Text

Description automatically generated

**Stock Market Prediction Analysis**

**by**

**Yash Purohit**

**19112020**

**Under the guidance of**

**Dr. Jitendra Kaushik**

**An Industry Project Report submitted in partial**

**fulfillment of the requirements for the award of**

**Degree of Bachelor of Science ( Data Science) of**

**CHRIST ( Deemed to be University),**

**Pune Lavasa Campus**

**April - 2022**

Text

Description automatically generated

CERTIFICATE

*This is to certify that the report titled* ***Stock Market Prediction Analysis*** *is a bona fide record of work done by* ***Yash Purohit (19112020)*** *of CHRIST (Deemed* to be University)*, Pune Lavasa Campus, in partial fulfilment of the requirements of 6 Semester B.Sc. (Data Science) during the year 2022.*

.

**Head of the Department Guide & Co- Guide**

Valued- by:

|  |  |  |
| --- | --- | --- |
|  | Name | : ***Yash Purohit*** |
| 1. | Register Number | : ***19112020*** |
|  | Examination Centre | : ***Pune Lavasa Campus*** |
| 2. | Date of Exam | : 29-Apr-2022 |

### ACKNOWLEDGEMENT

**`**

Words fall short to express my deep sense of gratitude towards them all who have imparted their valuable time, energy and intellect towards the beautification of the analysis project entitled as, **“*Stock Market Prediction Analysis*”.**

It gives me great pleasure in presenting this report. Its justification that it will never sound good if I do not express my vote of thanks to the **Fr. Jossy P. George, Dr. Samiksha Shukla (H.O.D.)** & my guide ***Dr. Jitendra Kaushik*** without whose help this thesis would have neither began nicely nor would have reached a fine ending.

Never can we forget the efforts taken by **All Professors** who’s hard – working nature, sophisticated teaching & guidance helped us framing & building this project & lastly, also thankful to my classmates for being directly or indirectly helpful in building the project to be presented to the college.

**Abstract**

Evolving and opportunistic markets, such as India, are rapidly transforming into machines of future success, but; Only a small portion of Indian household savings are currently invested in the local stock market, but simultaneously the GDP is expanding at 7% to 8% yearly for the last few years, but in the 6% range for 2018 and 2019, and a healthy financial environment, we may see more money enter the race. Perhaps now is the time for international investors to seriously consider investing in India.

A stock market is a marketplace where a company's stocks and derivatives can be traded at a set price. The stock market is driven by supply and demand for shares. The stock market is one of the most rapidly developing sectors in any country. Many people are now involved in this industry, either indirectly or directly.

As a result, understanding market trends becomes critical; again, as its result individuals are interested in stock price predictions as the stock market develops. However, because of the dynamic character of the stock market and its proclivity for rapid price movements, stock price prediction is a difficult assignment. The stock market is primarily a chaotic, non-parametric, non-linear, noisy, and deterministic system. This project tries to undergo all these challenges to reach near prediction of stock prices.

**Content**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No. | | | Content | Page No. |
|  | | | **Acknowledgements** | i |
|  |  | **Abstract** | | ii |
|  |  | **List of Figures** | | iii |
| **1** | | | **Introduction** |  |
|  | 1.1 | Project Introduction | | 1 |
|  | 1.2 | Problem Statement | | 3 |
| **2** | | | **Literature Review** | 4 |
| **3** | | | **Data & Preprocessing** | 6 |
|  | 3.1 | Data Collection | | 7 |
|  | 3.2 | Data Description | | 8 |
|  | 3.3 | Technologies Used | | 10 |
|  | 3.4 | Data Pre-Processing | | 11 |
|  | 3.5 | Data Exploration | | 13 |
| **3.6** | | | **Data Visualisation & Analysis** |  |
|  | **3.6.1** | **108-months Candlestick graph** | | 16 |
|  | **3.6.2** | **15-months Candlestick graph** | | 16 |
|  | **3.6.3** | **5-months Candlestick graph** | | 17 |
|  | 3.6.4 | **30-days Candlestick graph** | | 17 |
|  | 4 | Prediction | | 18 |
|  | 4.1 | ALgorithms | | 19 |
|  | 4.2 | Psudocode and Prediction | | 22 |
| **5** | | | **Conclusion And Future Work** | 28 |
|  | 5.1 | Conclusion and Expression | | 29 |
|  | 5.2 | Future Enhancement | | 30 |
| **6** | | | **Bibliography** | 31 |

### LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **P. No** | **Figure No.** | **Chart Name** |
| 15 | 1 | HDFC BANK PRICE ACTION |
| 16 | 2 | **108-months Candlestick graph** |
| 16 | 3 | **15-months Candlestick graph** |
| 17 | 4 | **5-months Candlestick graph** |
| 17 | 5 | **30-days Candlestick graph** |
| 26 | 6 | closingVSprediction |
| 26 | 7 | Gap\_chart |
| 27 | 8 | PRED\_CHART |
| 28 | 9 | Nifty Bank Chart |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

# INTRODUCTION

### INTRODUCTION:

The efficient market hypothesis asserts that stock prices cannot be predicted and that stock prices follow a random walk. It appears to be quite difficult to substitute an experienced trader's professionalism in anticipating stock prices.

Stock price prediction is very important as it is used by most of the business people as well as common people. People will either gain money or lose their entire life savings in stock market activity. It is a chaos system. Building accurate model is difficult as variation in price depends on multiple factors such as news, social media data, fundamentals, production of the company, government bonds, historical price and country’s economics.

As technology advances, stock traders are increasingly turning to Intelligent Trading Systems rather than basic analysis to predict stock values, allowing them to make quick investment decisions.

5paisa, Zerodha, Upstox, Angel One, Groww, and others are just a few examples. The greatest consumers of trading/investing in the share equity share market are such companies. Their faith is founded on their convenience of use, less time-consuming method, and, most importantly, accurate real-time graphs and charts. We know these firms had an algorithm that could forecast future stock prices on the basis of historical data. The goal of this project is to use Artificial Neural Network Long Short-Term Memory (ANN-LSTM) to predict stock prices, for personal use.

### 1.2 PROBLEM STATEMENT:

As we have seen, the pandemic and the Ukraine - Russia war; the impact of these situations on the stock market was brutal, and though, even listening to the news only partially benefitted the investors, as the market hit severe low. But as the trading techniques have evolved over the year market recovered quickly. The only problem was the unpredictability of the market and the increased uncertainty with absurd volumes, and not the emergencies itself.

At this time, not only the sentiment of the market is shifted but the investors sentiments are also disturbed.

It is a say in the market that when not in balanced sentiment, one should rest and not speculate. The idea of working on a predicting algorithm is induced by these situations, where the person cannot rely on its expertise, he/she takes help of self-tuned program which can be trusted.

One should not solely rely on the algorithm but the program will help setting up a resistance or a support or a stop loss.

Because, as the abstract points out, there is a dearth of a function that many trading platforms should incorporate, namely the forecast of a company's share price, this project was chosen to be undertaken in this area. It has the appearance of something that one should not put their trust in at all. However, as we progress in the field of artificial intelligence, collecting data points even from people's thoughts, it is becoming increasingly possible to anticipate the stock market using only logical and mathematical methods.

For those who are unfamiliar with the concept, it involves developing a model that forecasts the value of a company's stock for the following day by regressing historical data.

.

**CHAPTER 2**

# LITERATURE REVIEW

**2.1 LITERATURE REVIEW**

* The paper and subsequent study analyses the various fluctuating velocities, trends, effects, and other variables related to the global pandemic.
* To arrive at the said conclusions, various metrices and datasets were analysed that represented the vaccine reports, mortality rate, state-wise demographics.
* Lin et al. (2007) made stock market forecasts using the genetic algorithm. The choice of degrees for all factors and their combinations is a critical aspect in the effectiveness of a trading rule. The difficulty is that the range of parameters fluctuates across a wide area, and finding the best parameter combinations is difficult. This article solves the problem using a genetic algorithm.
* Share market trends are one of the most difficult questions to address in financial arguments, owing to the large number of factors that influence the functioning of the stock market and the complexity of the problem. Political events, the broader economic scenario, and even the traders expectatons are just a few examples of what is considered.
* Since artificial neural networks are one of the prediction methods available in this subject area that are more effective in comparison to others a number of studies have found that artificial neural networks are more precise in prediction than regression models, and they also provide distinct analysis.

