|  |
| --- |
| AIDI - 2004 |
| Bitcoin Price Prediction |
| FINAL PROJECT |

|  |
| --- |
| ABHISHEK JOSHI | DHEERAJ SAVANTHY | ESRA ERSOY  4-21-2022 |

**PROJECT REPORT**

**Objective:**

The objective of this project to create a web application that can be used to obtain the forecast of average weighted price of Bitcoin in the currency of US Dollars plotted on an interactive graph. The application is built using the python language, with the support of flask framework and has been deployed using Microsoft Azure Cloud Services.

**Tech Stack:**

1. Python: Is an interpreted, high-level, and general-purpose programming language. It is used to develop the whole system.









1. GitHub: It’s a software development platform to store, track, and collaborate on projects. It is used to manage the version control, share, and check the progress of the project.
2. Flask: Flask is used for developing web applications using python, implemented on Werkzeug and Jinja2. Advantages of using Flask framework are: There is a built-in development server and a fast debugger provided.
3. Plotly: The Plotly Python graphing library is a scientific graphing library. Graphs can be styled with Python and a GUI, and shared with a URL for others to view, collaborate, or save a copy.

**Application Access:**

The application is live at – https://crypforec.azurewebsites.net

**GitHub Link:** https://github.com/dheerajsavanthy/TimeSeriesVisualizer

**Time Series Forecasting:**

Time series forecasting occurs when you make scientific predictions based on historical time stamped data. It involves building models through historical analysis and using them to make observations and drive future strategic decision-making. An important distinction in forecasting is that at the time of the work, the future outcome is completely unavailable and can only be estimated through careful analysis and evidence-based priors

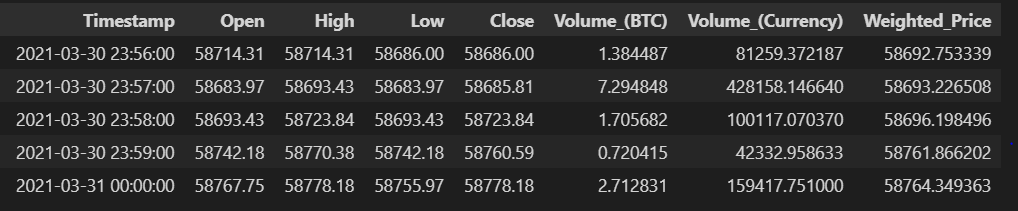
**Step 1: Data Collection**

Bitcoin is the longest running and most well known cryptocurrency, first released as open source in 2009 by the anonymous Satoshi Nakamoto. Bitcoin serves as a decentralized medium of digital exchange, with transactions verified and recorded in a public distributed ledger (the blockchain) without the need for a trusted record keeping authority or central intermediary. Transaction blocks contain a SHA-256 cryptographic hash of previous transaction blocks, and are thus "chained" together, serving as an immutable record of all transactions that have ever occurred. As with any currency/commodity on the market, bitcoin trading and financial instruments soon followed public adoption of bitcoin and continue to grow. Included here is historical bitcoin market data at 1-min intervals for select bitcoin exchanges where trading takes place.

