\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



***A Mini-Project Report On***

***“STOCK PREDICTION SYSTEM”***

**Submitted By**

*Piyush Ramesh Datir 1132200500*

*Aditya Gajanan Dorwekar 1132200246*

*Aditya Digambar Joshi 1132200230*

*Aniket Ratnakar Pawar 1132200111*

*Karan Uday Bhelkar 1132200658*

*Ashutosh Pangaonkar 1132200157*

***SY M.Sc. (Data Science and Big Data Analytics)***

***School of Computer Science Faculty of Science***

***MIT – World Peace University Pune - 411038***

***Academic Year 2021-2022***

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**MIT WORLD PEACE UNIVERSITY, PUNE SCHOOL OF COMPUTER SCIENCE**

***Certificate***

This is to certify that

**Aditya Gajanan Dorwekar**

Of ***M.Sc. (Data Science and Big Data Analytics)*** successfully completed his/her Mini Project in

**“STOCK PREDICTION SYSTEM”**

to our satisfaction and submitted the same during the academic year 2020- 2021 towards the partial fulfillment of degree of **Master of Science in Data Science and Big Data Analytics** of MIT World Peace University under the School of Computer Science, MIT WPU, Pune.

|  |  |
| --- | --- |
| **Prof. Dr. Shubhalaxmi Joshi** | **Prof. Prajakta Soman** |
| **Associate Dean** | **Program Head** |
| **Faculty of Science** | **School of Computer Science** |
| **MITWPU** | **MIT WPU** |

Contents

[**Introduction**](#_bookmark0) 4

[PROBLEM STATEMENT](#_bookmark1) 4

[DOMAIN NAME](#_bookmark2) 4

MOTIVATION ……………………………………………5

[**Literature Survey**](#_bookmark3) 5

[**Solution Design**](#_bookmark4) 8

[SOLUTION APPROACH](#_bookmark5) 8

[TECHNOLOGY STACK](#_bookmark6) 8

[DESIGN MODEL 8](#_bookmark7)

[**Solution Implementation and Results**](#_bookmark8) 10

[DETAILS OF DATASET BEING USED](#_bookmark9) 10

[DATA EXTRACTION USING YAHOO FINANCE](#_bookmark10) 11

OBSTACLES/TECHNICAL CHALLENGES FACED ………………………………………….11

**Results** [1](#_bookmark13)2

LSTM[1](#_bookmark14)2

[*SENTIMENT ANALYSIS*](#_bookmark15) 14

[*USER INTERFACE*](#_bookmark16) 16

[**Conclusion and Future Scope**](#_bookmark21) 18

CONCLUSION ………………………………………….18

[FUTURE SCOPE](#_bookmark22) 18

[**References**](#_bookmark23) 19

**INTRODUCTION**

***PROBLEM STATEMENT: STOCK PREDICTION USING DEEP LEARNING AND SENTIMENTAL ANALYSIS***

Stock price prediction is a classic and important problem. With a successful model for stock prediction, we can gain insight about market behaviour over time, spotting trends that would otherwise not have been noticed. With the increasingly computational power of the computer, machine learning will be an efficient method to solve this problem. However, the public stock dataset is too limited for many machine learning algorithms to work with, while asking for more features may cost thousands of dollars every day.

Stock market prediction methods are divided into two main categories: technical and sentimental analysis. Technical analysis focuses on analyzing historical stock prices to predict future stock values (i.e., it focuses on the direction of prices). On the other hand, sentimental analysis relies mostly on analyzing unstructured textual information like financial news and tweets on twitter using twitter API.

**DESCRIPTION OF PROBLEM STATEMENT:**

We propose a method for stock prediction based on a set of technical trading rules, which are optimized by an optimization algorithm.  The aim of the research is to check if it is possible to obtain a set of trading rules deriving from technical indicators and Sentiments of people in the market, which could be used to make decisions such as Buy, Wait or Sell

**DOMAIN NAME**: Deep learning for prediction system some features like the latest announcements about an organization, their quarterly revenue results, etc., deep learning techniques have the potential to unearth patterns and insights we didn’t see before, and these can be used to make unerringly accurate predictions.

we can work with historical data about the stock prices of a publicly listed company. We implement a mix of deep learning algorithms to predict the future stock price of this company, starting with simple algorithms like averaging and linear regression, and then move on to advanced techniques like Auto ARIMA and LSTM.

**MOTIVATION:**

* The country faced lockdown for months due to the pandemic.
* It Created issues like unemployment.
* To develop a Stock price prediction Model using DL.
* To reduce human efforts
* Hence to Maintain the financial income using the stock market.
* In order to do so, our aim is to help people in stock market-related decisions.

**LITERATURE SURVEY**

**Stock Price Prediction Using Long Short-Term Memory** - The results of comparison between Long Short-Term Memory (LSTM) and Artificial Neural Network (ANN) show that LSTM has a better prediction accuracy than ANN. An analysis of the results also indicates that both models give better accuracy when the size of the dataset increases. With more data, more patterns can be fleshed out by the model, and the weights of the layers can be better adjusted.

**Stock Market Prediction Using LSTM Recurrent Neural Network**- It has never been easy to invest in a set of assets, the abnormally of financial market does not allow simple models to predict future asset values with higher accuracy. Machine learning, which consist of making computers perform tasks that normally requiring human intelligence is currently the dominant trend in scientific research. This article aims to build a model using

Recurrent Neural Networks (RNN) and especially Long-Short Term Memory model (LSTM) to predict future stock market values. The main objective of this paper is to see in which precision a Machine learning algorithm can predict and how much the epochs can improve our model

This paper proposes RNN based on LSTM built to forecast future values for both GOOGL and NKE assets, the result of our model has shown some promising result. The testing result conform that our model is capable of tracing the evolution of opening prices for both assets. For our future work we will try to find the best sets for bout data length and number of training epochs that beater suit our assets and maximize our predictions accuracy.