**CHAPTER 3**

**DATA AND PREPROCESSING**

### Data Sources

* For the Stock market data, I scrapped data from Yahoo finance and NSE

1. <https://finance.yahoo.com/>
2. <https://www1.nseindia.com/live_market/dynaContent/live_watch/equities_stock_watch.htm>

* for observing and learning, I used Money control, and trading view

1. <https://www.moneycontrol.com/indian-indices/nifty-bank-23.html>
2. <https://in.tradingview.com/chart/9Af4vLau/>

### Data Description

### The data scrapped was dynamic time series data availed by NSE India, but for better python support yahoo finance was scrapped.

### 

### Info:

### 

### 

### 

### 

### *The Data Sets Columns:*

### ‘zip', 'sector', 'fulltimeEmployees', 'longBusinessSummary', 'city', 'phone', 'country',

### 'companyOfficers', 'website', 'maxAge', 'address1', 'fax', 'industry', 'address2', 'ebitdaMargins',

### 'profitMargins', 'grossMargins', 'operatingCashflow', 'revenueGrowth', 'operatingMargins', 'ebitda',

### 'targetLowPrice', 'recommendationKey', 'grossProfits', 'freeCashflow', 'targetMedianPrice', 'currentPrice', 'earningsGrowth', 'currentRatio', 'returnOnAssets', 'numberOfAnalystOpinions',

### 'targetMeanPrice', 'debtToEquity', 'returnOnEquity', 'targetHighPrice', 'totalCash', 'totalDebt',

### 'totalRevenue', 'totalCashPerShare', 'financialCurrency', 'revenuePerShare', 'quickRatio',

### 'recommendationMean', 'exchange', 'shortName', 'longName', 'exchangeTimezoneName',

### 'exchangeTimezoneShortName', 'isEsgPopulated', 'gmtOffSetMilliseconds', 'quoteType', 'symbol',

### 'messageBoardId', 'market', 'annualHoldingsTurnover', 'enterpriseToRevenue', 'beta3Year',

### 'enterpriseToEbitda', '52WeekChange', 'morningStarRiskRating', 'forwardEps',

### 'revenueQuarterlyGrowth', 'sharesOutstanding', 'fundInceptionDate', 'annualReportExpenseRatio',

### 'totalAssets', 'bookValue', 'sharesShort', 'sharesPercentSharesOut', 'fundFamily', 'lastFiscalYearEnd',

### 'heldPercentInstitutions', 'netIncomeToCommon', 'trailingEps', 'lastDividendValue',

### 'SandP52WeekChange', 'priceToBook', 'heldPercentInsiders', 'nextFiscalYearEnd', 'yield',

### 'mostRecentQuarter', 'shortRatio', 'sharesShortPreviousMonthDate', 'floatShares', 'beta',

### 'enterpriseValue', 'priceHint', 'threeYearAverageReturn', 'lastSplitDate', 'lastSplitFactor', 'legalType',

### 'lastDividendDate', 'morningStarOverallRating', 'earningsQuarterlyGrowth',

### 'priceToSalesTrailing12Months', 'dateShortInterest', 'pegRatio', 'ytdReturn', 'forwardPE',

### 'lastCapGain', 'shortPercentOfFloat', 'sharesShortPriorMonth', 'impliedSharesOutstanding', 'category',

### 'fiveYearAverageReturn', 'previousClose', 'regularMarketOpen', 'twoHundredDayAverage',

### 'trailingAnnualDividendYield', 'payoutRatio', 'volume24Hr', 'regularMarketDayHigh', 'navPrice',

### 'averageDailyVolume10Day', 'regularMarketPreviousClose', 'fiftyDayAverage',

### 'trailingAnnualDividendRate', 'open', 'toCurrency', 'averageVolume10days', 'expireDate', 'algorithm',

### 'dividendRate', 'exDividendDate', 'circulatingSupply', 'startDate', 'regularMarketDayLow', 'currency',

### 'trailingPE', 'regularMarketVolume', 'lastMarket', 'maxSupply', 'openInterest', 'marketCap',

### 'volumeAllCurrencies', 'strikePrice', 'averageVolume', 'dayLow', 'ask', 'askSize', 'volume',

### 'fiftyTwoWeekHigh', 'fromCurrency', 'fiveYearAvgDividendYield', 'fiftyTwoWeekLow', 'bid',

### 'tradeable', 'dividendYield', 'bidSize', 'dayHigh', 'regularMarketPrice', 'preMarketPrice', 'logo\_url'

### TECHNOLOGIES USED

Python for data scrapping, preprocessing, visualization etc.