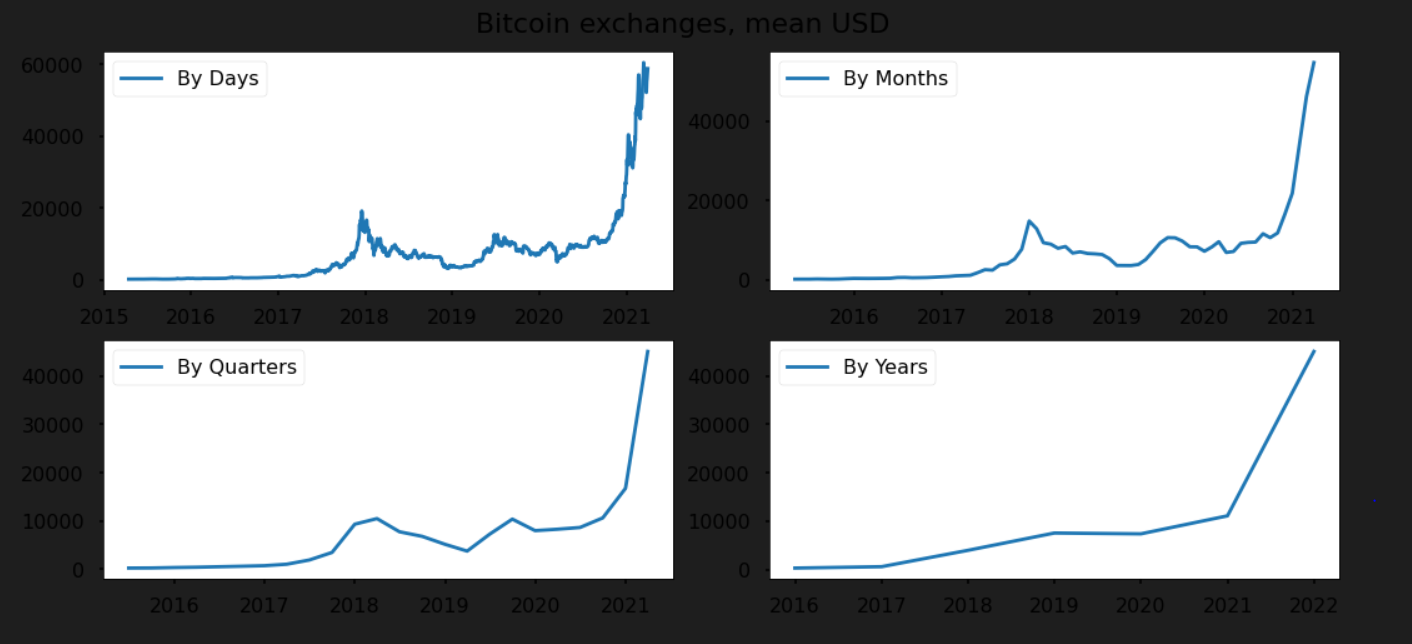
CSV files for select bitcoin exchanges for the time period of Jan 2012 to December March 2021, with minute to minute updates of OHLC (Open, High, Low, Close), Volume in BTC and indicated currency, and weighted bitcoin price. **Timestamps are in Unix time. Timestamps without any trades or activity have their data fields filled with NaNs.** If a timestamp is missing, or if there are jumps, this may be because the exchange (or its API) was down, the exchange (or its API) did not exist, or some other unforeseen technical error in data reporting or gathering.

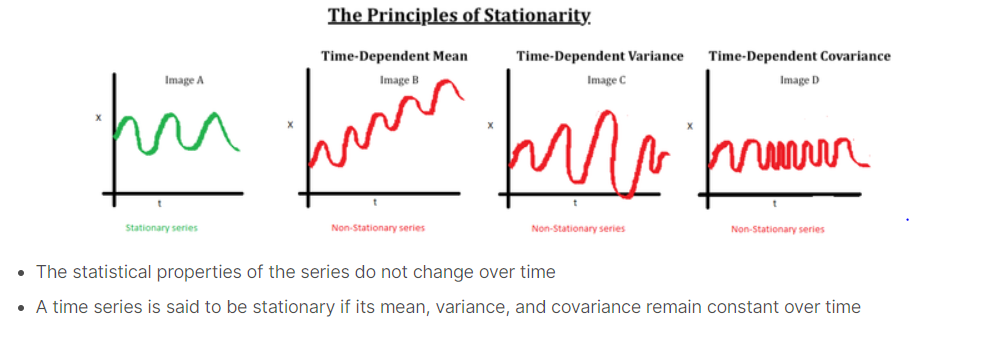
**Step 2: Importing of Data**

Imported the data using pandas and applied a transformation on Timestamp column to convert to date time format.