1. In this paper, Data mining techniques play important role in Stock market which can search and uncover hidden patterns and increase the certain level of accuracy, where traditional and statistical methods are lacking. Data mining is a part of the Knowledge Discovery in Databases (KDD). Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Polynomial regression is a nonlinear model in data but in statistical estimation problem, it is linear, in the sense that the regression function is linear in the unknown parameters that are estimated from the data.
2. In this paper, the classification of (i) approaches with respect to features/techniques and advantages/limitations and (ii) tools with respect to the different techniques used for sentiment analysis. Sentiment analysis is a new field of research born in Natural Language Processing (NLP), aiming at detecting subjectivity in text and/or extracting and classifying opinions and sentiments. The sentiment classification is a task of classifying a target unit in a document to positive (favourable) or negative (unfavourable) class. There are three main classification levels mainly, (i) Document Level (ii) Sentence Level (iii) Aspect Level. Emoticons have been often used in combination of other techniques like Linguistic Enquiry and Word Count for building a training dataset in supervised machine learning technique. There are three approaches explored in sentimental classification, (i) Machine Learning based (ii) Lexicon based (iii) Hybrid. Tools used for sentiment analysis are (i) Emoticons (ii) LIWC (iii) SentiStrength (iv) Senti WordNet (v) SenticNet (vi) Happiness Index.
3. In this paper, they proposed crypto-currencies have been highly attracted by investors. Due to the lack of a central money authority, the stock markets are experiencing large price fluctuations based on speculation. The Random Forest was observed to perform the most successful and fastest sentiment classification process using a combination of semi-supervised feature According to our experiments we recommend at least 5,000 labelled comments for the first system training for better learning rate result pare increasing. It should not be forgotten that this proposed system design is a self-learning method First plan was to determine social media price sentiments. Later on, determination of date intervals thought to have been manipulated in the BTC market and determination of comments made at these dates test datasets as a vector will be evaluated by ML and DL methods. All these researches are planned to completed and shared results in the year 2020.
4. In this paper, they Predict of stock market is a long-time attractive topic to researchers from different fields. In particular, numerous studies have been conducted to predict the movement of stock market using machine learning algorithms such as support vector machine (SVM) and reinforcement learning. In this project, we propose a new prediction algorithm that exploits the temporal correlation among global stock markets and various financial products to predict the next-day stock trend with the aid of SVM. Numerical results indicate a prediction accuracy of 74.4% in NASDAQ, 76% in S&P500 and 77.6% in DJIA. The same algorithm is also applied with different regression algorithms to trace the actual increment in the markets. Finally, a simple trading model is established to study the performance of the proposed prediction algorithm against other benchmarks.

**SOLUTION DESIGN**

**SOLUTION APPROACH:**

For example, we used Apple’s daily stock price for the last 10 years downloaded in CSV format from Yahoo Finance. This dataset is made by the High, Low, Open and Adjusted close price, in addition to the volume.

The price itself is not enough to produce useful predictions. For example, we added the following indicators that will be used as inputs for our models:

* Moving averages (the average price of the last N days)
* The moving average plus and minus a certain number of standard deviations.
* The rolling maximum and minimum prices of the last N days.

As usual, we’ll split our dataset into training and test sets. The training set will be used to train our models and tune the hyperparameters, while the test set will be used only for performance calculations.

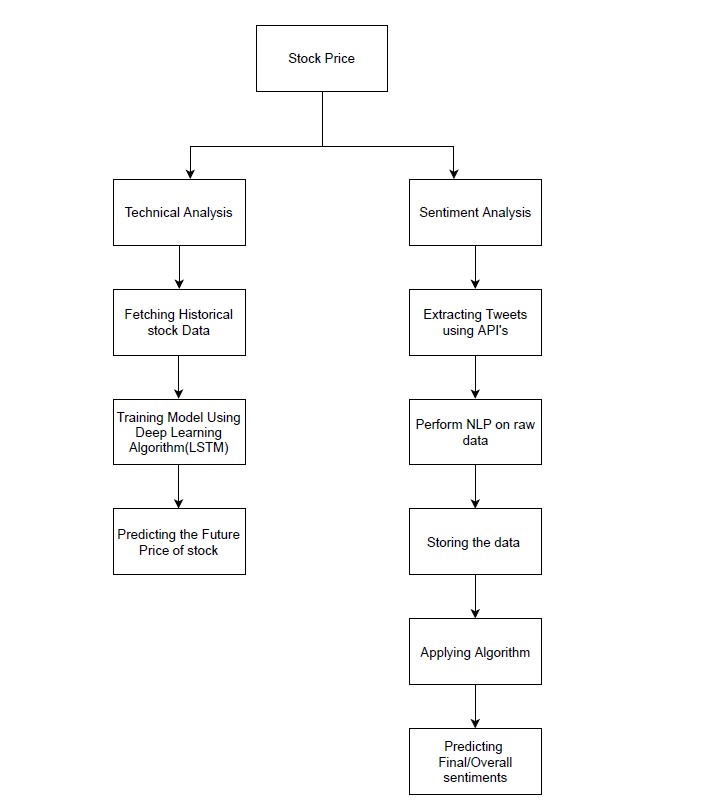
After applying desired algorithms, we get our desired output that is the predicted stock price.

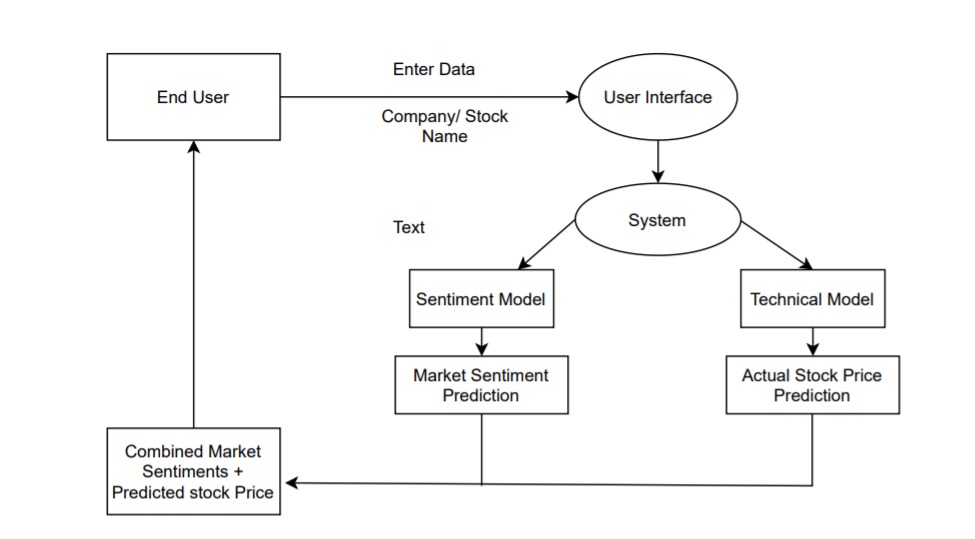
**TECHNOLOGY STACK:**

|  |  |
| --- | --- |
| SOFTWARES USED | JUPYTER NOTEBOOK, PYCHARM |
| LANGUAGES USED | PYTHON 3.7 |
| LIBRARIES USED | PANDAS, SKLEARN, MATPLOTLIB, KERAS, NLTK, NUMPY |
| FRAMEWORK USED | FLASK |
| MODEL USED | LSTM |