Libraries Used :

* + 1. Pandas: Working with data files
    2. Numpy : for Scientific Calculation
    3. Matplotlib : Basic Visualization
    4. Plotly: Advance Visualization
    5. Math: splitting the data, ceiling
    6. Yfinance: ticker and data set
    7. Sklearn: for scaling
    8. Keras for saving models
    9. Seaborn: data visualization

import yfinance as yf

import math

import numpy as np

import pandas as pd

from sklearn.preprocessing import MinMaxScaler

from keras.models import Sequential  # layer by layer

from keras.layers import Dense, LSTM # Dense - helps in carrying neurons from one layer to another  # LSTM - Recurrent Neural Networks and it has loops to think from scratch

import matplotlib.pyplot as plt

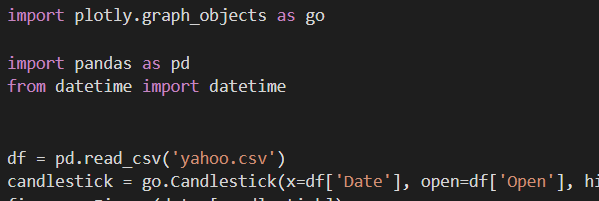
import seaborn as

### DATA PRE-PROCESSING

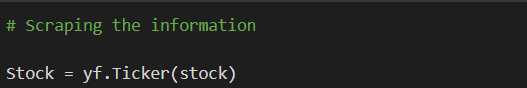
An important data mining approach is data preprocessing, which is the transformation of raw data into a format that can be understood. In real-world situations, data is frequently insufficient, inconsistent, and/or deficient in specific behaviors or trends, and it is likely to contain a large number of mistakes. Such concerns can be resolved by the use of data preparation, which has been demonstrated to be effective. A method of preparing raw data for later processing is known as data preprocessing.

Steps Used:

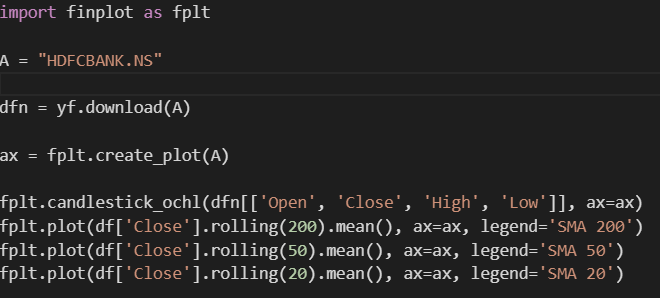
The dynamic datasets were converted in the form of csv files in order to be extracted later.



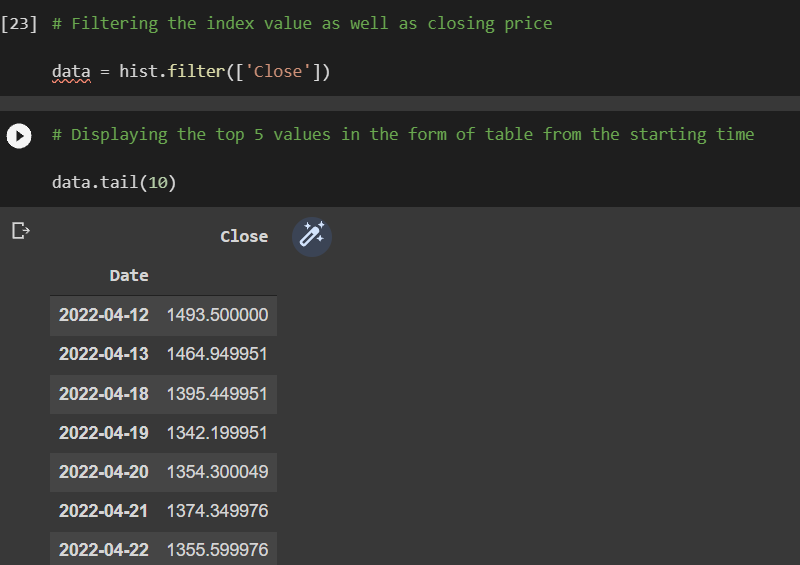
1. Using Yfinance.ticker to import the dataset onto python for data analysis.



