTRAO CHAVAN INSTITUTE OF SCIENCE, SATARA (AUTONOMOUS)

**DEPARTMENT OF STATISTICS**

A Project Report On

**ANALYSIS OF SHARE PRICES OF SELECTED IT COMPANIES**



Submitted by,

## Mane Arati Balasaheb

M.Sc. II

Roll No.202 UID No. 202097102

|  |  |  |
| --- | --- | --- |
| **Dr. D. S. Jadhav** | **Dr. D. S. Jadhav** | **Dr. H. P. Umap** |
| Teacher-In-charge | P.G. Coordinator | Head  Department of Statistics |

|  |  |
| --- | --- |
| **Roll No.: 202** | **UID No.: 202097102** |

RAYAT SHIKSHAN SANSTHA’S

**YASHAVANTRAO CHAVAN INSTITUTE OF SCIENCE, SATARA (AUTONOMOUS)**

## DEPARTMENT OF STATISTICS

(M. Sc. II)

**CERTIFICATE**

This is to certify that the Project study report entitled “**Analysis of Share Prices of Selected IT Companies”** being submitted by **Mane Arati Balasaheb** as a partial fulfilment for the award of degree of M.Sc. in Statistics of Yashavantrao Chavan Institute of Science, Satara (Autonomous) is a record of bonafide work carried out by her under my supervision and guidance.

To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

Date:

Place: Satara

|  |  |  |
| --- | --- | --- |
| **Dr. D. S. Jadhav** |  | **Dr. H. P. Umap** |
| P.G. Coordinator | **Examiner** | Head  Department of Statistics |

|  |  |
| --- | --- |
|  |  |
|  |  |

# ACKNOWLEDGEMENT

I am privileged to express my sincere thanks with great respect and gratitude to Dr. H. P. Umap (Head, Department of Statistics), Dr. D. S. Jadhav, Mr. P. H. Mahadik, Miss. A. D. Chavan, Dr. S.M. Nimbale Miss. K. R. Kadam for their aspiring guidance. They all helped with kind of co-operation and constant encouragement. I am grateful to thank them for providing me with all necessary facilities.

Also, I would like to thank all the non-teaching staff of the department for their help and co- operation. I thank all my friends and the teaching staff for their co-operation and help which I received from them during the work throughout.

I am indebted to my parents for their encouragement and patience throughout my study and for the trust they have on me.

Your sincerely,

Miss. Arati Balasaheb Mane

M. Sc. II Department of Statistics

# INDEX

* Introduction …………… 5
* Objective …………… 6
* Methodology and Data Description …………… 7
* Statistical Tools & Software’s …………… 8
* Data Analysis …………… 9-20
* Major Findings …………… 21
* Limitations & Scope ……………. 22
* References ……………. 22

# INTRODUCTION

Forecasting stock market prices have always been challenging task for many business analysts. In fact, stock market price prediction is an interesting area of research for investors. For successful investment, many investors are interested in knowing about the future situation of the market. Effective prediction systems indirectly help traders by providing supportive information such as the future market direction. Data mining techniques are effective for forecasting future by applying various algorithms to data. Modeling and Forecasting of the financial market have been an attractive topic to scholars and researchers from various academic fields. In the present scenario of the financial market world, especially in the stock market, forecasting the trend or the price of stocks using machine learning techniques and artificial neural networks are the most attractive issue to be investigated.

An index that helps to measure the prices in stock market in order to help the investor to know and compare their current prices with their past prices in order to analyze their market performance is known as Stock Index or Stock Market Index. Sensex and Nifty are the two top most stock market indices in India. Sensex contains top 30 stocks of public listed companies and Nifty contains 50 stocks of public listed companies.

The Indian IT sector has been playing a huge and eminent role in developing the economic growth of the country. The importance of IT sector is not only for the Nation but also for the whole world. Today, every other sector in the country is depending on IT sectors because it not only plays a major role at work place but also in our everyday life say for example, we can take computers which is an appliance based on IT. The IT’s footprints can be seen everywhere. While comparing with other developed countries, skilled professionals in India are considered as highest cost effective.

The stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock. Previous methods of stock predictions involve the use of Artificial Neural Networks and Convolution Neural Networks which has an error loss at an average of 20%. In this report, we will see if there is a possibility of devising a model using Recurrent Neural Network which will predict stock price with a less percentage of error.