**DESIGN MODEL:**

Flow diagram / Architecture of the system:

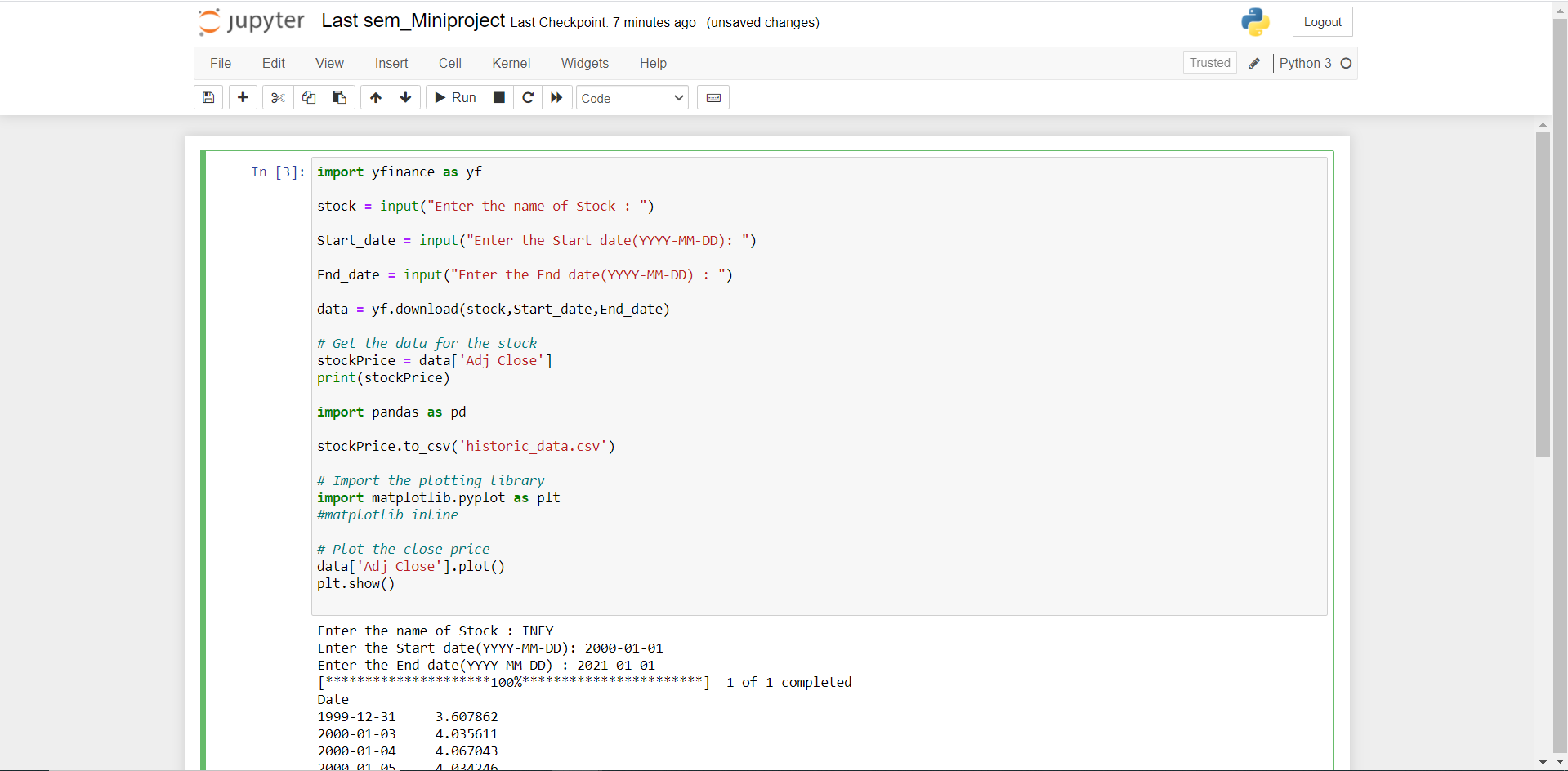


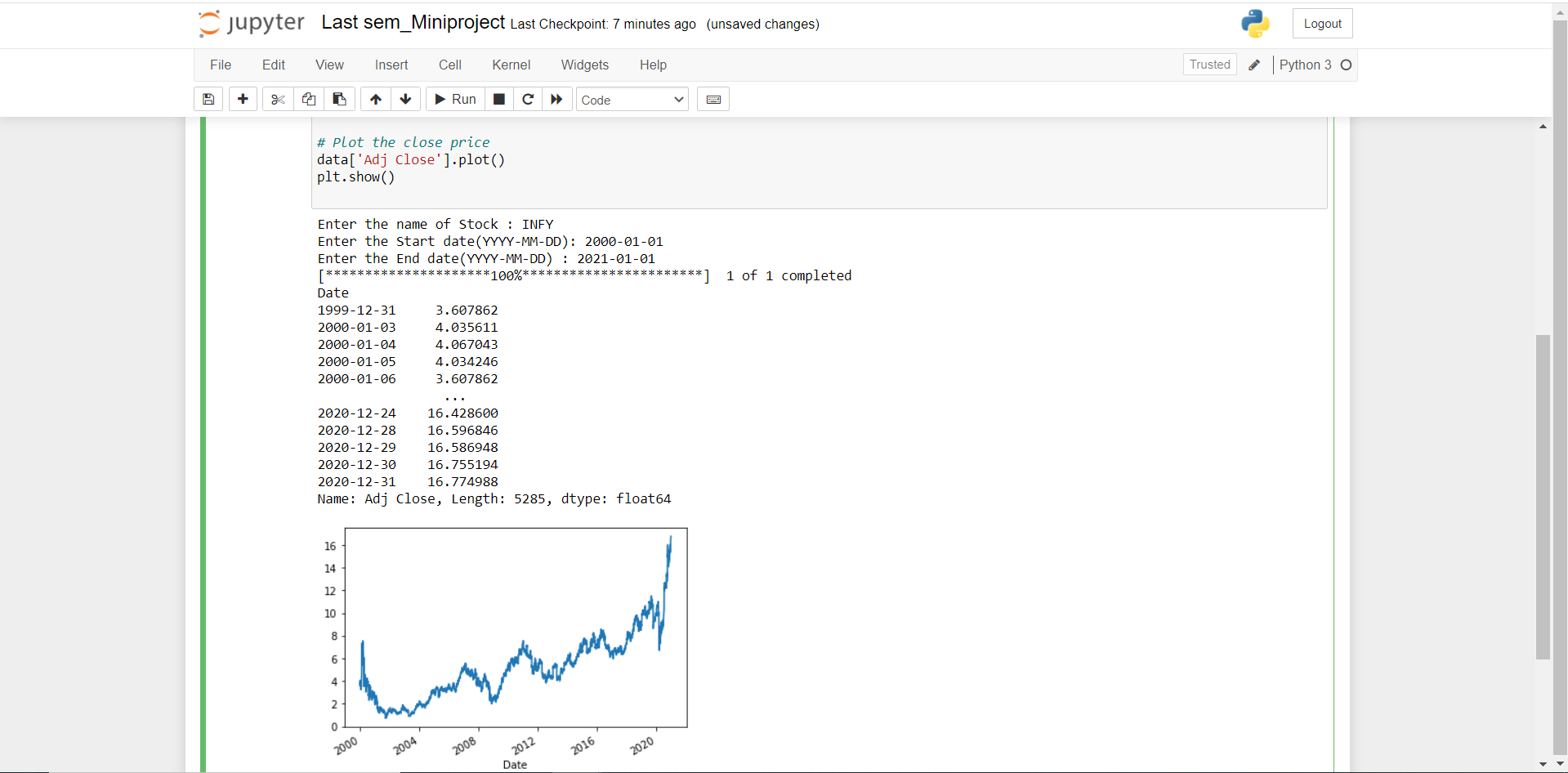


**DETAILS OF DATASET BEING USED:**

We manually tried extracting stock data for a particular company. So, it was showing a limited number of companies for prediction. So, we decided to go with stock APIs. Moreover, most of the API’s were paid APIs. Finally, we decided to fetch our stock data using Yahoo Finance API

**DATA EXTRACTION USING YAHOO FINANCE:**





**OBSTACLES / TECHNICAL CHALLENGES FACED:**

We manually tried extracting stock data for a particular company. So, it was showing a limited number of companies for prediction. So, we decided to go with stock APIs. Moreover, most of the API’s were paid APIs