1. Using Yfinance.download to import the dataset onto python for data analysis.



1. Filtering out columns based on necessity.

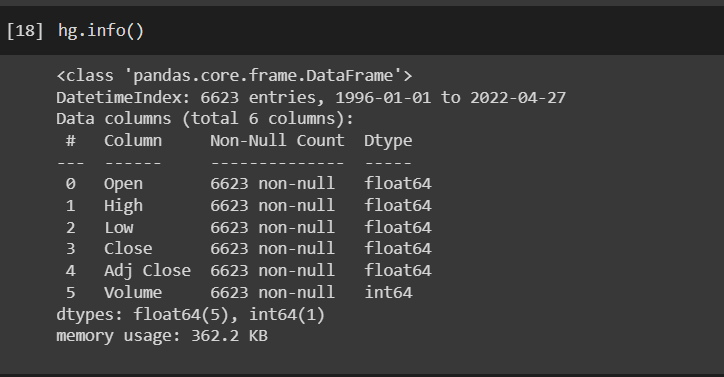


### DATA EXPLORATION

#### EDA

#### 

#### 



Disc: the population size is 6623

The market parameters are open , high , low, close , adjusted close, volume.

#### 3.6 VISUALISATION & ANALYSIS

#### 

plt.figure(figsize=(16,8))

plt.title('Close Price History')

plt.plot(hist['Close'])

plt.xlabel('Date',fontsize=18)

plt.ylabel('Close Price INR ($)', fontsize=18)

plt.legend([stock], loc='upper center', fontsize=10)

plt.show()

This graph here shows a trend line graph of HDFC bank ltd. The line in the graph is closing price action of the hdfc share over the years from 2013, till date.

The graph of the stock shows a good uptrend over these 9 years.

And now its seems to consolidate between a 1400 to 1600 INR range. The code takes the closing price of the stock

**GRAPHS:**

**3.6.1 108-months Candlestick graph**



A weak support is observed at 700 mark and 1300 in the chart above.



**3.6.2 15-month candlegraph**

A down trend

****

A very strong support is observed at 1380,

Also, a strong resistance can be seen at 1610.

A reverse head and shoulder & a consolidation right after

**3.6.3 5-month candlegraph**

A down trend

****

It’s a down trend in the 5-month chart

Also, a strong resistance can be seen at 1590.

And a weak support at 1400

**3.6.4 30-day candlegraph**

****

The 30-day graph shows pure consolidation.

Here also the line is resisted at around 1550 and supported around 1400.

All the graphs are representing the same line at different time frames,

But as we observed: 108 m | 15 m | 5m | 30d

Resistance: | 1600 | 1590 | 1550

Support: 1300 | 1380 | 1400 | 1400

**CHAPTER 4**

**PREDICTION**

**4.1 THE ALGORITHMS APPLIED**

The Long Short-Term Memory of an Artificial Neural Network (ANN-LSTM)

ANN is analogous to the human brain. There are millions and billions of cells (referred to as neurons) in the brain that process information in the form of electrical signals to generate memories. In a similar way, an ANN algorithm comprises millions and billions of cells, and all of the cells are connected to one another through the use of nodes. It is sometimes referred to as a multi-layer perceptron since it has three levels (an input layer, a hidden layer (also known as a distillation layer), and an output layer. Using these layers, you may better grasp the data's essential patterns and patterns of interest. Before sending the information to its second level of processing, the first layer assigns different numbers to it (ranging from 0 to 1), and the information is then sent to the second layer, where it is again assigned certain numbers by the algorithm (ranging 0 - 1). The third and final layer likewise assigns certain values to the data and then makes the final prediction/forecast by examining various patterns in the previous layers. Furthermore, the LSTM (Long Short-Term Memory) algorithm will be used to estimate the closing price of a firm based on historical data, which will be obtained from the Yahoo Finance website.

