Khulna University of Engineering & Technology
খুলনা প্রকৌশল ও প্রযুক্তি বিশ্ববিদ্যালয়

CSE-3200: System Development Project

Stock market price prediction using ARIMA-LSTM hybrid model

by

Umme Israt Afroz
Roll: 1807043
Pushan Paul
Roll: 1807065

Supervisor:
Mr. Sunanda Das
Assistant Professor

Department of Computer Science and Engineering
Khulna University of Engineering & Technology

December 28, 2022

Approval

This Project Report has been submitted for examination with the approval of our supervisor.

Mr. Sunanda Das
Assistant Professor
Dept. of Computer Science and Engineering
Khulna University of Engineering Technology

Acknowledgement

With heartfelt gratitude, We, **Umme Israt Afroz** and **Pushan Paul** would like to express our deepest appreciation to all those who provided us the possibility to complete this project on “A Hybrid Approach for Predicting Stock Price using ARIMA-LSTM Network” as a part of CSE 3200: System Development Project. Our appreciation also goes out to our family and friends for their encouragement and support all through our studies.

A special gratitude to our project supervisor,**Mr. Sunanda Das**, whose contribution in stimulating suggestions and encouragement, helped us to coordinate our project especially in writing this report

Abstract

Inspite of its great importance, there has been no general consensus on how to model the trend and the seasonal component in time-series data. Box and Jenkins auto-regressive integrated moving average (ARIMA) is one of the more popular linear models in time series forecasting of the past four decades. Since time- series data is sequential, we will be focusing more on recurrent neural networks (RNN) because they are the most suited to the problem . Both the trend and the seasonal component are important to model in order to build a robust prediction. As bitcoin's price fluctuates a lot in a short period, it is very challenging to predict the bitcoin price accurately. In this paper, a hybrid methodology that combines ARIMA and LSTM models is proposed to take advantage of the unique strength of ARIMA. Experimental results with real data sets indicate that the hybrid modeling approach can be an effective way to make forecasting accuracy higher than what it would have been by either of the models used separately.

Contents

Approval **i**

Acknowledgement **ii**

Abstract **iii**

1 Introduction **1**
1.1 Background 1
1.2 Problem Statement 2
1.3 Main Objective 2

2 Related Works **3**
2.1 Introduction 3

3 Methodology **4**
3.1 Introduction 4
3.2 Data Collection and Description 4
3.2.1 Normalization 5
3.3 **The Proposed Model** 5
3.3.1 Autoregressive Integrated Moving Average(ARIMA): . 5
3.3.2 Long Short Term Memory (LSTM): 6
3.4 **Performance Evaluation** 7

4 Result Analysis **8**

5 User Interface **9**
5.1 Flow Chart of UI: 10
5.2 Implementation 11

6 Conclusion **20**

List of Figures

3.1 Work flow of the proposed method. 4
3.2 Closing Price vs Time. 5
3.3 LSTM Architecture. 7

4.1 Loss vs Epochs Curve 8

5.1 Work flow of the user interface. 10
5.2 Home Screen. 11
5.3 displaying top values of dataframe. 11
5.4 Closing price vs time plot. 12
5.5 Splitting the train and test data(manually). 12
5.6 Shape of test data. 13
5.7 Plotting ACF and PACF. 13
5.8 Observed vs forecast of ARIMA model. 14
5.9 Arima prediction closing price. 14
5.10 Arima Prediction for the train data. 15
5.11 Arima prediction against test data. 15
5.12 Enter range for prediction. 16
5.13 Showing prediction for the given range 16
5.14 Prediction vs test for the given range. 17
5.15 Loss vs epoch curve. 17
5.16 Query Params showing and details about company. 18
5.17 Providing Information and queries. 18
5.18 Model overview by clicking from the navigation bar. 19
5.19 Related Work. 19

Chapter 1

Introduction

1.1 Background

The word "stock market" refers to a number of exchanges where shares of publicly traded businesses can be bought and sold. Such financial transactions take place on formal exchanges and in over-the-counter (OTC) markets, which are governed by a specific set of rules. Frequently, the terms "stock exchange" and "stock market" are considered synonymous. On one or more of the stock exchanges that make up the larger stock market, investors purchase and sell shares of stock. The Dutch East India Corporation, the first company to trade on a stock exchange, founded the first modern stock market in Amsterdam. Securities buyers and sellers can connect, communicate, and conduct business on the stock market. The markets provide price discovery for stock in firms and act as a gauge for the state of the national economy. Because market participants compete on an open market, buyers and sellers may be sure that they will receive a fair price, a high level of liquidity, and transparency. Philadelphia hosted the country's first stock exchange in 1790. [4] The Buttonwood Agreement, which gave its name after the buttonwood tree under which it was signed, opened New York's Wall Street in 1792.The document, which was signed by 24 merchants, established the first securities trading association in the United States. In 1817, the traders changed the name of their company to the New York Stock and Exchange Board. Market participants can confidently trade shares and other permissible financial products on stock exchanges in a secure and regulated environment with little to no operational risk. The stock markets serve as primary markets and secondary markets, operating in accordance with the regulations set forth by the regulator.

