A PROJECT PHASE-I REPORT

ON

STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING
IN
ELECTRONICS AND TELECOMMUNICATION

BY

Dhananjay Patil	**ROLL No. E41055**
Pranav Nerkar	**ROLL No. E41056**
Harshwardhan Patil	**ROLL No. E41057**

UNDER THE GUIDANCE OF

Prof P.S. KOKARE

Sinhgad Institutes

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION ENGG.
STES'S
SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING
VADGAON BK., OFF.SINHGAD ROAD
PUNE 411041
2022 – 2023

CERTIFICATE

This is to certify that the project phase-I report entitled

"**STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING**"

Submitted by

Dhananjay Patil	Roll.No. E41055
Pranav Nerkar	Roll No. E41056
Harshwardhan Patil	Roll No. E41057

Is a bonafide work carried out by them under the supervision of **Mr. P.S. Kokare** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University for the award of the Degree of Bachelor of Engineering (Electronics and Telecommunication Engineering)

This project phase-I report has not been earlier submitted to any other Institute or University for the award of any degree or diploma.

Mr./Ms.----------------------   Mr./Ms.----------------------
Co-guide, Department of E & TC   **Guide, Department of E & TC**

Mr. P.S. Kokare/Ms. M. M. Sonkhaskar	Dr. S.K. Jagtap	Dr. A.V. Deshpande
Project Coordinator	**Head**	**Principal**
Department of E&TC	**Department of E&TC**	**S.K.N.C.O.E, Pune-41**

Place: Pune
Date :

CONTENTS

CHAPTER NO.		TITLE	PAGE

ACKNOWLEDGEMENT

ABSTRACT

In this project we attempt to implement a 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. 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. Introduction

1.1 Background

The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening a new business or the need for a high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analysed.

Time-series prediction is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit. Many time series prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on Recurrent Neural Networks (RNN), as well as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study it and propose various models. In this project, the LSTM model is used to predict the stock price.

1.2 PROBLEM STATEMENT

In order to predict the stock prices in future markets, we have analyzed papers and have given an overview on how these algorithms give precise and accurate future predictions. In this paper, we used several algorithms from which we observed that not all the algorithms implemented can predict data we need. There has been a basic requirement for computerized and automated ways to deal with powerful and proficient usage of huge measures of money related information to help organizations and people in vital arranging and decision making on investments.

1.3 OBJECTIVE

- To Develop a prototype to estimate the price of a given stock in the near future.
- To use Machine Learning Conception to develop the proposed model.
- To use the LTSM model for the planned system for maximum accuracy.
- To develop an algorithm with maximum efficiency.
- To develop a deep understanding of all the concepts used in the proposed model.
- To resolve errors or problems if encountered any while making the proposed model

CHAPTER 2
LITERATURE SURVEY

2. Literature Survey

2.1 Introduction

The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analysed.

Businesses primarily run over customer's satisfaction, customer reviews about their products. Shifts in sentiment on social media have been shown to correlate with shifts in stock markets. Identifying customer grievances thereby resolving them leads to customer satisfaction as well as trustworthiness of an organization. Hence there is a necessity of an un biased automated system to classify customer reviews regarding any problem. In today's environment where we're justifiably suffering from data overload (although this does not mean better or deeper insights), companies might have mountains of customer feedback collected; but for mere humans, it's still impossible to analyse it manually without any sort of error or bias. Oftentimes, companies with the best intentions find themselves in an insights vacuum. You know you need insights to inform your decision making and you know that you're lacking them, but don't know how best to get them. Sentiment analysis provides some answers into what the most important issues are, from the perspective of customers, at least. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition.

2.2 Backgroud Study

The prediction of stock value is a complex task which needs a robust algorithm background in order to compute the longer-term share prices. Stock prices are correlated within the nature of market; hence it will be difficult to predict the costs. The proposed algorithm using the market data to predict the share price using machine learning techniques like recurrent neural network named as Long Short-Term Memory, 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. The network is trained and evaluated with various sizes of input data to urge the graphical outcomes [1].