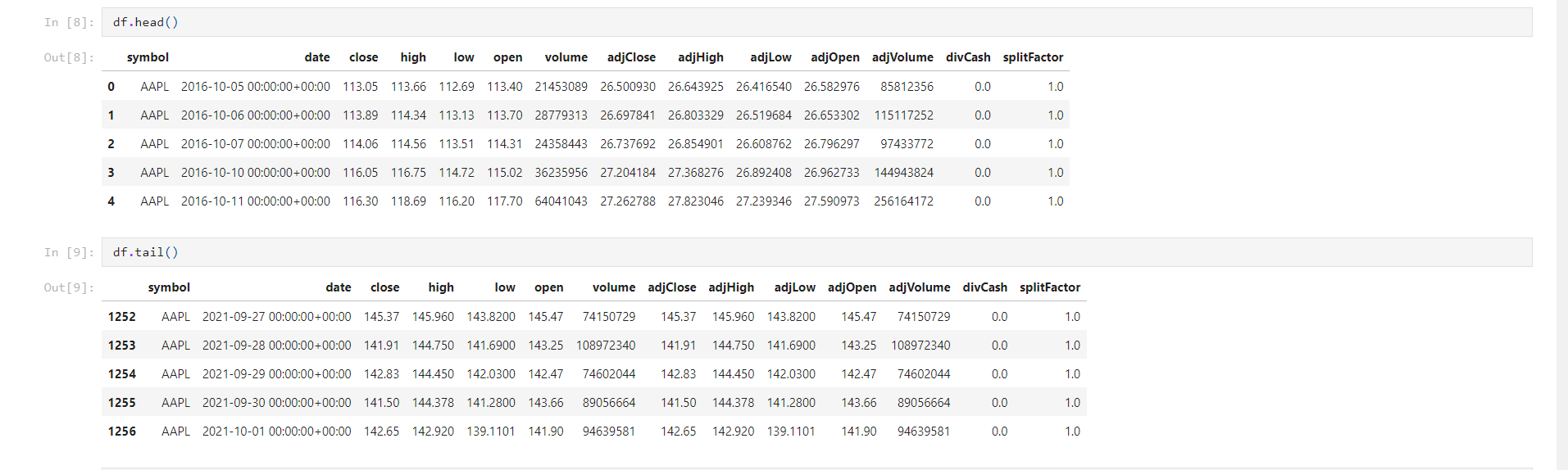
**LSTM:**

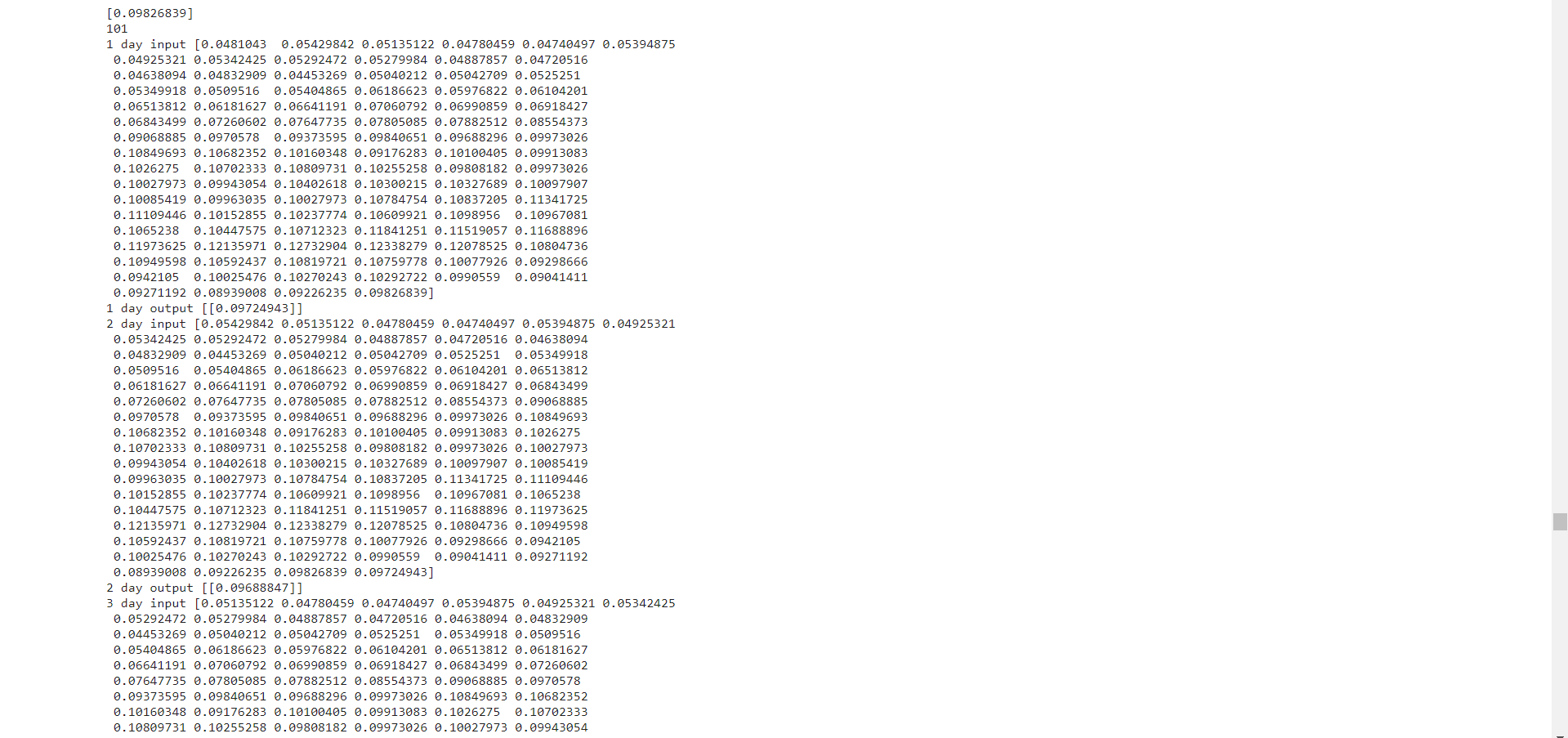
* **Long short-term memory** (**LSTM**) *is an artificial recurrent neural network (RNN) architectureused in the field of deep learning*

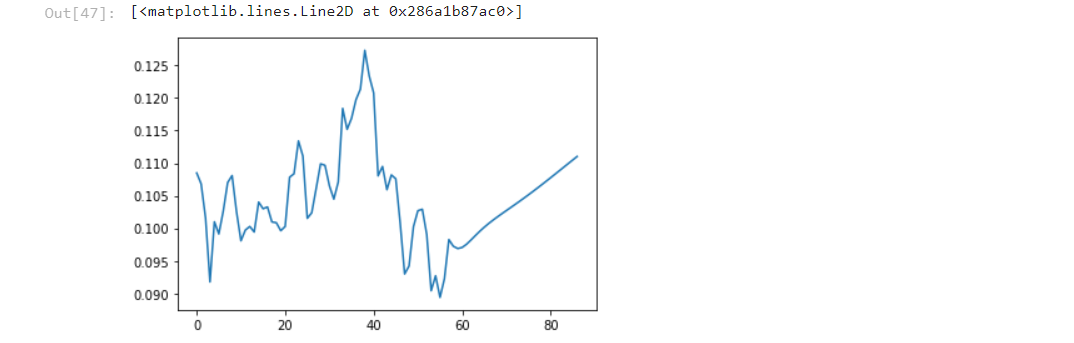