In this project, we study the problem of stock market forecasting using Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM). The purpose of this project is to examine the feasibility and performance of LSTM in stock market forecasting.

# OBJECTIVES

### Primary Objective:

* To predict the share prices of selected IT companies.

### Secondary Objectives:

* To understand the stock market behavior by studying the price movement of the selected IT companies share prices.
* Comparative study of share prices of IT companies.
* To know the selected companies stock price pattern in particular trend or trend reversal.

# PROBLEM STATEMENT

The stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock.

In this report, we will see if there is a possibility of devising a model using Recurrent Neural Network which will predict stock price with a less percentage of error.



# METHODOLOGY

This study is aimed at analyzing the share price movement in three National Stock Exchange (NSE) listed IT companies namely Tata Consultancy Services, Infosys Limited and WIPRO Limited. The three companies were selected based upon the market capitalization of the companies, minimum age of thirty years of the company and its reputation in industry.

### Sources of Data

The main sources of data for the present study used are secondary in nature. The data is the price history and trading volumes at a day-level with pricing and trading values split across .csv files. The data spans from 1st March, 2004 to 31st July, 2021.

The dataset using from [www.Kaggle.com](http://www.kaggle.com/) Also, we use direct link:

<https://www.kaggle.com/rohanrao/nifty50-stock-market-data?select=INFY.csv>

### Data Description:

**Name and Symbol:** This column tells you the company name (usually abbreviated) and the stock symbol assigned to it. Financial tables list stocks in alphabetical order by symbol, and you need to use them in all stock communications.

**Day High and Low:** The day high and low simply show the highest and lowest prices at which the stock traded throughout the day, from market open to market close. However, the day high and low may not be the open and close prices – those are separate figures.

**Open Price:** The open price is simply the price at which the stock opened trading on any given day.

**Close Price:** The close price is perhaps more significant than the open price for most stocks. The close is the price at which the stock stopped trading during normal trading hours (after-hours trading can impact the stock price as well). If a stock closes above the previous close, it is considered an upward movement for the stock (and will impact things like candlestick charts, which we'll get to later). Vice versa, if a stock's close price is below the previous day’s close, the stock is showing a downward movement.

**Prev. Close:** The prev. close, or previous close, is the price at which the stock closed the previous day (24 hours before).

**Volume:** This column tells you how many shares of that particular stock were traded that day. If only 100 shares are traded in a day, the trading volume is 100.

**Turnover:** Share turnover is a measure of stock liquidity, calculated by dividing the total number of shares traded during some period by the average number of shares outstanding for the same period. The higher the share turnover, the more liquid company shares are.

# STATISTICAL TOOLS & SOFTWARE’S

### Statistical Tools:

* Exploratory Data Analysis: Line charts, Pair Plot.
* Moving Average.
* Time Series: Long Short-Term Memory (LSTM) Model.

### Statistical Software’s:

* Python 
* MS-Excel Description: microsoft excel logo primary resized2

### Software Specifications:

**NumPy:** NumPy is basically a library that is available in python for scientific computing. It contains a powerful dimensional array object then 10 tools for integrating with C, C++. It is also very useful in linear algebra Fourier transform and random number capabilities. It’s adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The multi-dimensional array has various elements that are stored in their respective memory locations.

**Matplotlib:** Matplotlib is a very useful library for the Python programming language and its statistical discipline extension NumPy. It gives a critique settled API to implanting plots into submissions developing totally beneficial.

**Pandas:** It is likewise a Python library for information control and examination. Specifically, it offers information structures and activities for controlling numerical table and time arrangement information.

**Pandas-Data Reader:** Use to remove information from large number of web sources.

**Scikit-Learn/SkLearn:** It is Machine learning python library which chiefly includes arrangement, relapse, and bunching calculations like help vector machines or SVM, inclination boosting, irregular woods, and k-Nearest Neighbors.