In the era of big data, deep learning for predicting stock market prices and trends has become even more popular than before. We collected 2 years of data from Chinese stock market and proposed a comprehensive customization of feature engineering and deep learning-based model for predicting price trend of stock markets. The proposed solution is comprehensive as it includes pre-processing of the stock market dataset, utilization of multiple feature engineering techniques, combined with a customized deep learning-based system for stock market price trend prediction.

We conducted comprehensive evaluations on frequently used machine learning models and conclude that our proposed solution outperforms due to the comprehensive feature engineering that we built. The system achieves overall high accuracy for stock market trend prediction. With the detailed design and evaluation of prediction term lengths, feature engineering, and data pre-processing methods, this work contributes to the stock analysis research community both in the financial and technical domains [2].

A stock market, equity market or share market is the aggregation of buyers and sellers of stocks (also called shares), which represent ownership claims on businesses. The task of predicting stock prices is one of the difficult tasks for many analysts and in fact for investors. For a successful investment, many investors are very keen in predicting the future ups and down of share in the market. Good and effective prediction models help investors andanalysts to predict the future of the stock market. In this project, I had proposed Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) model by using Machine andDeep Learning models to predict stock market prediction. In present, there are several models to predict the stock market but they are less accurate. I had proposed a model that uses RNN and LSTM to predict the trend in stock prices that would be more accurate.

LSTM introduces the memory cell, a unit of computation that replaces traditional artificial neurons in the hidden layer of the network. In this work by increasing the Epochs and batch size, the accuracy of prediction is more. In proposed method, I am using a test data that is used to predict which gives results that are more accurate with the test data. The proposed method is capable of tracing and prediction of stock market and the prediction will produce higher and accurate results[3].

In recent years, machine learning and deep learning have become popular methods for financial data analysis, including financial textual data, numerical data, and graphical data. One of the most popular and complex deep learning in finance topics is future stock prediction. The difficulty that causes the future stock forecast is that there are too many different factors that affect the amplitude and frequency of the rise and fall of stocks at the same time. Some of the company-specific factors that can affect the share price like news releases on earnings and profits, future estimated earnings, the announcement of dividends, introduction of a new product or a product recall, secure a new large contract, employee layoffs, a major change of management, anticipated takeover or merger, and accounting errors or scandals. Furthermore, these factors are only company factors, and other factors affect the future trend of stocks, such as industry performance, investor sentiment, and economic factors. This paper proposes a novel deep learning approach to predict future stock movement. The model employs a blending ensemble learning method to combine two recurrent neural networks, followed by a fully connected neural network. In our research, we use the S&P 500 Index as our test case. Our experiments show that our blending ensemble deep learning model outperforms the best existing prediction model substantially using the same dataset, reducing the mean-squared error from 438.94 to 186.32, a 57.55% reduction, increasing precision rate by 40%, recall by 50%, F1-score by 44.78%, and movement direction accuracy by 33.34%, respectively. The purpose of this work is to explain our design philosophy and show that ensemble deep learning technologies can truly predict future stock price trends more effectively and can better assist investors in making the right investment decision than other traditional methods.

2.3 SURVEYS

[1] Pramod, & P.M. Mallikarjuna :-

• getting the historical data from market is mandatory step. 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.

• Using the proposed system has help us achieve an accuracy of 96 LSTM units.

• It does not take market sentiments into consideration.

[2] Shen, J. , Shafiq, M.O. :-

• The high-level architecture of our proposed solution could be separated into three parts. First is the feature selection part, to guarantee the selected features are highly effective. Second, we look into the data and perform the dimensionality reduction.

• The proposed system has a lot of features and has a decent accuracy. We can use combination of features by using some features and retaining others to furthermore increase our accuracy.

• The RFE algorithm is not sensitive to the term lengths other than 2-day, weekly, biweekly.