1.2 Problem Statement

A hybrid model based on ARIMA-LSTM has been suggested in this research to forecast the closing price of a stock over a 1-day period.A comparison with LSTM and ARIMA models using various errors, such as RMSE and MAE, has been done in order to assess the performance of the suggested model.

1.3 Main Objective

The objectives of the project are:

- To apply deep learning and neural network in real life solution.
- To know about time series analysis.
- To know about Stock market.
- To establish a system that can predict the price of stock as accurately as possible.
- To learn about convolutional neural network and long short term memory network and how to apply them for time series analysis.
- To implement a user interface of a system.

Chapter 2

Related Works

2.1 Introduction

As bitcoins-stock are becoming more popular day by day, many researchers develop several methods to predict the bitcoin-stock price. McNally et al. proposed deep learning-based models such as recurrent neural network (RNN), LSTM for predicting the closing price of bitcoin-stock where the simple moving average (SMA) is used for feature engineering. They applied a temporal window size of 20 days and concluded LSTM network produces better results compared to RNN and ARIMA model. Ji et al. [3] demonstrated a comparative analysis employing deep learning based networks such as deep neural network (DNN), LSTM, CNN, deep residual network, and their combinations for forecasting bitcoin price. They shared that for the regression task, i.e., for the prediction of bitcoin price, LSTM networks perform better and DNN-based networks are better suited for predicting the direction of price. Aggarwal et al. [1] investigated socio-economic factors like gold price, Twitter sentiment, and different cryptocurrencies for bitcoin price prediction and proposed root LSTM and gated recurrent unit (GRU) based network. They conferred gold price has less impact on the bitcoin price prediction while Twitter sentiment may give false information about the up-down of price. Also, the LSTM model performs more reliably compared to CNN or GRU-based models.

Chapter 3

Methodology

3.1 Introduction

The proposed method is divided into several stages, including dataset collection, normalization,dataset splitting, and so on, as shown in Fig. 3.1

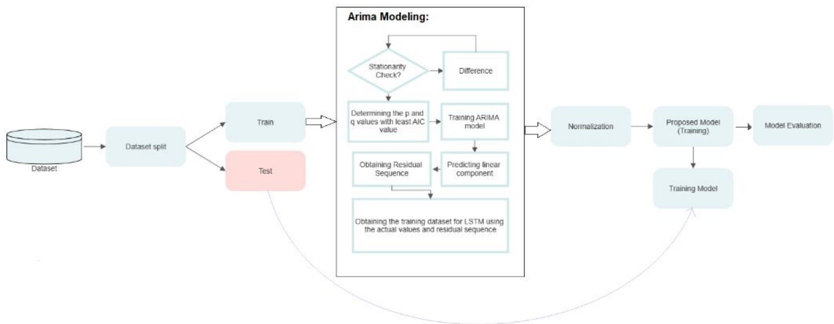

Figure 3.1: Work flow of the proposed method.

3.2 Data Collection and Description

We collected the bitcoin dataset from coincodex [5]. The dataset includes values necessary for predicting bitcoin from 10th December 2018 to 9th December 2022. The dataset consists of seven attributes, namely Date, Open, High, Low, Close, Volume, Market Cap .We have taken 'Close' columns as an independent feature for further processing, and other columns and null values are removed. An actual time series plot for bitcoin-stock the close price can be seen in Fig. 3.2

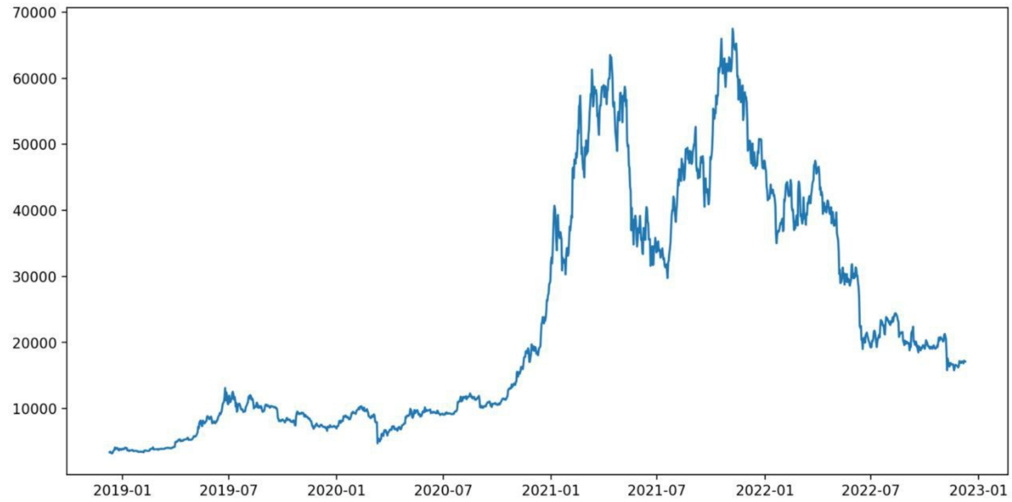

Figure 3.2: Closing Price vs Time.

3.2.1 Normalization

The min-max normalization technique has been applied to the dataset to keep the close price between the range of 0 to 1. To normalize a range between an arbitrary set of values [l, r], the min-max normalization formula is

$$p' = l + \frac{p - min(p)(r - l)}{max(p) - min(p)} \tag{3.1}$$

As $[l, r] = [0, 1]$ has been used in our method, the min-max normalization formula becomes

$$p' = \frac{p - min(p)}{max(p) - min(p)} \tag{3.2}$$

Where p' is normalized price, p is the actual price.