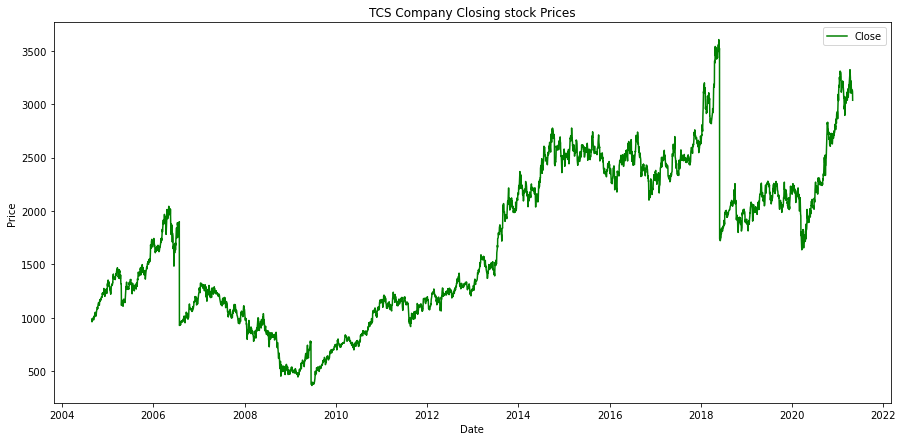
**TensorFlow:** It is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. TensorFlow is a symbolic math library based on dataflow and differentiable programming.

**Keras:** Keras is an open-source software library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

***EXPLORATORY DATA ANALYSIS***

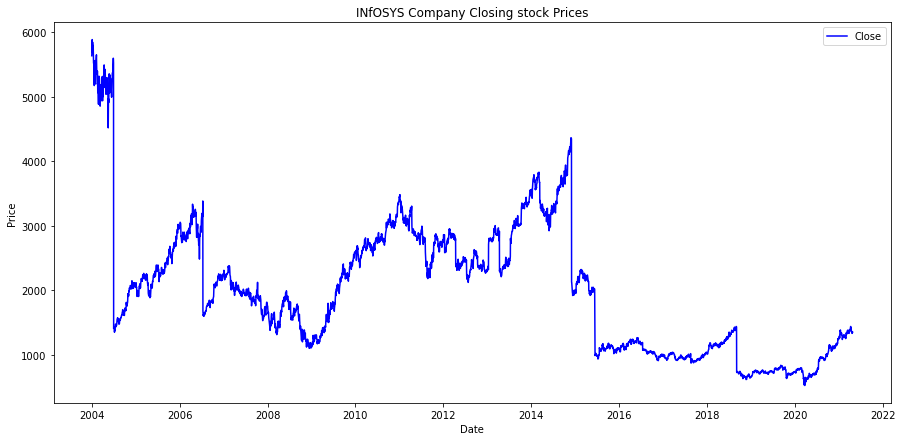
**Closing Prices of IT Companies:**

### Tata Consultancy services

****

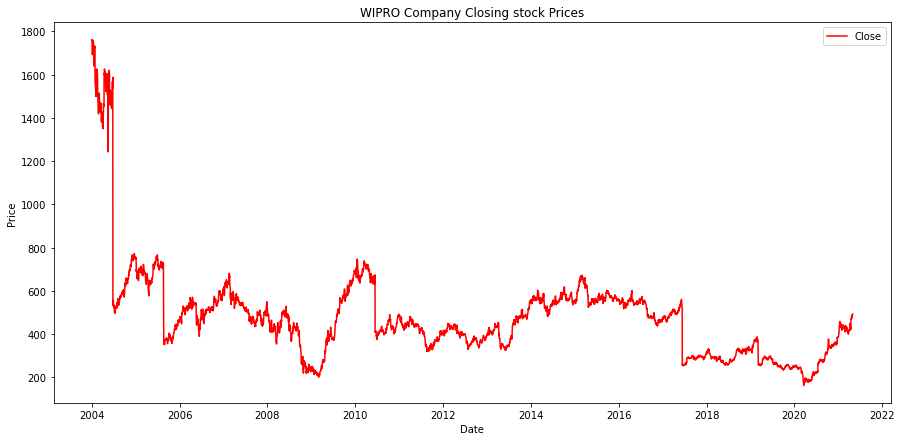
**Interpretation:** A price chart shows how a stock’s price has evolved over a given period of time, which plots the closing price of the Tata Consultancy Services over the beginning of 2004. The stock price of Tata Consultancy Services increases from 2004 constantly with small ups & down but in 2009 share price suddenly going ups nearly 40% because Money life report alleging fresh whistle-blower compliant made to market regulator, Securities Exchange Board of India (SEBI). From this COVID-19 pandemic stock price constantly increases.

### Infosys Limited

****

**Interpretation:** In above graph, the stock price of Infosys Limited suddenly decreases in 2004 but from 2004 to 2020 having many ups and downs started. After that the share price constantly increase highly.