[3].Shriram. S, Dr. K. Anuradha, Dr. K. P. Uma

• The Field of the proposed model is ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, DEEP LEARNING and my model is FUTURE STOCK PRICE PREDICTION USING RECURRENT NEURAL NETWORK, LSTM AND MACHINE LEARNING. The Stock Price prediction model can predict more accurate than other existing models and my project is also different in a way that I had created a Graphical User Interface (GUI) where we can upload the Train data, Test Data and we can get the result of the model and the future 30 days predicted graph with the stock prices

• The main Advantage is that since the model uses RNN, LSTM, Machine Learning and Deep Learning models the prediction of stock prices will be more accurate. And, in the model it can predict the future 30 days Stock Prices and it can show it in a graph. Also, the main feature is that the model can show an output of the Individual Predicted Close prices of the Predicted 30 days.

• The algorithm mostly runs good except when it comes to shadow or other interference on the road

[4] Li, Y., Pan

•   we decide to deploy a blending ensemble learning model that combines LSTM and GRU to accomplish this difficult task. The main differences between LSTM and the GRU are the exposure of memory content inside the unit and how new information is processed by each unit. For the LSTM unit, the amount of memory content that is seen controlled by the output gate (not all of the content are exposed to other units; the output gate decides what information will be used in the next LSTM unit).

•   High accuracy and efficiency

•   The existing approaches are failed to detect the market sentiments or any sudden news that will be taken into consideration.

CHAPTER 3

METHODOLOGY

3. Methodology

Stock is unpredicted and liberal in nature. The follow of the same is impressive and reluctant in nature. Finding the predictability and getting the nearest is the best hit goal for the same. The exact and accurate estimation of the same is never-less possible.

There are various constrains that in-fluctuate the pricing and the rate of stock. Those constrains had to be taken in consideration before jumping to the conclusion and report derivation.

Long short-term memory network :

Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

Working of LSTM:

LSTM is a special network structure with three "gate" structures. Three gates are placed in an LSTM unit, called input gate, forgetting gate and output gate. While information enters the LSTM's network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate.

3.1 Block Diagram

A structure chart (SC) in software engineering and organizational theory is a chart which shows the breakdown of a system to its lowest manageable levels. They are used in structured programming to arrange program modules into a tree. Each module is represented by a box, which contains the module's name.
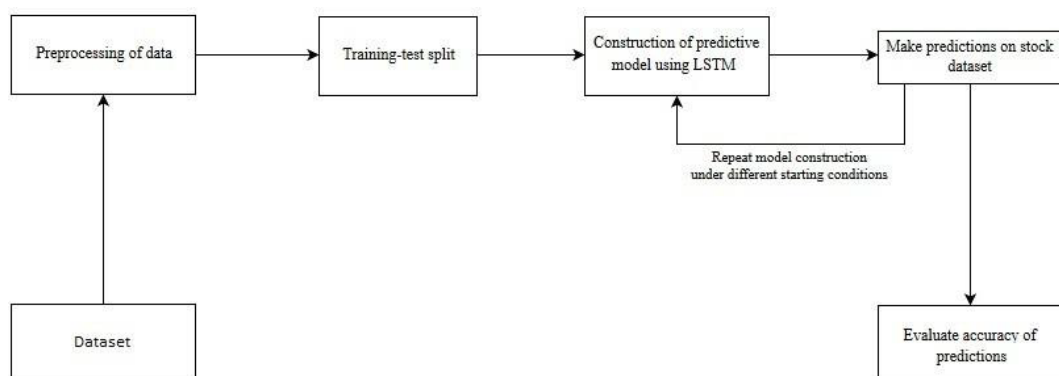
Figure 1 : Block Diagram

The proposed system will have an input from the dataset which will be extracted featured wise and Classified underneath. The classification technique used is supervised and the various techniques of machine level algorithms are implemented on the same.

Training Dataset are created for training the machine and the test cases are derived and implemented to carry out the visualization and the plotting's. The result generated are passed and visualized in the graphical form.