3.3 The Proposed Model

3.3.1 Autoregressive Integrated Moving Average(ARIMA):

ARIMA was first introduced by Box and Jenkins in 1976 in a book that received tremendous attention from the scientific community, working on research works oriented towards prediction at that time. This method is therefore applied in a wide variety of fields and remains one of the most robust models in data processing and operational prediction . ARIMA characterizes time series by going from three fundamental aspects:

- Autoregressive terms (AR) that model past process information.
- Integrated terms (I) that model the differences needed to make the process stationary.
- The moving average (MA) that controls the past information of noise around the process.Specifically, the terms AR give a representation of the series based on its p past observations.

3.3.2 Long Short Term Memory (LSTM):

The next layer of our proposed model is the LSTM layer [2], which is typically a gated Recurrent Neural Network as shown in Fig. 3.3.LSTMs work in a three-step process.

- The first step in LSTM is to decide which information to be omitted from the cell in that particular time step. It is decided with the help of a sigmoid function. It looks at the previous state (ht-1) and the current input xt and computes the function.
- There are two functions in the second layer. The first is the sigmoid function, and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function gives the weightage to the values passed, deciding their level of importance from -1 to 1.
- The third step is to decide what will be the final output. First, you need to run a sigmoid layer which determines what parts of the cell state make it to the output. Then, you must put the cell state through the tanh function to push the values between -1 and 1 and multiply it by the output of the sigmoid gate.

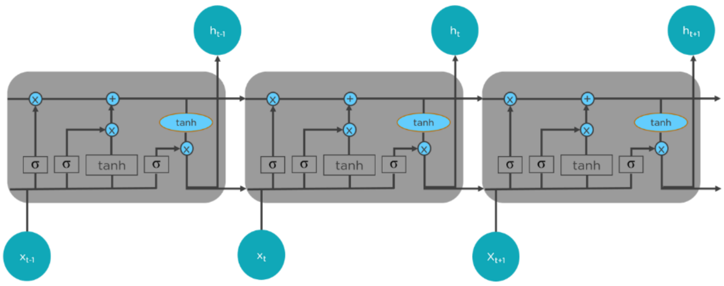

Figure 3.3: LSTM Architecture.

3.4 Performance Evaluation

To compare the proposed model with other models, we have used four types of errors, i.e., Root Mean Square Error (RMSE), Mean Absolute Error (MAE). The formulas for calculating the errors are given below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \hat{p}_i)^2} \tag{3.3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - \hat{p}_i| \tag{3.4}$$

where p is an actual value, $\hat{p}$ is the predicted value and n is the sample size.

Chapter 4

Result Analysis

After developing the model, we applied the adam as the optimizer and Mean Squared Error as the loss function for compiling the model. The model is trained for 80 epochs with a batch size of 50. The loss vs epoch curve, as shown in Fig 4.1a, indicates that the model is free from overfitting. For estimating the best learning rate for the adam optimizer, we vary the learning rate and calculate the respective loss of that rate.The loss vs epoch curve is shown in fig 4.1