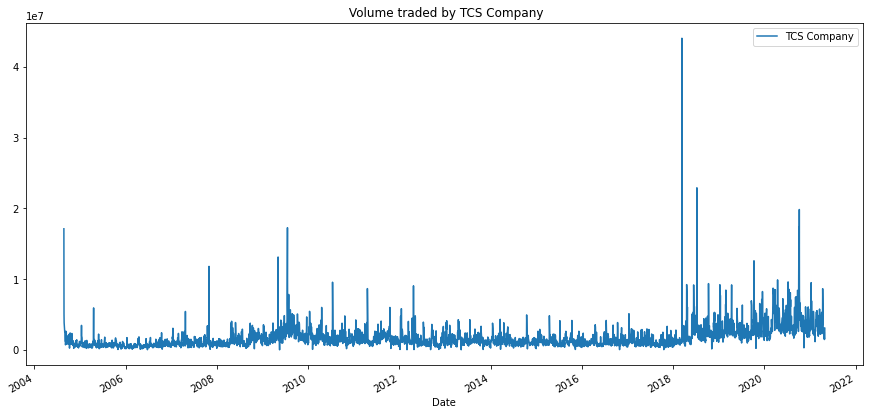
### Western India Palm Refined Oils Limited (Wipro)

****

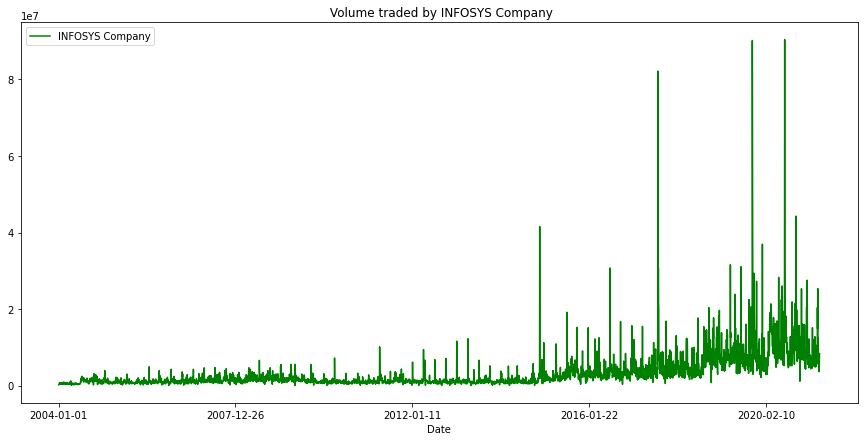
**Interpretation:** From above graph, in 2004 the stock price of WIPRO Limited suddenly falls down. After that stock prices are increase and decrease irregularly up to 2020.