**4.2COMPARISION OF OTHER ALGORITHMS**

LSTM allows us to incorporate more and more control knobs while also dividing the data into equal batches (as many batches as we wish - for example, there are 100 data points). It might be difficult and time-consuming to analyze 100 data points at once. We may divide the data points into multiple batches to speed up the process and lower the mistake rate. For example, let's divide the 100 data points into 10 batches to reduce the error rate. Now that we have 10 data points in a single batch, the procedure will be faster and the error rate will be lower). LSTMs also enable us to utilize the epochs option to feed data through neural networks as many times as we like, which is really convenient. It makes use of the dense function to transfer data from one neutron to another, and we can add as many layers as we like to this process by employing the sequential function in order to have a better outcome.

SVM, SGD, Random Forest, and ANN are algorithms that are related to one another since they can deal with the same kind of issues (regression and classification).

Regularization capabilities are provided by the SVM - Support Vector Machine. While this prevents the model from over-fitting, it may also be used to address both category and mathematical problems.

A little adjustment to the data does not have a significant impact on the SVM. As a result, the SVM model is stable.

An algorithm called SGD - Stochastic Gradient Descent - is used to limit the amount of time it takes to reach a specific point in a linear function. As each case is considered, the model is refreshed based on the inclination of the loss in order to mimic an open slope. M-estimator is one example of an indicator returned by regression as a minimizer of the aggregate. Regression is useful for both large and small datasets.

There are many trees in Random Forest, which makes it possible to make precise decisions because it has a large number of trees. It's a forest, as we might expect from the name. As a result, Random Forest requires more time to prepare than a single tree. To predict the outcome, each branch and leaf of the decision tree sifts through the random factors. Once this computation has gathered all predicted individual decision trees, it can also account for any missing attributes in order to come up with the final estimate.

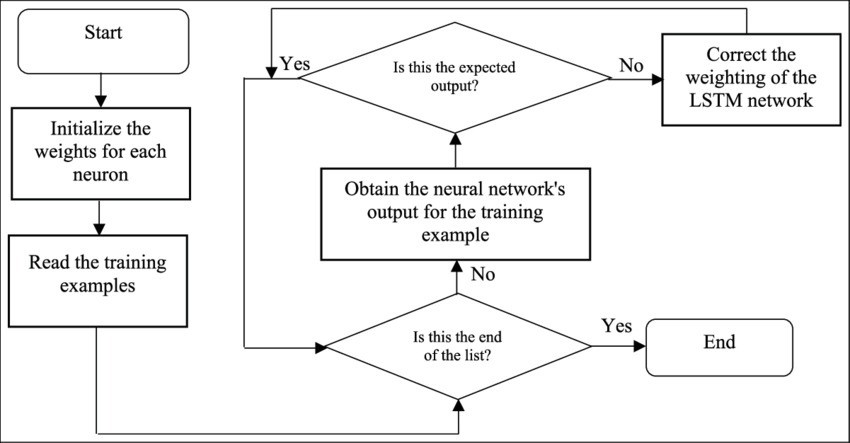
Then, as previously said, the ANN-LSTM operates in a manner similar to that of a human brain, with millions of bits of data being transmitted through numerous layers until the most optimal solution is discovered. Furthermore, LSTM, a form of recurrent ANN, is a fantastic technique for forecasting something as delicate as stock values, among other things.

We may argue that the LSTM-ANN model is superior than the other models since it separates the data and can have a large number of layers. The finest feature is that it starts the process from the beginning and does not use any pre-processing, which allows it to undertake more research than any other algorithm. When it comes to making predictions, SGD is also a good model because it helps in reducing the loss factor and has a good AUC. The second-best thing about LSTM is that it uses different optimizers to compile the results, and the best optimizer is the Adam optimizer because it uses two gradient descent methodologies to interpret the results and it is adaptive in learning (AUC - it is a measure to differentiate between two classes). That is why we chose LSTM-ANN, which performs significantly better than the previously described algorithms.