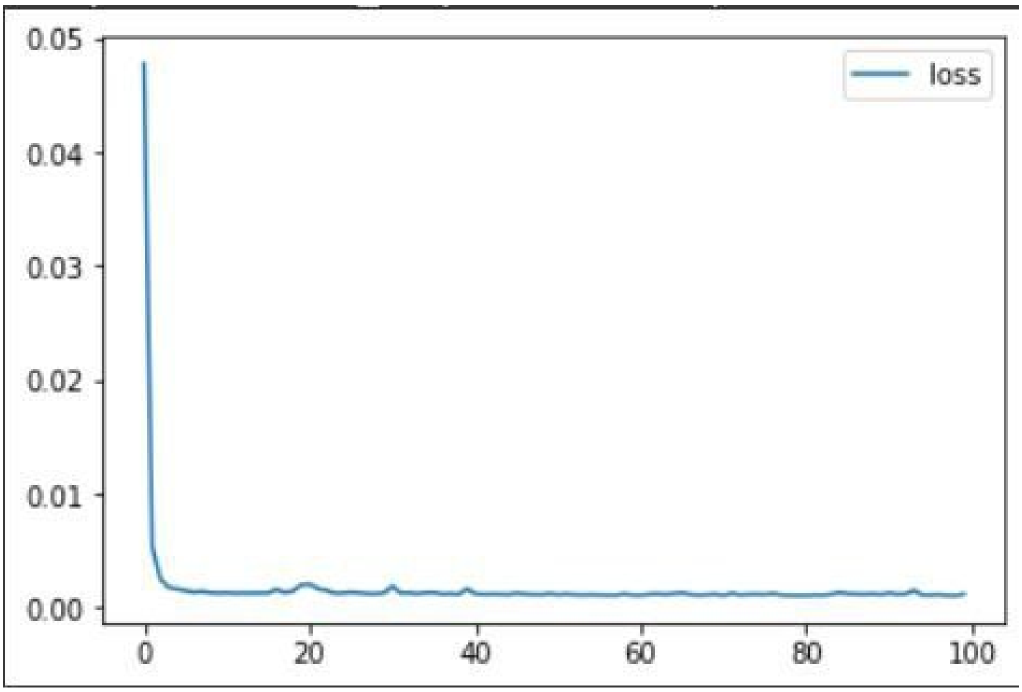

Figure 4.1: Loss vs Epochs Curve

Chapter 5

User Interface

5.1 Flow Chart of UI:

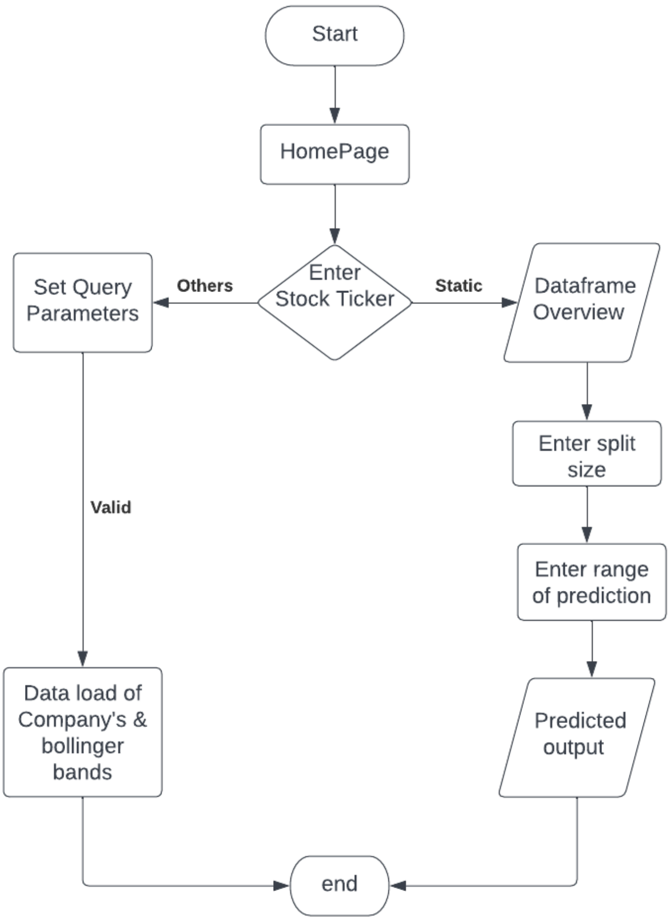

Figure 5.1: Work flow of the user interface.

5.2 Implementation

UI is made with Streamlit. Step by step user guideline is shown below:
Step 1: Start the system

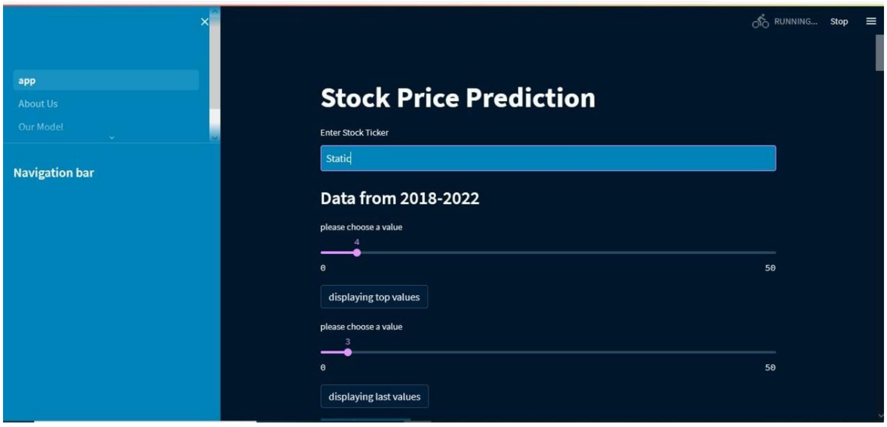

Figure 5.2: Home Screen.

Step 2 : Putting 'Static' keyword in the textfield

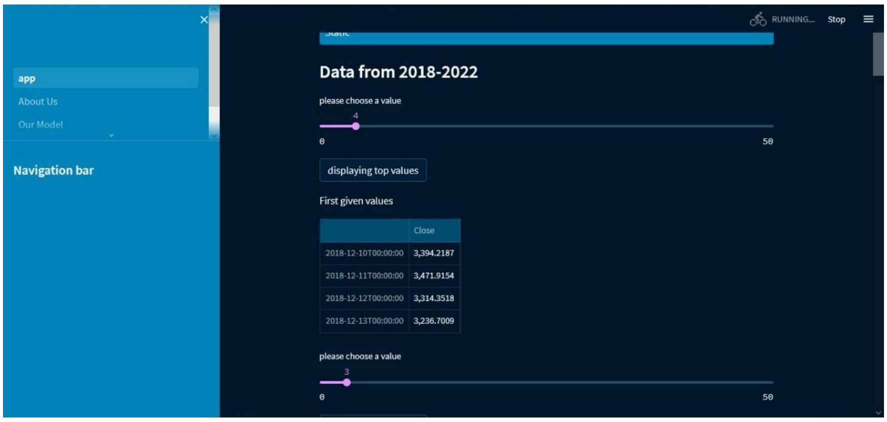

Figure 5.3: displaying top values of dataframe.