## Graphical Presentation of Volume of IT Companies:

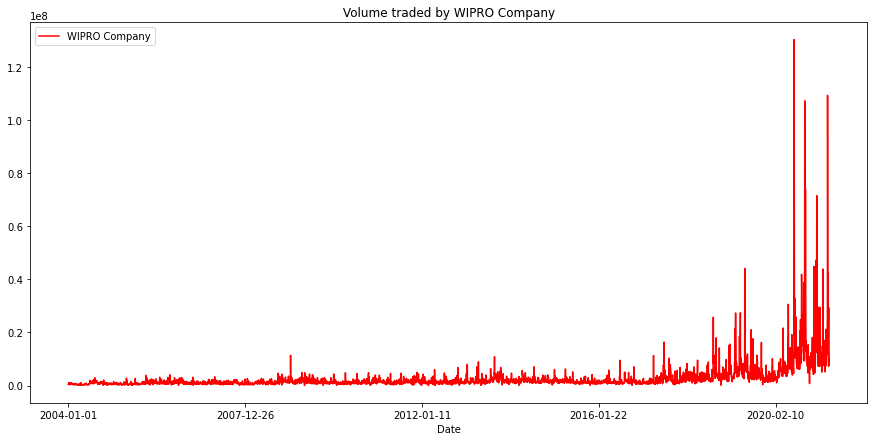
### Tata Consultancy Services

****

* **Infosys Limited**

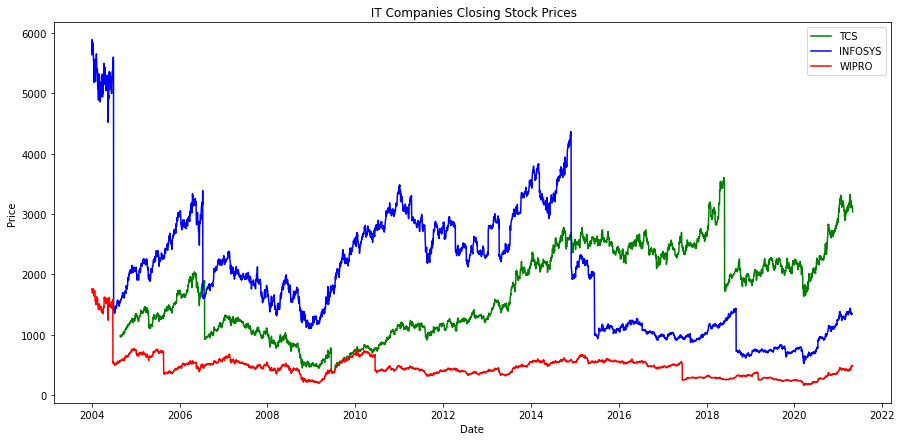
****

### Western India Palm Refined Oils Limited (Wipro)

****

**Interpretation:** A chart shows how a stock’s volume has evolved over a given period of time, which plots the traded volume of the Tata Consultancy Services over the beginning of 2004. All these three IT Industries volume increase constantly. The WIPRO Limited volume is higher as compare to Tata Consultancy Services & Infosys Limited.

## Graphical Presentation of Closing Prices of IT Companies:

****

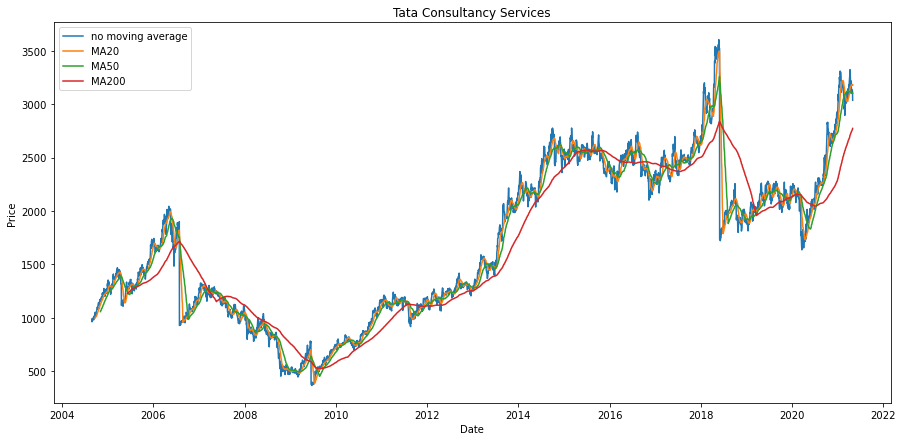
**Interpretation**: A price chart shows how a stock’s price has evolved over a given period of time, which plots the closing price of the Tata Consultancy Services over the beginning of 2004. The above graph shows stock price of all three IT Industries. The stock price of this companies near up to 2009. After that Tata Consultancy Services stock price highly increased as compare to Infosys Limited& WIPRO Limited.

## Moving Average:

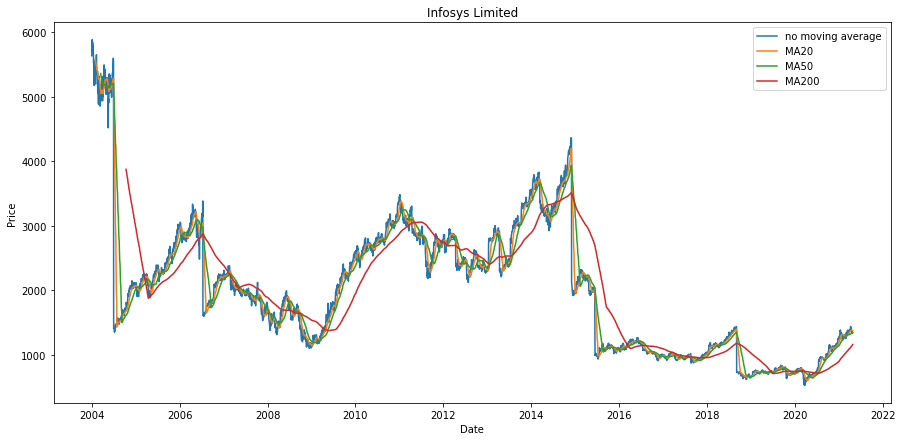
In a time, series, a moving average of period *N* at a certain time *t,* is the mean value of the *N* values before *t* (included). It’s defined for each time instant excluding the first *N* ones. In this particular case, we are talking about the Simple Moving Average (SMA) because every point of the average has the same weight. There are types of moving averages that weigh every point in a different way, giving more weight to the most recent data. It’s the case of the Exponential Moving Average (EMA) or the Linear Weighted Moving Average (LWMA).