3.2.1 Supervised Machine Learning

Supervised learning is the types of machine learning in which machines are trained using well "labeled" training data, and on basis of that data, machines predict the output. The labeled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.
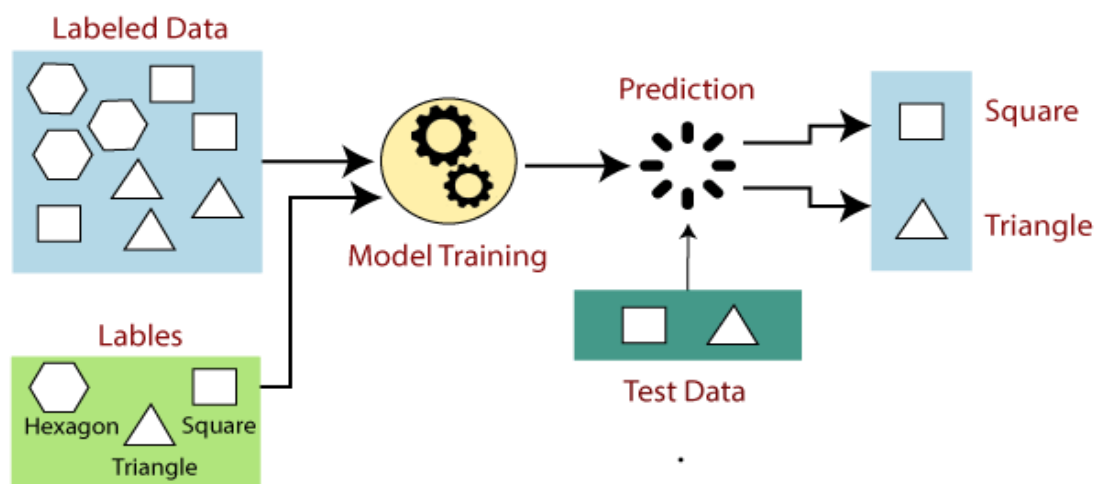

Figure 2 : Supervised machine learning

3.2.2 Unsupervised Machine Learning

Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

Example: Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs. The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset. The task of the unsupervised learning algorithm is to identify the image features on their own. Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.
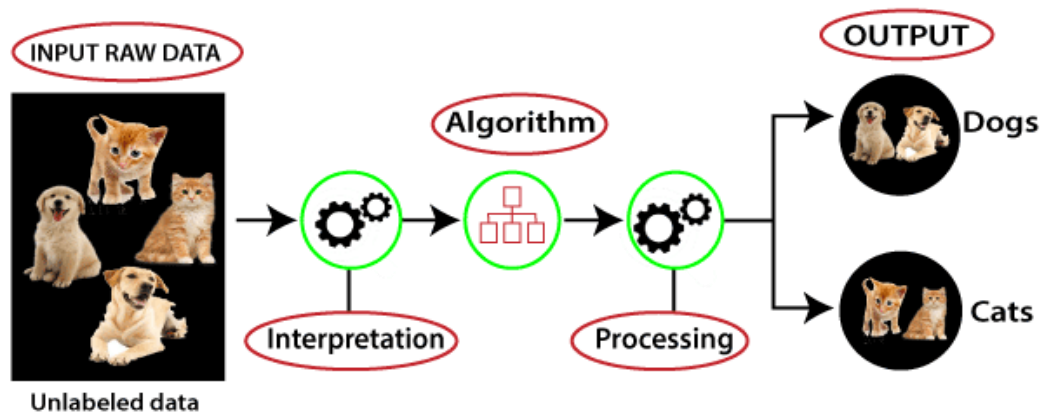

Figure 3: Un-Supervised machine learning

CHAPTER 4

SYSTEM REQUIREMENT

4. System Requirements

4.1 Hardware Requirements

- RAM : 4 GB (Minimum)
- Storage : 500 GB
- CPU : 2 GHz or faster
- Architecture : 32-bit or 64-bit

4.2 Software Requirements

- Python 3.5 in Google colab is used for data pre-processing, model training and prediction.
- Operating System: windows 7 and above or Linux based OS or MAC OS.

Python : -

Python is a general-purpose object-oriented programming language with high- level programming capabilities. It has become famous because of its apparent and easily understandable syntax, portability and easy to learn. Python is a programming language that includes features of C and Java. It provides the style of writing an elegant code like C, and for object-oriented programming, it offers classes and objects like Java.