Step 3 : Showing the dataframe closing price

Figure 5.4: Closing price vs time plot.

Step 4 : Splitting the size

Figure 5.5: Splitting the train and test data(manually).

Step 5 : Extract Train and Test data

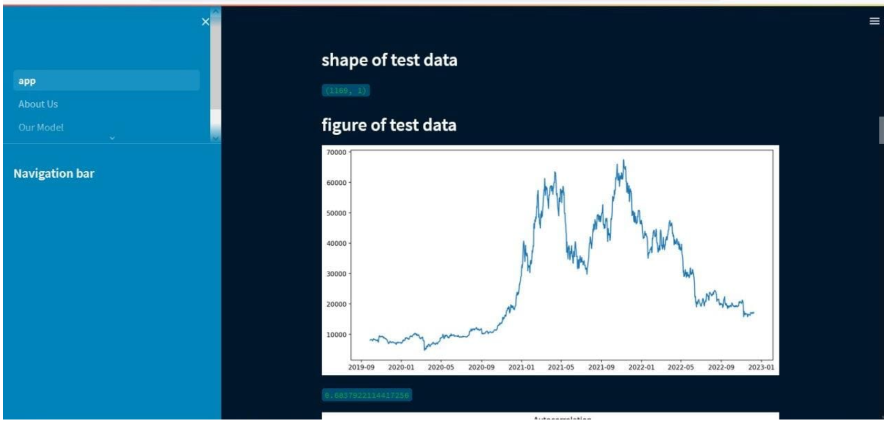

Figure 5.6: Shape of test data.

Step 6 : ACF and PACF plot

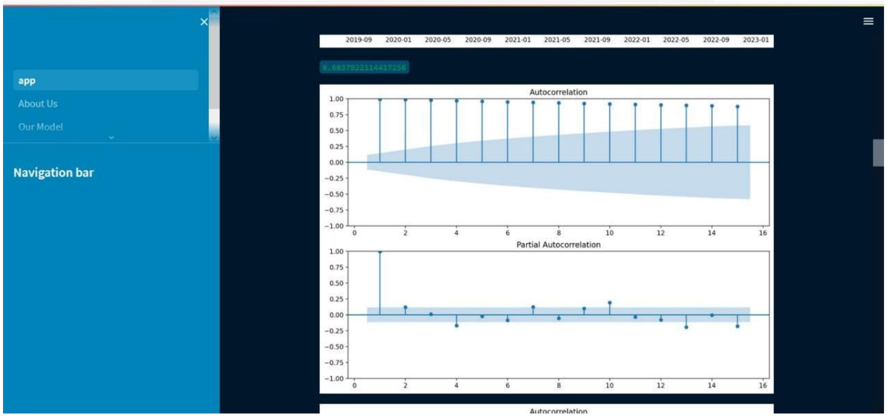

Figure 5.7: Plotting ACF and PACF.

Step 7 : Forecast of ARIMA model

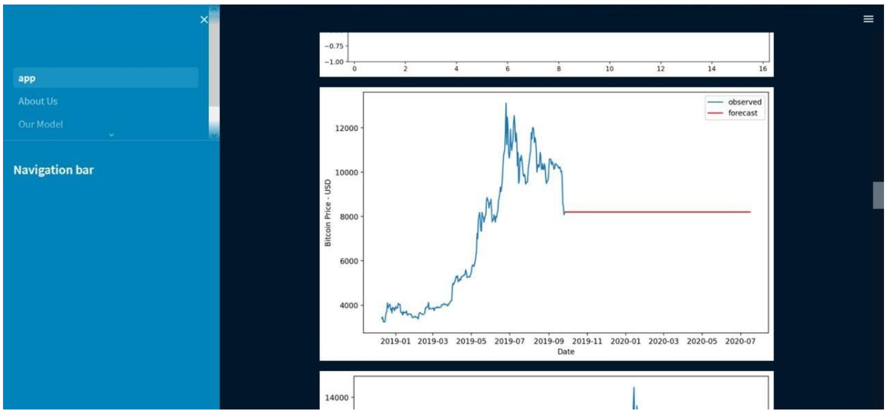

Figure 5.8: Observed vs forecast of ARIMA model.

Step 8 :ARIMA closing price prediction

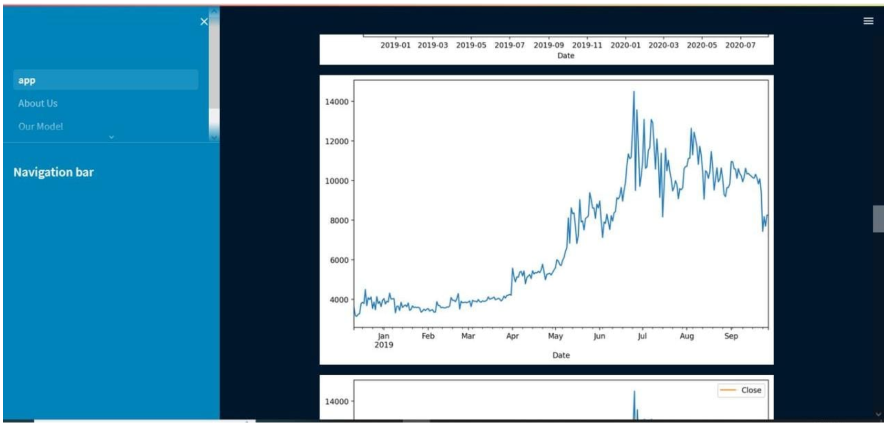

Figure 5.9: Arima prediction closing price.