Moving averages are often used to detect a trend. It’s very common to assume that if the stock price is above its moving average, it will likely continue rising in an uptrend.

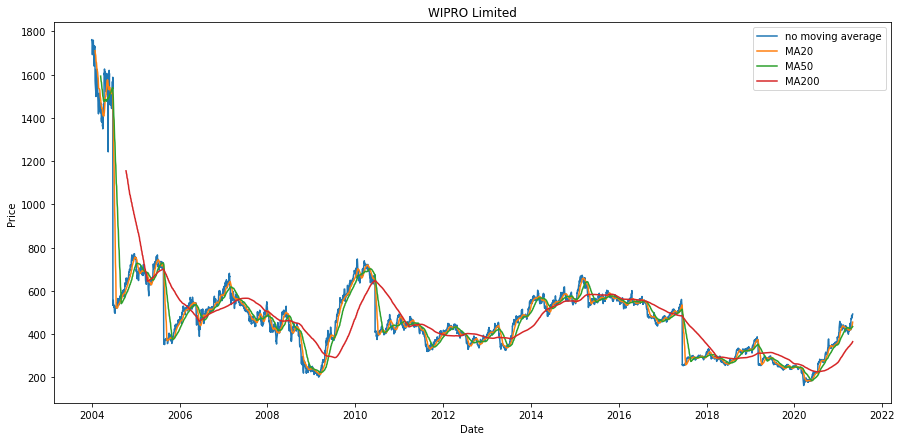
### Tata Consultancy Services

****

* **Infosys Limited**

****

### Western India Palm Refined Oils Limited (Wipro)

****

**Interpretation:** The short moving averages are useful to catch short-term movements, while the 200- period moving average is able to detect a long-term trend.

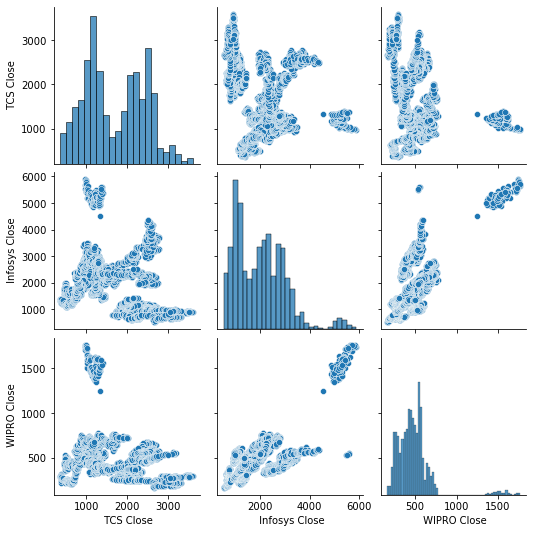
Generally, the most used moving average periods in trading are:

* + 20 for swing trading
  + 50 for medium-term trading
  + 200 for long-term trading

If a stock price is above 200-days moving average, the trend is bullish (i.e., the price rises). So, the investors are often looking for stocks whose price is above the 200-periods moving average.

## Visualizing Data with Pairs Plot:

**Pair Plot:** The pairs plot builds on two basic figures, the histogram and the scatter plot. The histogram on the diagonal allows us to see the distribution of a single variable while the scatter plots on the upper and lower triangle show the relationship between two variables.



**Interpretation:** The default pairs plot by itself often gives us valuable insights. We see that TCS Consultancy Services, Infosys Limited and WIPRO Limited are positively correlated. From the histograms, we see that WIPRO Limited are heavily right-skewed.

## Stock Price Prediction Using LSTM:

### Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) capable of addressing linear problems. LSTM is a deep learning technique. Long-term Memory Units are enforced to learn very long sequences. This is a more general version of the gated recurrent system. LSTM is more benign than other deep learning methods like RNN or traditional feed forward.

This model considers the historical share price of a company price and applies Recurrent Neural Networks (RNN) technique called Long Short-Term Memory (LSTM). The proposed approach uses the time series analysis in order to predict a share price.