- Python was developed in the late eighties, i.e., late 1980's by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in the Netherlands as a successor of ABC language capable of exception handling and interfacing.
- Python is derived from programming languages such as ABC, Modula 3, small talk, Algol-68.
- Van Rossum picked the name Python for the new language from a TV show, Monty Python's Flying Circus.
- Python page is a file with a .py extension that contains could be the combination of HTML Tags and Python scripts.
- In December 1989 the creator developed the 1st python interpreter as a hobby and then on 16 October 2000, Python 2.0 was released with many new features.
- On 3rd December 2008, Python 3.0 was released with more testing and includes new features.
- Python is an open source scripting language.
- Python is one of the official languages at Google.

CHAPTER 5

IMPLEMENTATION

5.1 Implementation

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables. More specifically, Regression analysis helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed. It predicts continuous/real values such as temperature, age, salary, price, etc.

This project can run on commodity hardware. We ran entire project on an Intel I5 processor with 8 GB Ram, 2 GB Nvidia Graphic Processor, It also has 2 cores which runs at 1.7 GHz, 2.1 GHz respectively. First part of the is training phase which takes 10-15 mins of time and the second part is testing part which only takes few seconds to make predictions and calculate accuracy.
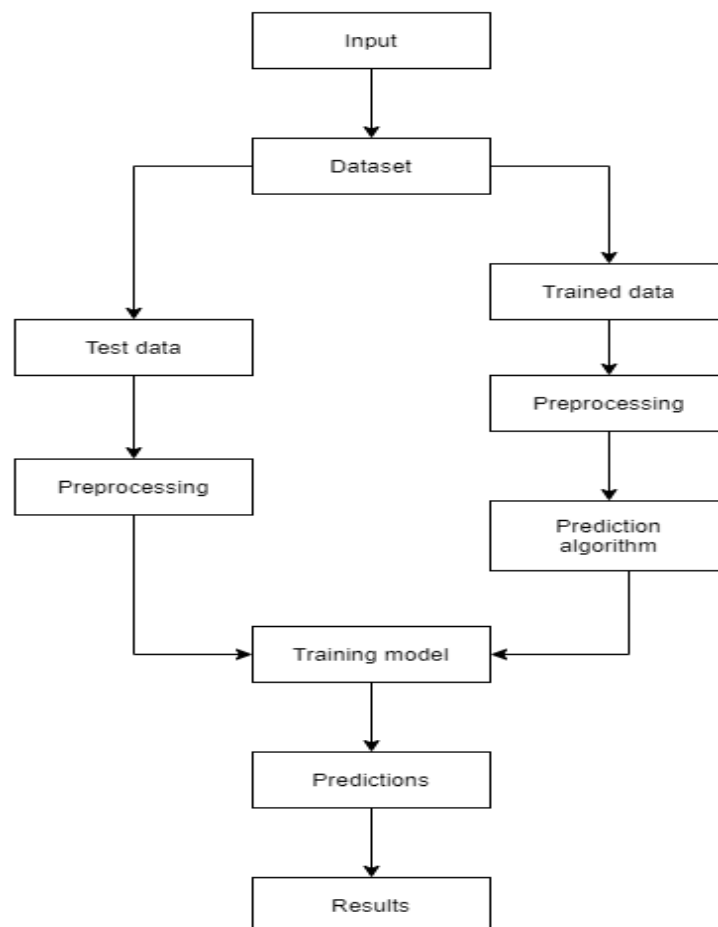
Figure 4: Flowchart