Step 9 :ARIMA prediction against train and test data

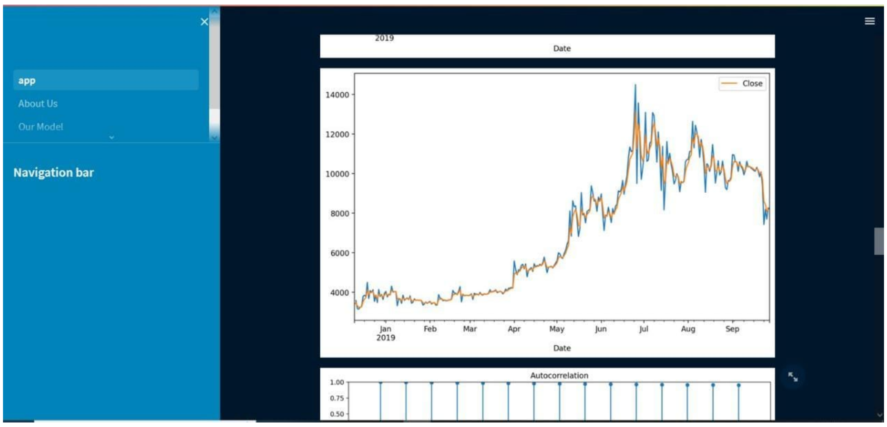

Figure 5.10: Arima Prediction for the train data.

Step 10 : Plotting

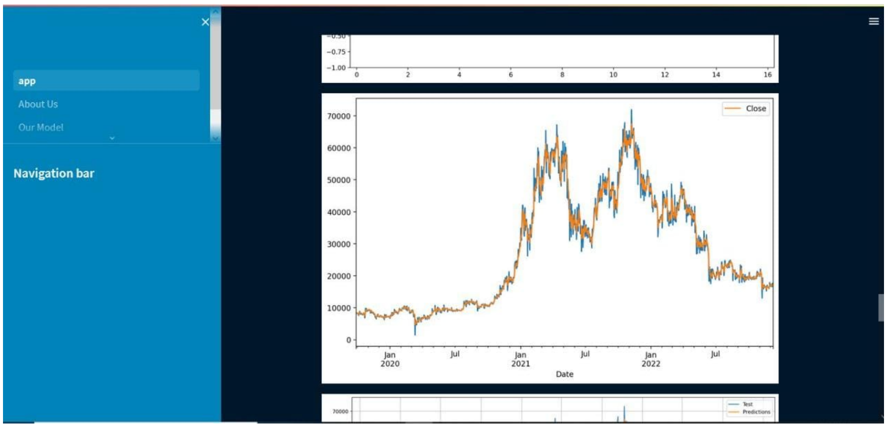

Figure 5.11: Arima prediction against test data.

Step 11 : Prediction for next given days

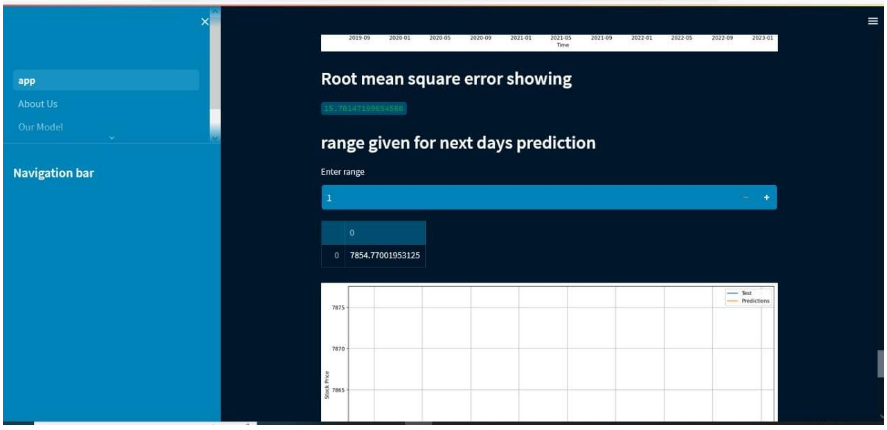

Figure 5.12: Enter range for prediction.

Step 12 : Plot

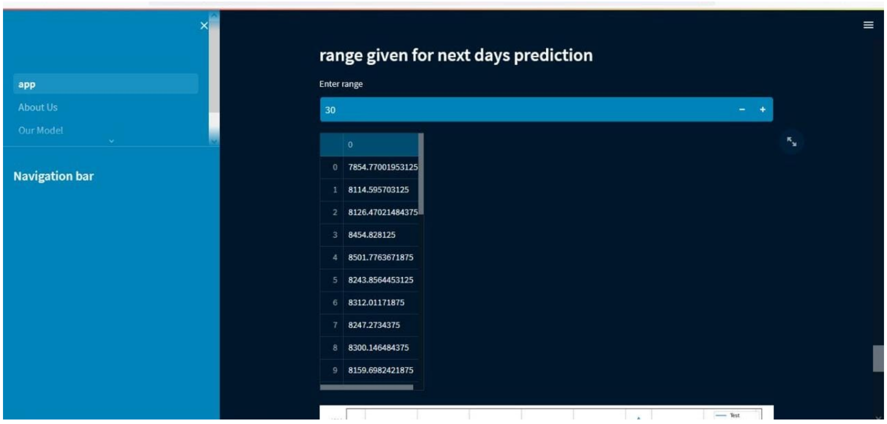

Figure 5.13: Showing prediction for the given range .