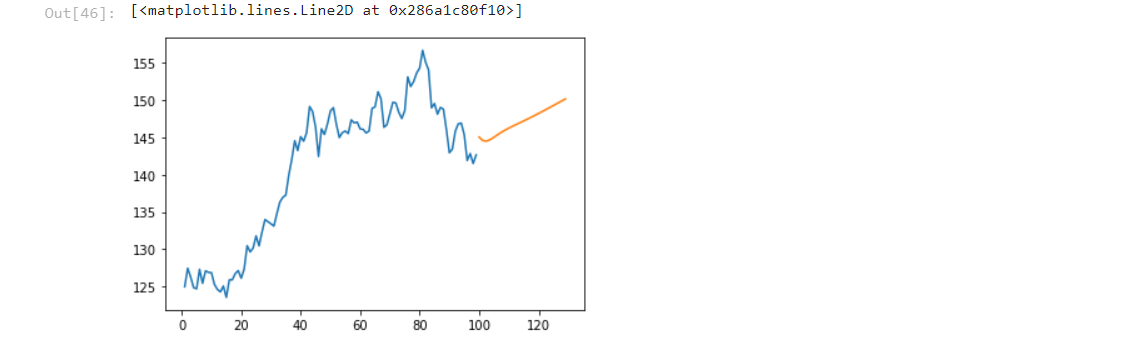
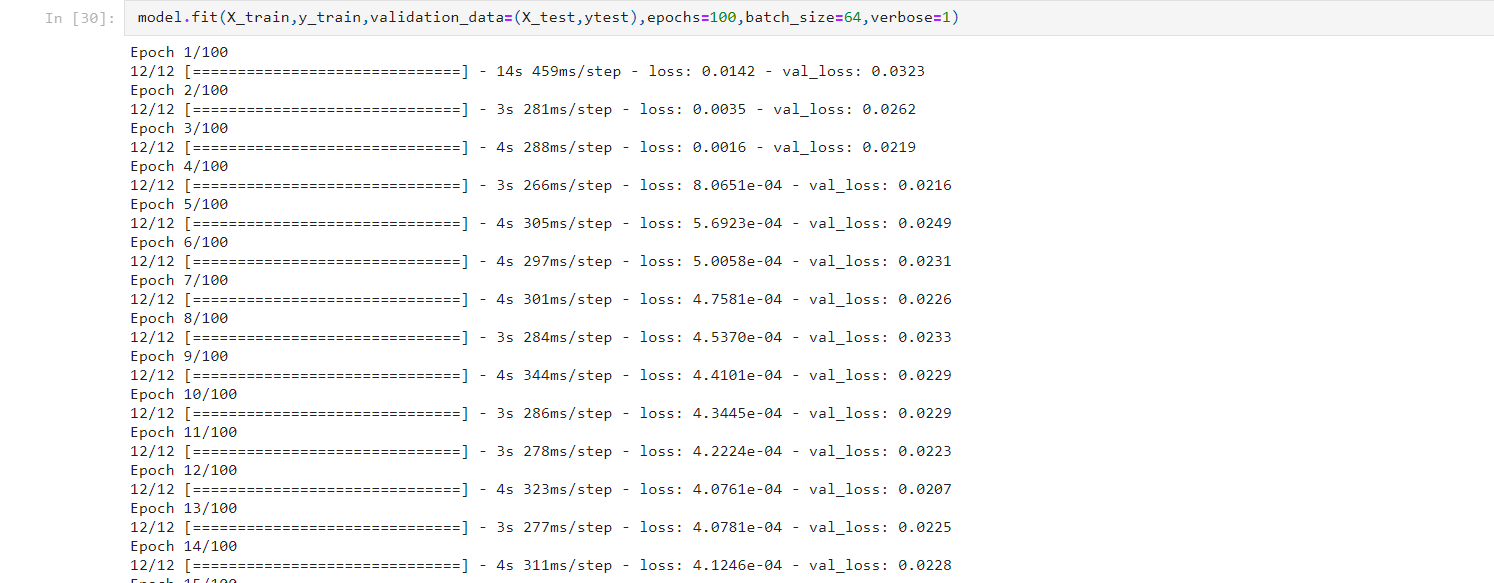
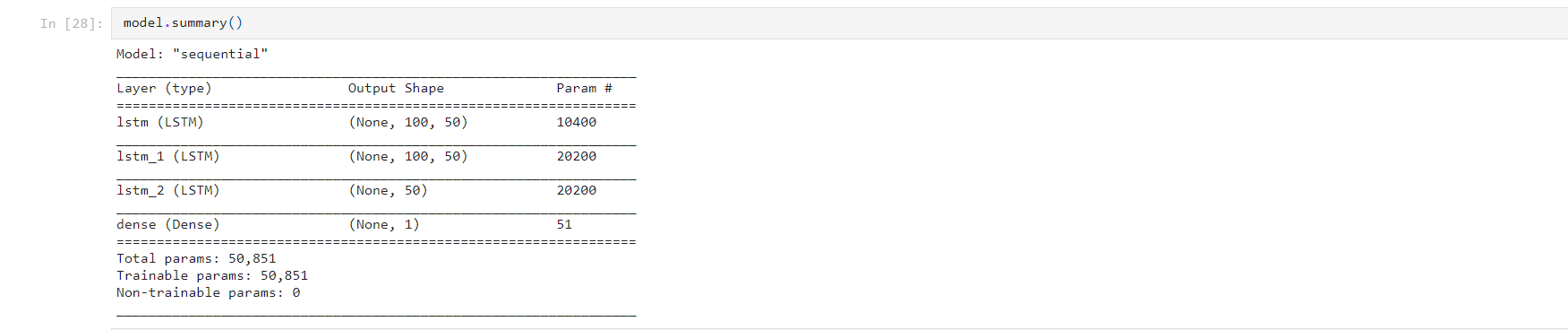
*LSTMs are very powerful in sequence prediction problems because they’re able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price*