5.2 Algorithms

5.2.1 Linear Regression :

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image :
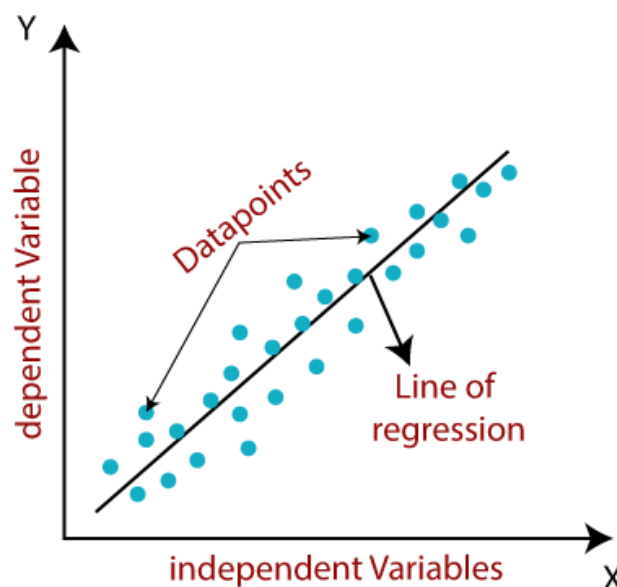

Figure 5 : Linear Regression

5.2.1 LSTM (Long Short Term Memory) :

Long Short Term Memory is kind of recurrent neural network. In RNN output From the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicrting, and classifying on the basis of time-series data.
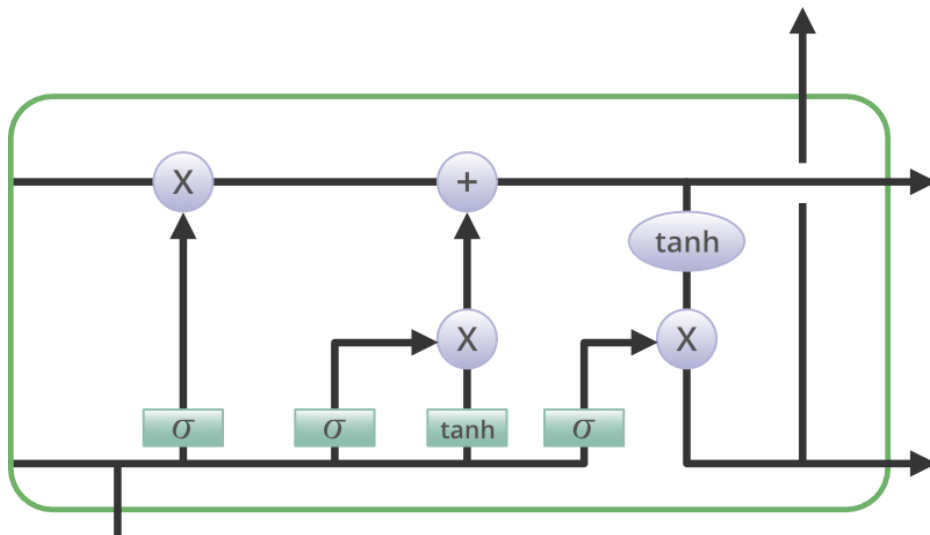
Figure 6 : LSTM Architecture

CHAPTER 6

RESULTS

6.1 Results

Visualization of Actual Price and Predicted Price

```
In [36]: graph= dfr.head(30)
```
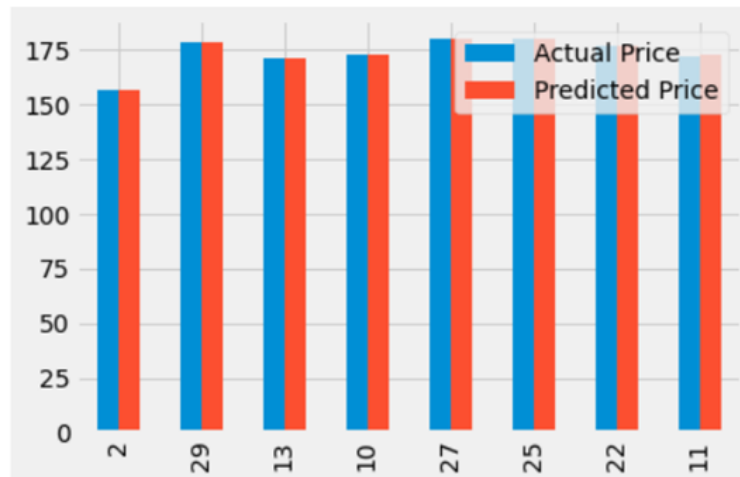
```
In [37]: graph.plot(kind='bar')
```