Step 8 : Plot

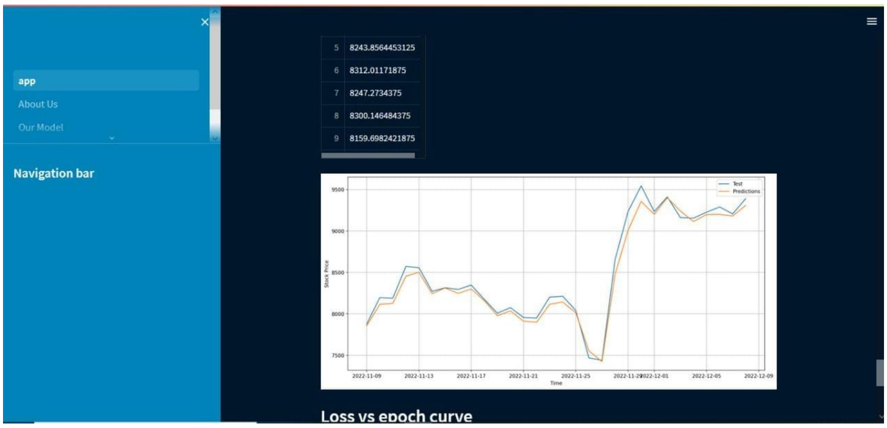

Figure 5.14: Prediction vs test for the given range.

Step 8 : Loss vs epoch curve plot

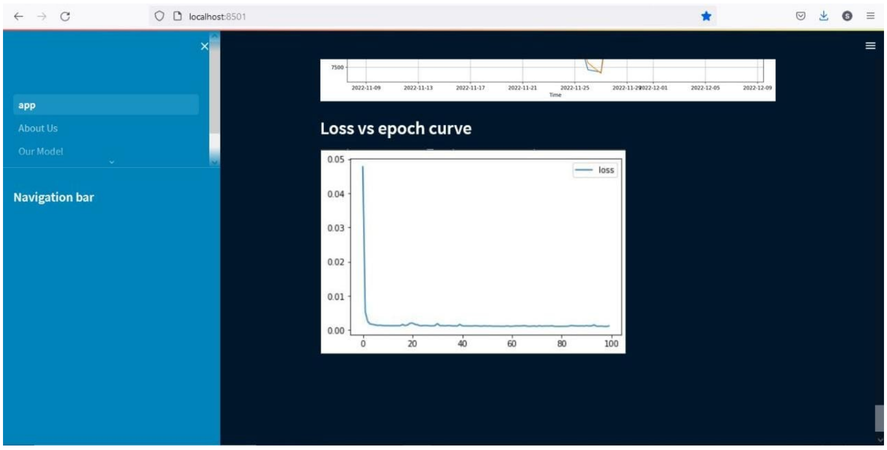

Figure 5.15: Loss vs epoch curve.

Step 8 : Query Params showing and details about company

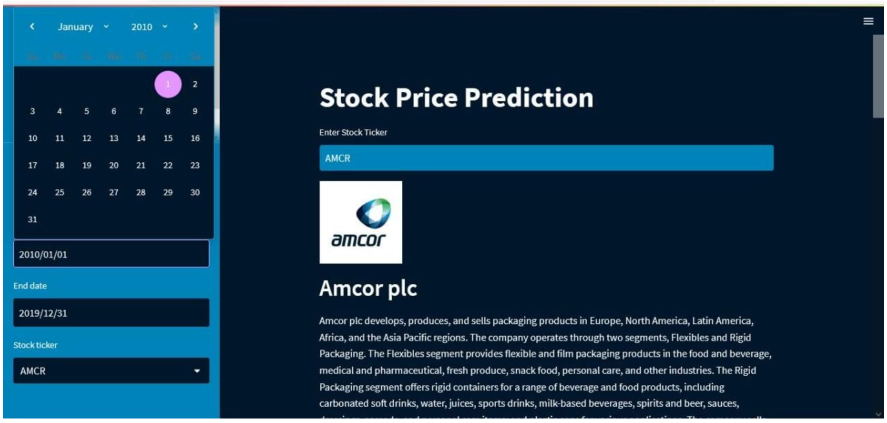

Figure 5.16: Query Params showing and details about company.