### Proposed System

After getting the data then there is a need to extract the feature which is required for data analysis, then divide it as testing and training data, training the algorithm to predict the price and the final step it to visualize the data. The following figure represents the Architecture of the proposed system.



RAW DATA

FEATURE EXTRACTION

TRAINING DATA

TESTING DATA

TRAINING DATA RESULT

PREDICTED STOCK PRICE

The typical LSTM units consists of a cell, an info door, an entrance door and a door with view. The cell collects values over discretionary time intervals, and the three inputs manage the progress of data into and out of the cell. The main advantage of the LSTM is its ability to learn context-specific temporal dependence. Each LSTM unit collects information for either a long or short period of time without explicitly using the activation function within the recurrent components.

### Analysis

The proposed framework that learns online anticipating the close costs of the stock with the assistance of Long Short-Term Memory (LSTM). The Long Short-Term Memory (LSTM) is a counterfeit intermittent neural system (RNN) design used in the field of deep learning, unlike standard feed forward neural systems, LSTM has input associations.

|  |
| --- |
| **Algorithm:** Stock prediction using LSTM |
| **Input:** Historic stock data  **Output:** Prediction of stock price using price variation |
| Step 1: Start.  Step 2: Data Preprocessing after getting the historic data for a particular share. Step 3: Import the dataset to the data structure and read the close price.  Step 4: Do a feature scaling on the data so that the data values will vary from 0 and 1. Step 5: Creating a data structure with 60 timestamps and 1 output.  Step 6: Building the RNN (Recurrent neural network) for Step 5 data set and Initialize the RNN by using sequential repressor.  Step 7: Adding the first LSTM layer and some Dropout regularization for removing unwanted values.  Step 8: Adding the output layer.  Step 9: Compiling the RNN by adding Adam optimization and the loss as mean\_squared\_error.  Step 10: Making the predictions and visualizing the results using plotting techniques. |

### IT Industries Stock Price Prediction

### Train & Test Data and Prediction

The implementation of proposed LSTM model using python which predict the future price of IT companies share based on its historical data. The below visualization figures show the visualization of IT Industries prediction. The implementation of an algorithm which predicts the stock price of a share for given period of time, the below graphs from our algorithm will show the predicted price of IT Industries share.

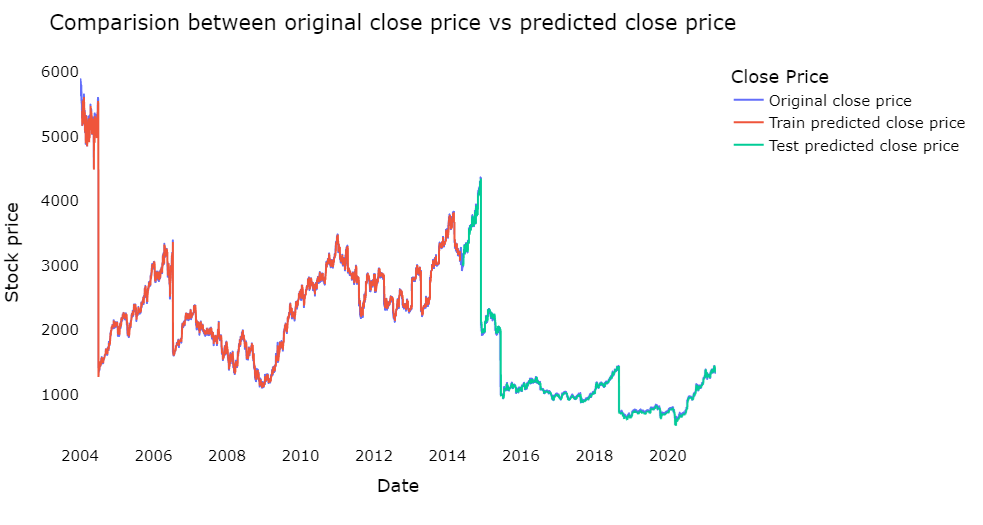
### Tata Consultancy Services

### 

|  |  |  |
| --- | --- | --- |
|  | MSE | RMSE |
| Train Data | 35.28 | 1244.72 |
| Test Data | 51.87 | 1828.18 |

**Interpretation:** The 1st graph has been plotted for whole data set along with some part of trained and test data. The graph is showing the close price of Tata Consultancy Services share for 4139th day’s closing price. The 4139th day close price was 3035 rupees INR and my predicted price is 3121 rupees per share.