Out[37]: <AxesSubplot:>

Figure 7 : Candalistic Output

```
In [34]: plt.plot(df1)
```

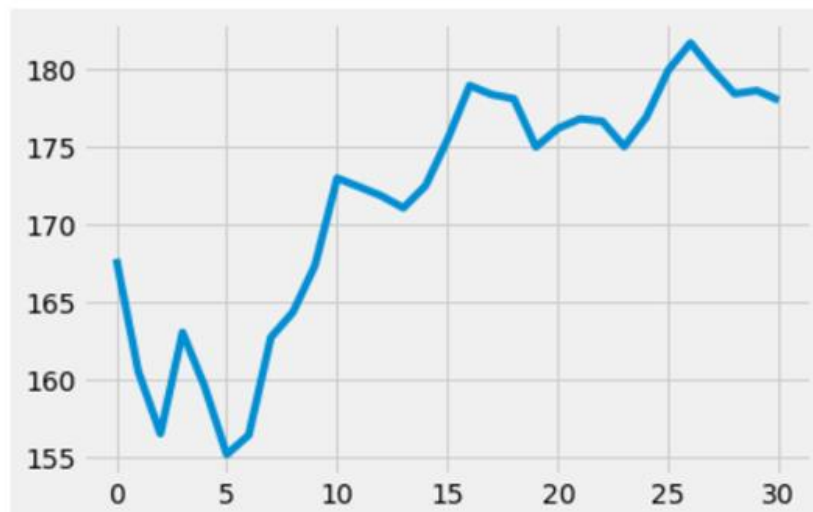Out[34]: [<matplotlib.lines.Line2D at 0x24d0e833188>]

Figure 8 : Graphical Output

CHAPTER 7
CONCLUSION

7.1 Conclusion

- The smarter solutions help our world to become smarter and have a secure and bright future ahead.

- With an optimistic vision our project will help in reduction of losses and maximization of profit.

- In this Stock Market Price Prediction project, we use Machine Learning and LSTM model to analyze the past prices of a stock , process it and then forecast the price in upcoming future.

- This concept is used to describe the path for Stock Market Trading.

CHAPTER 8
REFERENCES

8. References

[1] Pramod, & Pm, Mallikarjuna. . Stock Price Prediction Using LSTM. Test Engineering and Management. 83. 5246-5251. (2021).

[2] Shen, J., Shafiq, M.O. Short-term stock market price trend prediction using a comprehensive deep learning system. J Big Data 7, 66 (2020).

[3] Shriram. S, Dr. K. Anuradha, Dr. K. P. Uma, 2021, Future Stock Price Prediction using Recurrent Neural Network, LSTM and Machine Learning,(2021).

[4] Li, Y., Pan, Y. A novel ensemble deep learning model for stock prediction based on stock prices and news. (2021).

[5] Seethalakshmi, R. Analysis of Stock Market Predictor Variables using Linear Regression Analysis. Int. J. Pure Appl. Math. (2020).

[6] Krishna, V. ScienceDirect ScienceDirect NSE Stock Stock Market Market Prediction Prediction Using Using Deep-Learning (2021)

[7] Chandar, S.K. Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction. (2019)

[8] Strader, T.J.; Rozycki, J.J.; Root, T.H.; Huang, Y.H.J. Machine Learning Stock Market Prediction Studies: Review and Research Directions. J. Int. Technol. Inf. Manag.(2020)

[9] Zhang, C.; Cheng, J.; Tang, X.; Sheng, V.S.; Dong, Z.; Li, J. Novel DDoS Feature Representation Model Combining Deep Belief Network and Canonical Correlation Analysis. Comput. Mater. Contin. (2019)

[10] Ganser, A.; Hollaus, B.; Stabinger, S. Classification of Tennis Shots with a Neural Network Approach. 2021