In order to build the LSTM, we need to import a couple of modules from Keras:

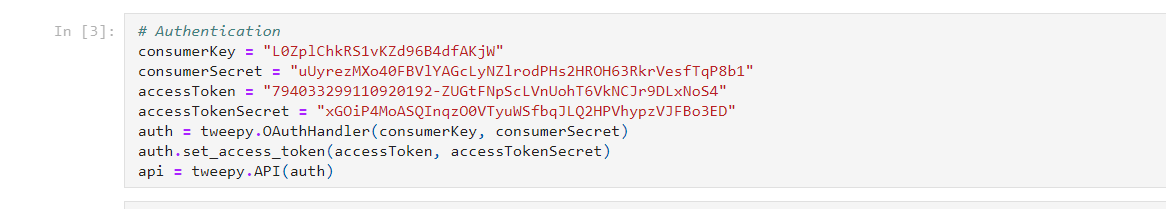
1. Sequential for initializing the neural network
2. Dense for adding a densely connected neural network layer
3. LSTM for adding the Long Short-Term Memory layer
4. Dropout for adding dropout layers that prevent overfitting

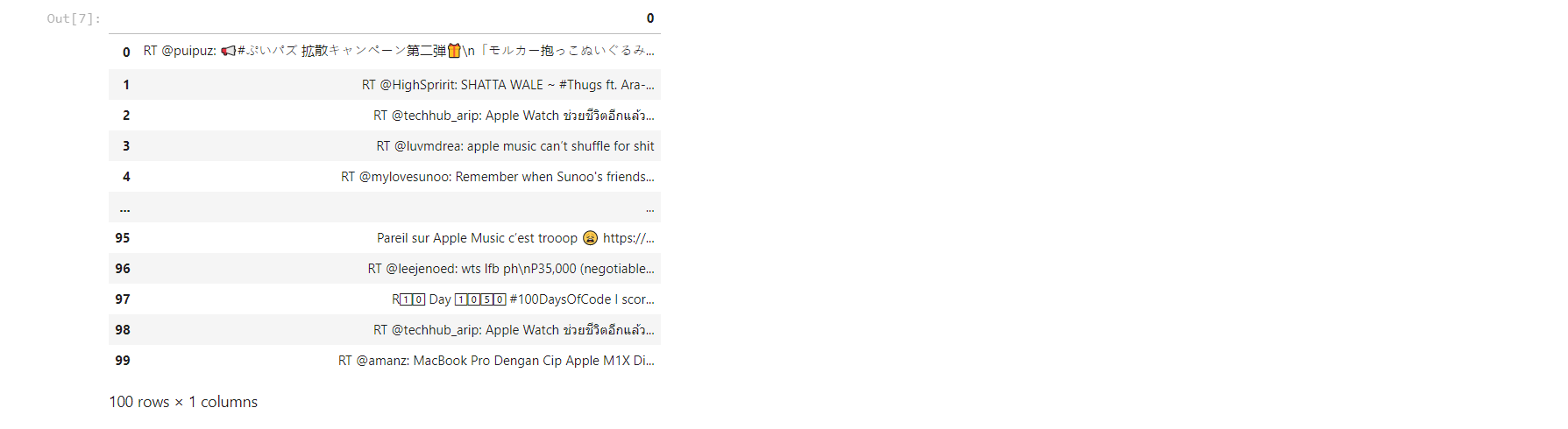






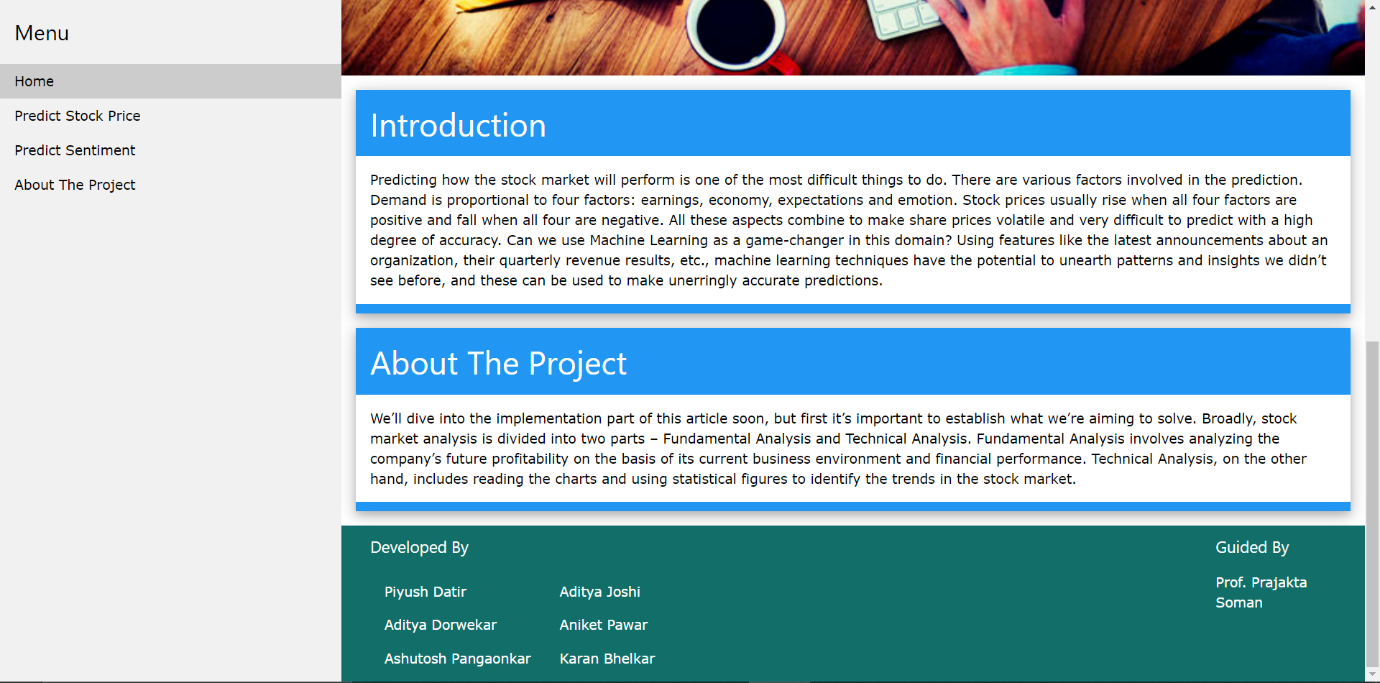
**SENTIMENT ANALYSIS:**

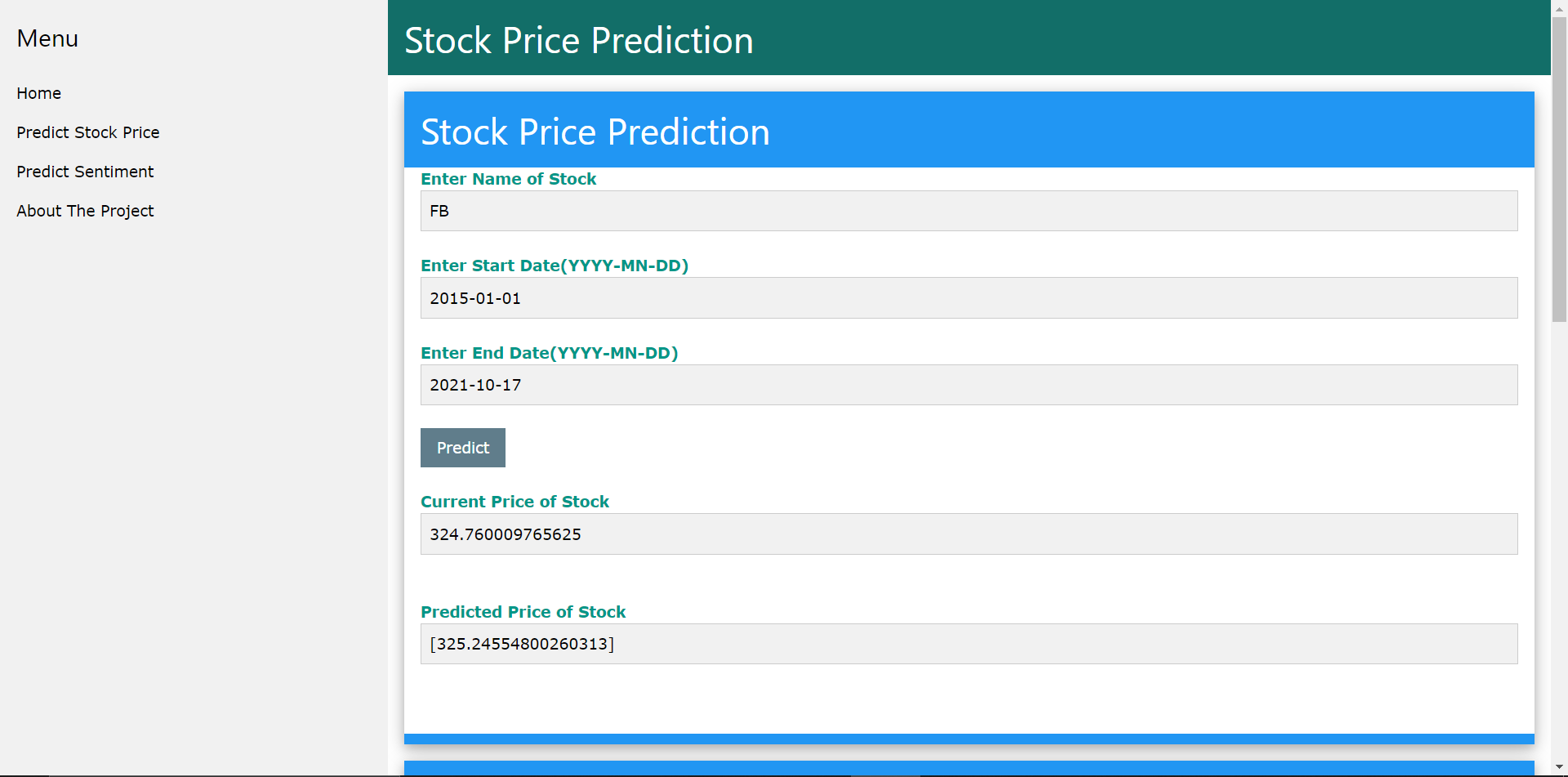
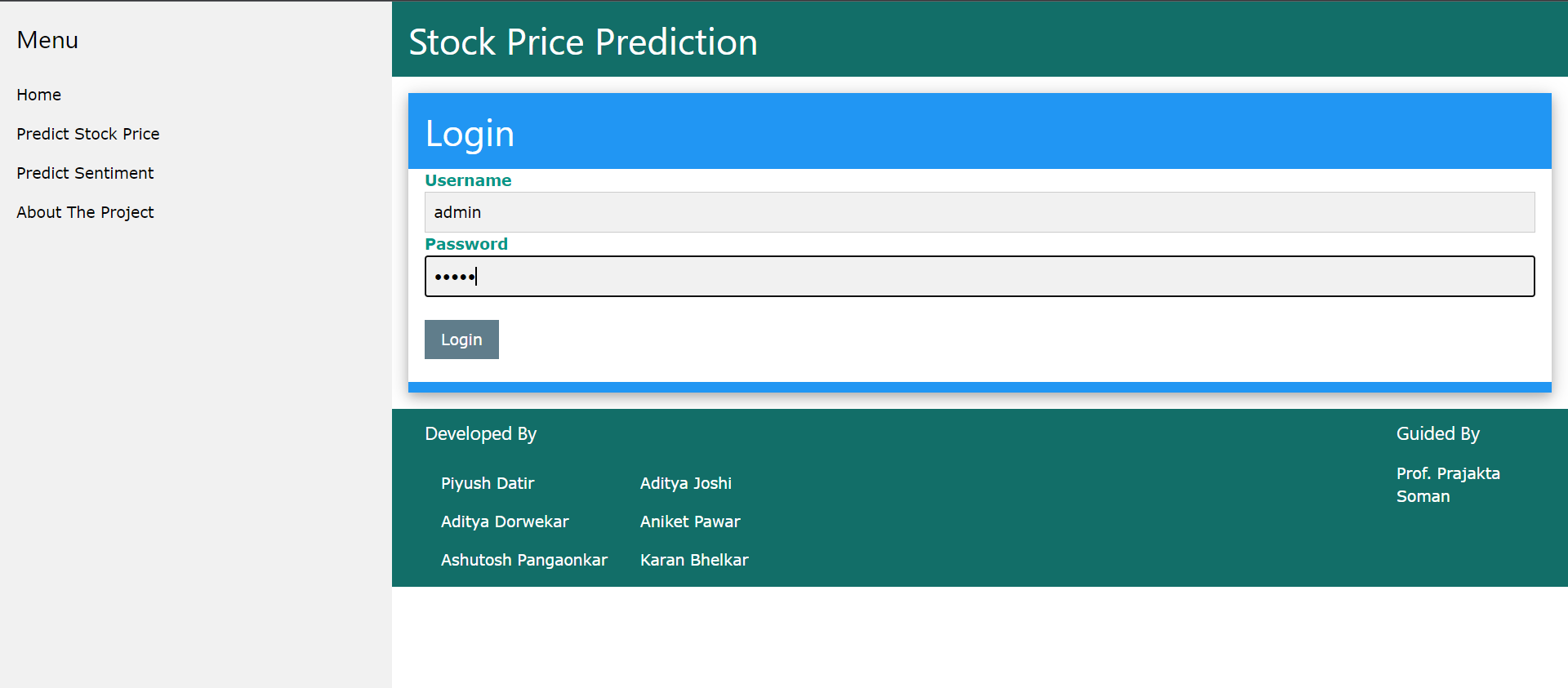




USER INTERFACE:







**CONCLUSION:**

The proposed system will help beginner traders as a decision support tool and

help them take decisions accordingly. Collective analysis of news regarding

market have been made easy using deep learning algorithms.

**FUTURE SCOPE:**

1. We have considered only twitter data for analysing people’s sentiments

which may be biased because not all the people who trade share their

opinions on Twitter. The study can be extended by incorporating data