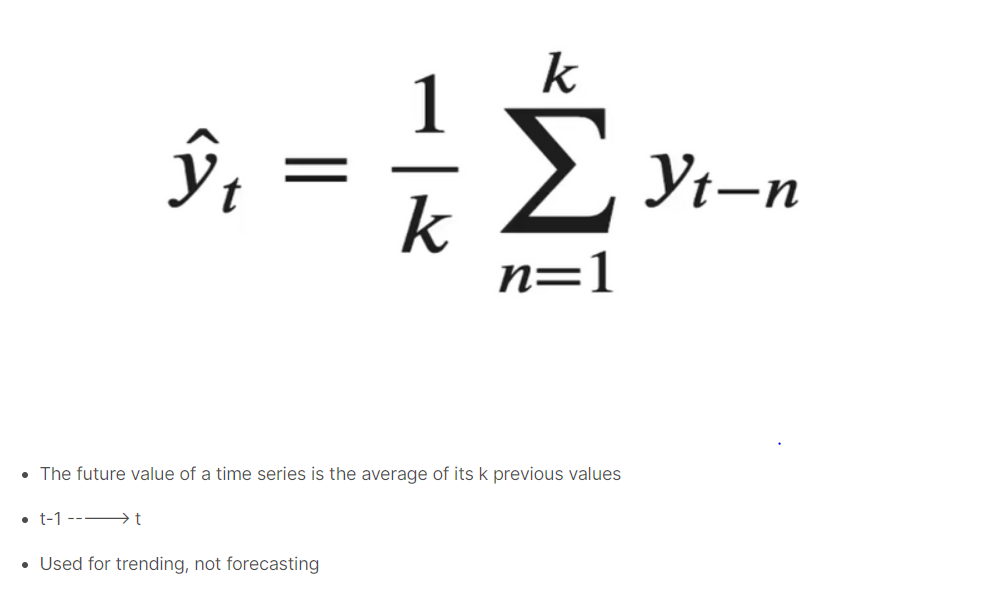


Then plotted the average price over the period recorded in the dataset on basis of days, months, quarters and years.

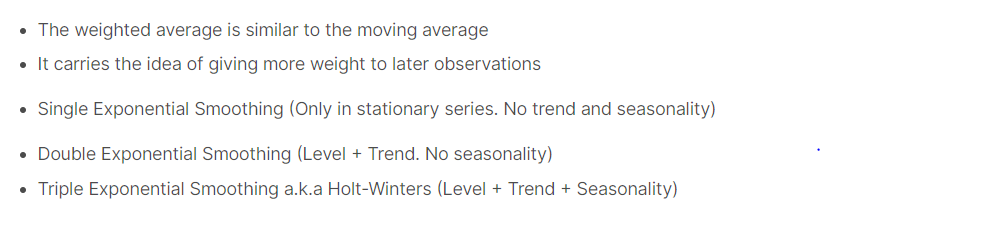


**Stationarity Check:** 

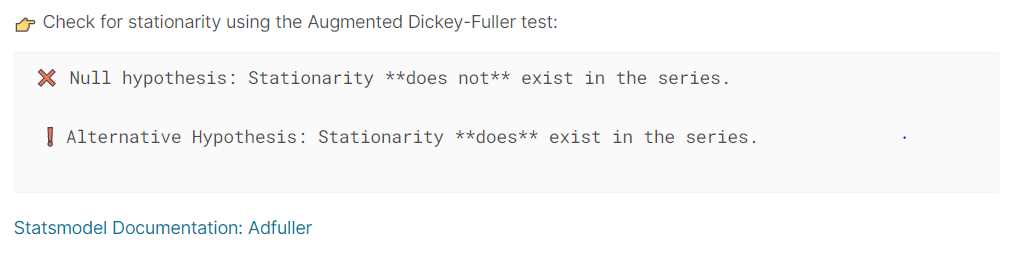
**Moving Average:**

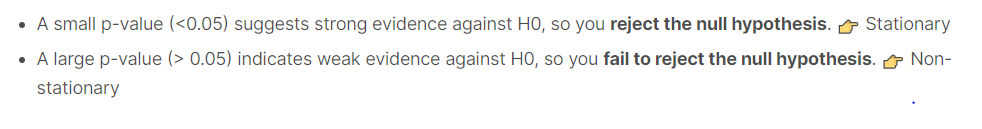


**Weighted Average:**

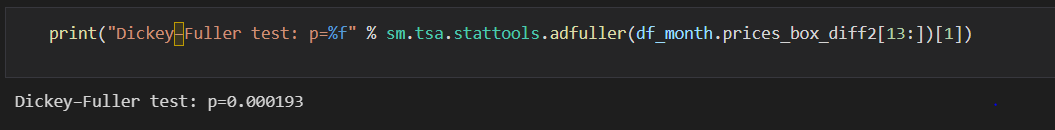


Then we have checked the stationarity using Dickey Fuller Test:





Desired result is found after transforming the data:



**Step 2 - ARIMA:**

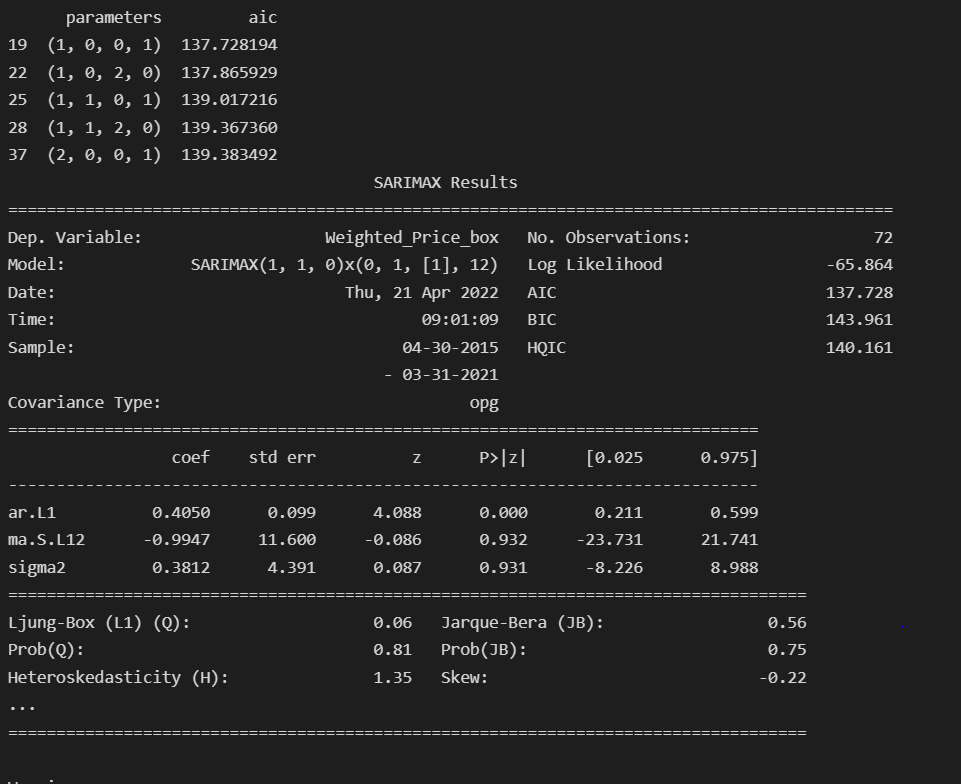
An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses [time series data](https://www.investopedia.com/terms/t/timeseries.asp) to either better understand the data set or to predict future trends.

A statistical model is autoregressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods.

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

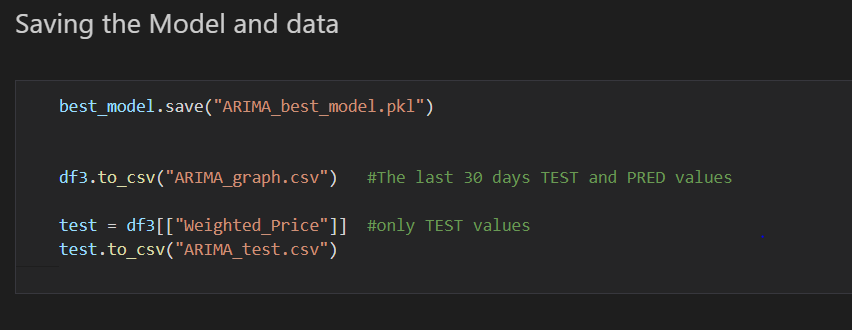
* *p*: the number of lag observations in the model; also known as the lag order.
* *d*: the number of times that the raw observations are differenced; also known as the degree of differencing.
* q: the size of the moving average window; also known as the order of the moving average.

**Model Fitting & Evaluation:**

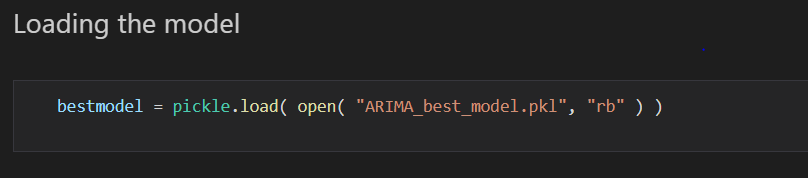


**Saving the Model and Data Files:**

The model and the necessary data files have been saved. The model has been stored in the pickle format in order to use it later during the deployment so that it can be further used for prediction without having to fit the data to the model again. This storage provides convenience in cloud deployments of application.