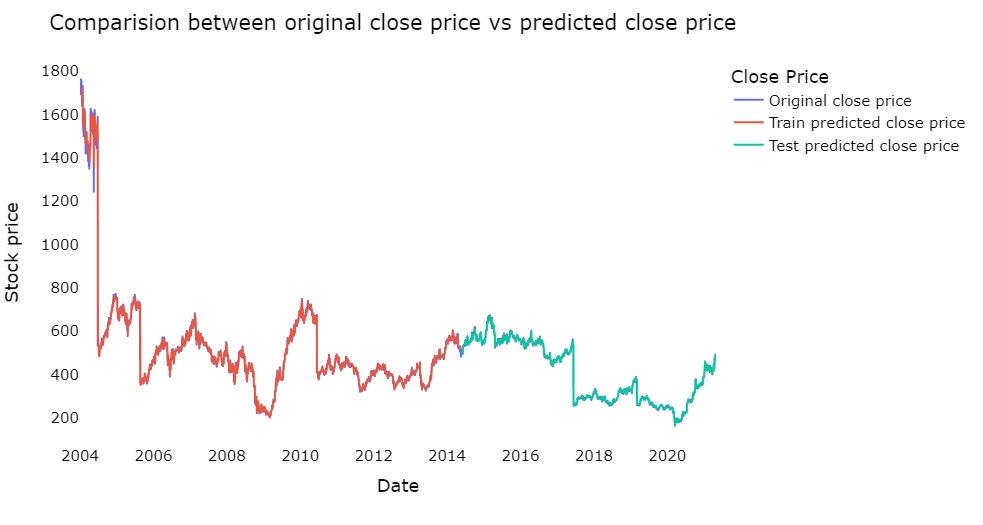
* **Infosys Limited**

****

|  |  |  |
| --- | --- | --- |
|  | MSE | RMSE |
| Train Data | 666076.13 | 1430.4111 |
| Test Data | 672226.50 | 919.8942 |

**Interpretation:** The 1st graph has been plotted for whole data set along with some part of trained and test data. The graph is showing the close price of Infosys Limited share for 4303th days. The 4303th days close price was 1354 rupees INR and my predicted price is 2470 rupees per share.

### Western India Palm Refined Oils Limited (Wipro)

****

|  |  |  |
| --- | --- | --- |
|  | MSE | RMSE |
| Train Data | 719.49 | 26.82 |
| Test Data | 834.89 | 19.34 |

**Interpretation:** The 1st graph has been plotted for whole data set along with some part of trained and test data. The graph is showing the close price of WIPRO Limited share for 4303th days. The 4303th days close price was 492 rupees INR and my predicted price is 498 rupees per share.

**MAJOR FINDINGS**

* + A stock’s price has evolved over a given period of time, which plots the closing price of the IT Industries over the beginning of 2004. In some years the share price falls suddenly but it’s because of split the face value shares. From this COVID-19 Pandemic every IT Industries share price increase constantly.
  + All IT Industries volume increase constantly. The WIPRO Limited volume is higher as compare to Infosys Limited & Tata Consultancy Services.
  + The stock price of this companies near up to 2010. After that Tata Consultancy Services price highly increased as compare to Infosys Limited and WIPRO Limited.
  + A simple algorithm to find the best Simple Moving Average for stock trading. It can be easily applied every trading day in order to find, day by day, the best moving average. In this way, a trader can easily adapt to market changes and to volatility fluctuations.
  + TCS Consultancy Services, Infosys Limited and WIPRO Limited are positively correlated. From the histograms, we see that WIPRO Limited are heavily right-skewed.
  + The implementation of proposed LSTM model which predict the future price of IT Industries share based on its historical data. The proposed model is able to predict the share price with very low loss and error rate with increase the epoch batch rates.

# LIMITATIONS & SCOPE OF THE STUDY

In future enhancement the inclusion of sentiment analysis from social media to understand what the market thinks about the price variation for a particular share and it can be implement this by adding twitter and Facebook API to our program as Facebook is a leading social media which has lots of market trend information posted by users.

# REFERENCES

1. Functional Programming in Python – David Mertz
2. The Art of Data Science – Roger D Peng & Elizabeth Matsui
3. LSTM Recurrent Neural Networks for Signature Verification – Conrad Tiflin
4. Stock Price Prediction Using LSTM – Pramod B. S.1\*, Mallikarjuna Shastry P. M.2