# 4.3 PSEUDOCODE

1. Following the installation of the essential libraries, a single firm will be chosen for stock price prediction.
2. Then, all stock-related data is retrieved, including the current price, volume, and profit.
3. Then, a certain time period is chosen for obtaining historical data, in this case, 15 months.
4. Then, a few graphs will be created to help visualize the stock's present position.
5. Following that, all columns will be deleted except for data and closing price, and all analysis will be performed on closing prices alone.
6. The model will then be trained using a subset of the dataset.
7. The training set's output will be transformed to an array so that it may be compared to the original dataset using the Adam optimizer.
8. After converting the data using Adam optimizer, the data is checked using rooted mean; if the error rate is between 0 and 0.5, the model is considered to be perfect; if the rate is between 0.5 and 0.9, the model is considered to be good. However, if it is greater than 1, it indicates that significant changes should be made to make it more realistic.
9. After combining predicted and actual closing prices, a graph was created to help visualise the pattern.
10. The second final step is to determine the difference between the actual closing price and the predicted closing price, as well as the returns on both. The final step is to run the model for the last 60 days to determine the stock's predicted price for the following day.

PROJECT FLOWCHART



*ANN-LSTM*

(*Image Source: https://www.researchgate.net/figure/Flowchart-for-LSTM-training\_fig3\_328881984)*

# PRIMARY ALGORITHM’S CODE

LSTM – ANN

# Initiating the Model using Sequential() function (ANN Model type as it also have layers)

# adding layer using LSTM and making conection with another layer using Dense

# 2-layer model - LSTM

model = Sequential()

model.add(LSTM(50, return\_sequences=True, input\_shape = (x\_train.shape[1],1)))

model.add(LSTM(50, return\_sequences = False)) model.add(Dense(25))

model.add(Dense(1))

# using adam optimizer to compile the data (to be in accordance with requirements)

model.compile(optimizer='adam',loss='mean\_squared\_error')

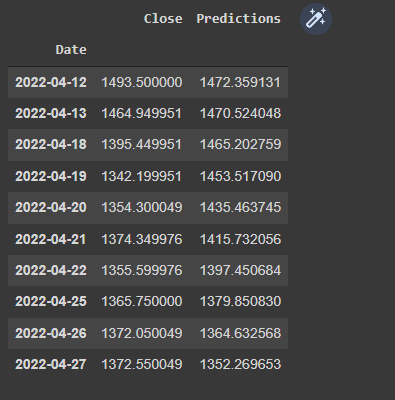
# Fitting the model

# Creating batch as to divide the data equally

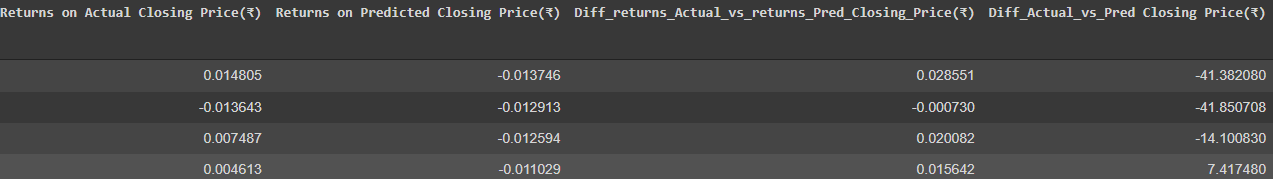
# allowing it to pass through neural networks using epochs model.fit(x\_train, y\_train, batch\_size = 30, epochs = 20)

(This is the main code block of LSTM after this the model is trained using the output of the main algorithm)

****

1. 

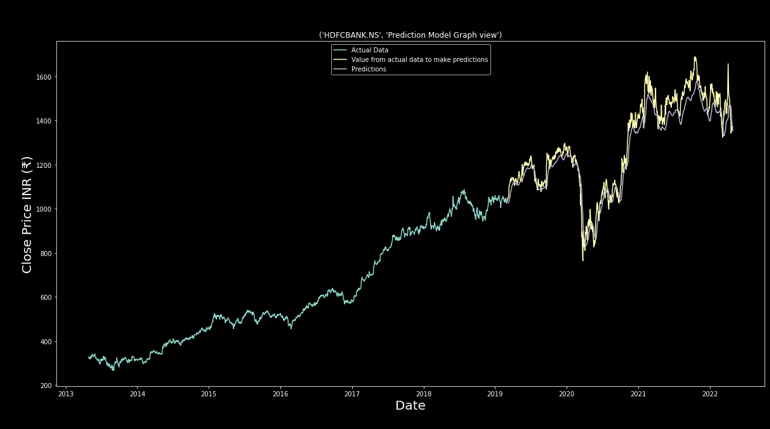
**Close\_pred**

****

**The gap charts**

**As we see the variance is insignificant**

**Calling our prediction, a good fit.**

****

Pred\_chart

In this pred\_chart:

The green line in the chart represent the closing price action line.

The yellow line also represents the part of the the closing price line, but the part which is taken for prediction

And the grey line is mapping the prediction over this yellow line.

**CHAPTER 5**

**CONCLUSION & FUTURE ENHANCEMENT**

### CONCLUSION

### The project was focused on predicting price of shares

### With a successful result we can make out few conlucions

### The share we used was hdbc bank listed in NSE.

### The price action was easily speculatable by human fundamental and technical analysis

### Second:

### The price action;

### I would like to highlight:

### 

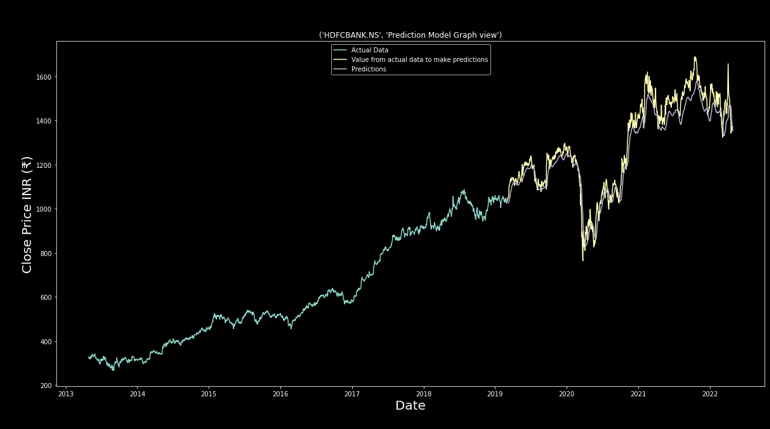
### This is a chart of nifty bank

### Which shows a very similar price action to the

### Price action of hdfc bank

>> we can guess but conclude that hdfc is the most dominant bank in bank nifty

>> it also is a big movable in nifty 50 index.

****

5.2.1 Expression

This project entails an inconceivable number of lines of code, as well as a significant deal of thinking and effort. From the beginning of the process, there has been a single goal: to be able to accomplish something that people genuinely desire, but that is most likely not already available. Investors will be quite enthusiastic about the development of a share price predictor in the near future.

### FUTURE ENHANCEMENT

##### Parameter tuning

##### highly enhanced tuning of parameters is an future enhancement.

##### Feature engineering

##### This project only focused on only one parameter and the historical data on it, but stock market forces comprise many other variables that helps predicts and creating an ideal algorithm, to create or near idea how a machine can speculate market action like a human.

##### Many features need to be added using news market rates and Interest rates, etc.

##### Sentiment Analysis

##### Not only technical analysis but market sentimental analysis plays equally important role in predating or speculating.

##### The major difference between a human and an algo is the sentiment analysis part, it cant get close to good when it comes to modeled sentiment analysis.

##### This is again is a wing of study making it to future enhancement of this project

**CHAPTER 6**

# BIBLIOGRAPHY

References

Candlestick. (n.d.). Plotly: The front end for ML and data science models. <https://plotly.com/python/candlestick-charts/>

LSTM architecture | Understanding the LSTM architecture. (2021, January 21). Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/01/understanding-architecture-of-lstm/>

Mousazadeh Abbassi, N., Ali Aghaei, M., & Moradzadeh Fard, M. (2014). An integrated system based on fuzzy genetic algorithm and neural networks for stock price forecasting. International Journal of Quality & Reliability Management, 31(3), 281-292. <https://doi.org/10.1108/ijqrm-06-2012-0085>

Recurrent neural network | Fundamentals of deep learning. (2020, November 28). Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/>