Step 8 : Providing Information and queries

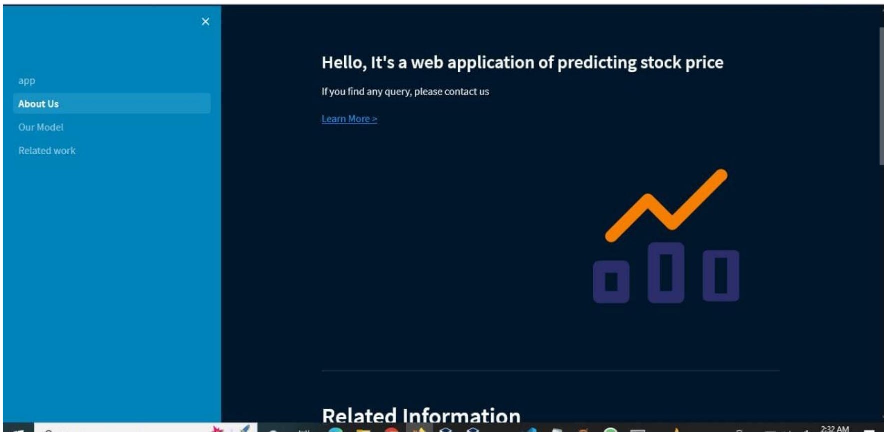

Figure 5.17: Providing Information and queries.

Step 8 : Model overview showing

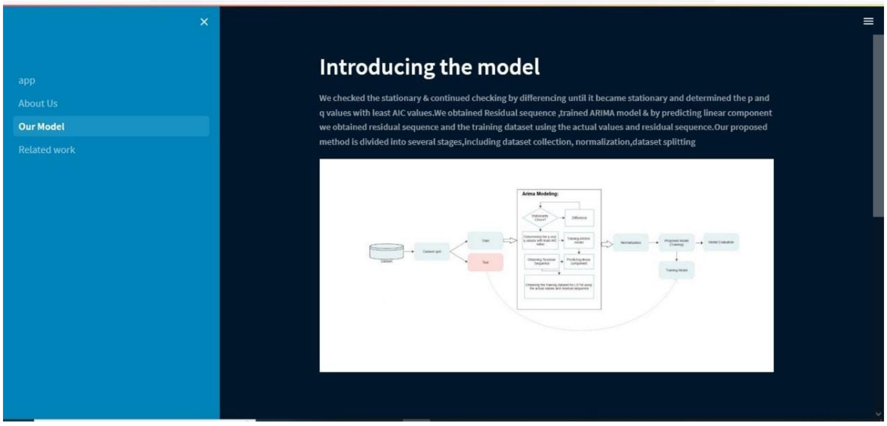

Figure 5.18: Model overview by clicking from the navigation bar.

Step 8 : Related Work

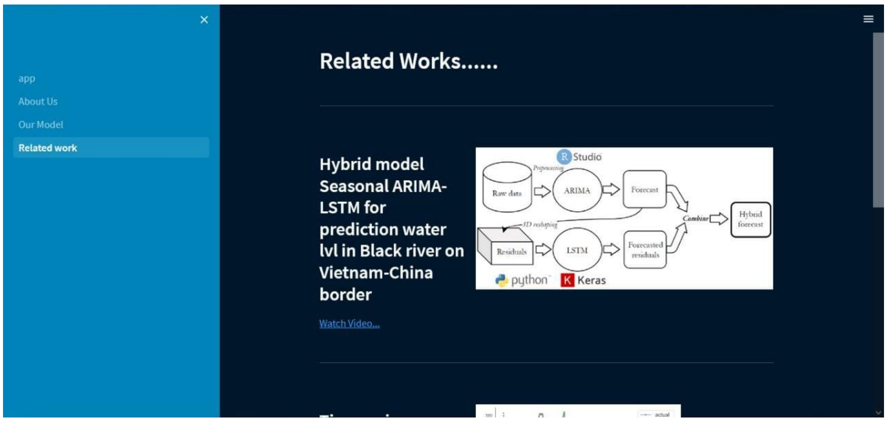

Figure 5.19: Related Work.

Chapter 6

Conclusion

There has been a lot of media coverage and public interest in bitcoin stock as a result of its recent price explosion and crash. With less risk for users and investors, this research attempted to establish a reliable approach for predicting the price of bitcoin (cryptocurrency) stock using a hybrid ARIMA-LSTM deep learning model. In this study, a dense layer is merged with ARIMA and LSTM networks to create a hybrid model. For comparison, other state-of-the-art predictive deep learning networks for prediction, including the conventional ARIMA model, were also created. The performance of long-term forecasts using AR forecasts is worse since they are both a linear function of the coefficients and a linear function of historical data. In comparison to other individual models, the detailed result analysis supports the suggested hybrid model's dependability and stability. The suggested hybrid technique for predicting the price of bitcoin stock would thereby promote consumer growth while also drastically lowering risk for potential investors. Write to Israt Afroz Isha

References

[1] Apoorva Aggarwal, Isha Gupta, Novesh Garg, and Anurag Goel. Deep learning approach to determine the impact of socio economic factors on bitcoin price prediction. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–5, 2019.

[2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[3] Suhwan Ji, Jongmin Kim, and Hyeonseung Im. A comparative study of bitcoin price prediction using deep learning. *Mathematics*, 7(10):898, 2019.

[4] Richard Sylla. The Philadelphia Stock Exchange and the City It Made. By Domenic Vitiello, with George E. Thomas. *The Pennsylvania Magazine of History and Biography*, 60322, 2017.

[5] CoinCodex `https://coincodex.com/crypto/bitcoin/historical-data/`.