from various platforms like moneycontrol.com, stock twits, Yahoo

Finance, etc.

2. This project is the initial phase of development where the algorithms

are selected and are being developed. As we proceed further the algorithm will be optimized for more technical parameters and more com- plex

parameters, to get much more real-world output values. And also, we can

extend this project for other markets like foreign stock markets,

commodities, Forex trading, etc

**REFERENCE:**

[1]Pahwa, K., & Agarwal, N. (2019). Stock Market Analysis using Supervised

Machine Learning.

2019 International Conference on Machine Learning, Big Data, Cloud and Parallel

Computing.

[2]Akba, F., Medeni, I. T., Guzel, M. S., & Askerzade, I. (2020). Assessment of

Iterative SemiSupervised Feature Selection Learning for Sentiment Analyses: Digital Currency

Markets.

2020 IEEE 14th International Conference on Semantic Computing

[3]Sarode, S., Tolani, H. G., Kak, P., & Lifna, C. S. (2019). Stock Price Prediction

Using Machine

Learning Techniques. 2019 International Conference on Intelligent Sustainable

Systems

[4] Ravikumar, S., & Saraf, P. (2020). Prediction of Stock Prices using Machine

Learning

(Regression, Classification) Algorithms. 2020 International Conference for Emerging

Technology

[5] Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., & Lifna, C. . (2019).

Stock Market

Prediction based on Social Sentiments using Machine Learning. 2018 International

Conference

on Smart City and Emerging Technology

[6]Kim, J., Seo, J., Lee, M., & Seok, J. (2019). Stock Price Prediction Through the

Sentimental

Analysis of News Articles. 2019 Eleventh International Conference on

Ubiquitous and Future

Networks

[7]Gupta, R., & Chen, M. (2020). Sentiment Analysis for Stock Price Prediction.

2020 IEEE

Conference on Multimedia Information Processing and Retrieval

[8]Moghaddam A H , Moghaddam M H , Esfandyari M . Stock market index

prediction using

artificial neural network:[J]. Journal of Economics Finance & Administrative

Science.

Thank You.