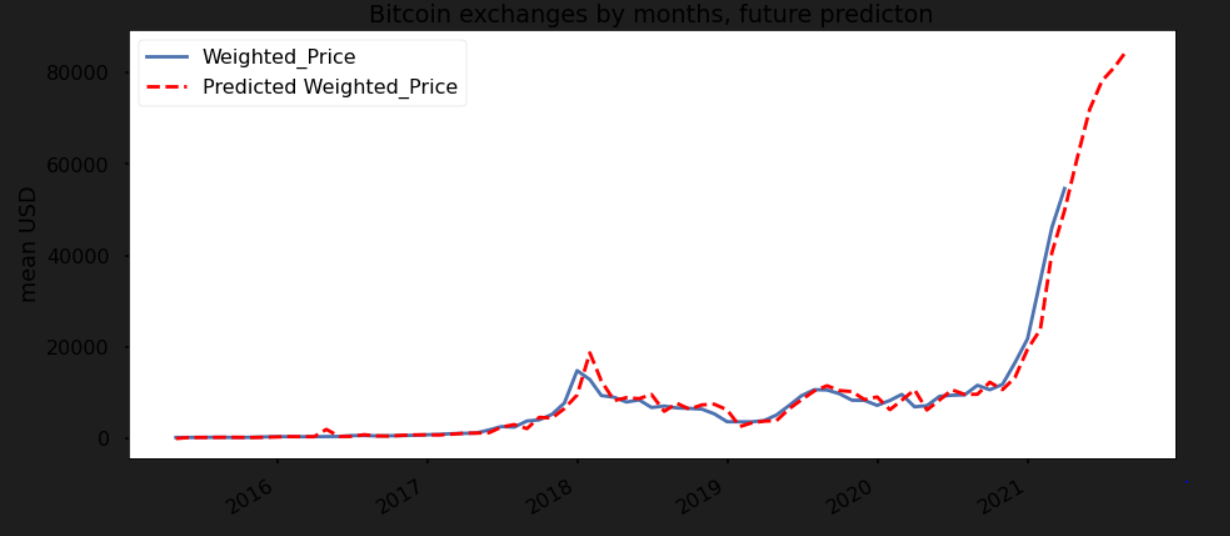
**Loading the Model:**

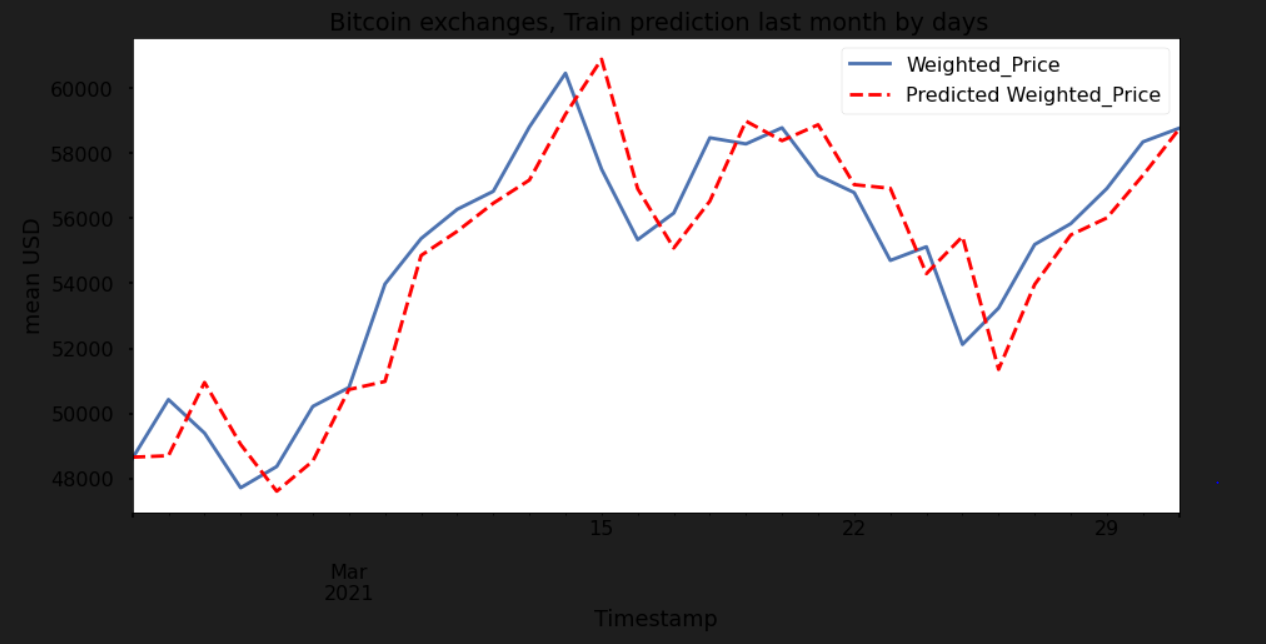


**Step 4 - Prediction:**

Now that we have successfully established all the dependencies and trained the model and saved the file ready to use, we will look at the actual task which has been the primary objective of the project in the first place, which is, prediction or forecasting of the Bitcoin’s weighted average values in comparison with the actual values.

We have plotted the forecast in terms of months and days. First we have plotted the predicted forecast on the monthly basis to a desired period in future. Then we used the last 30 days of the data in the dataset to predict the price for the very next 3 days in the future. Below are the graphs:







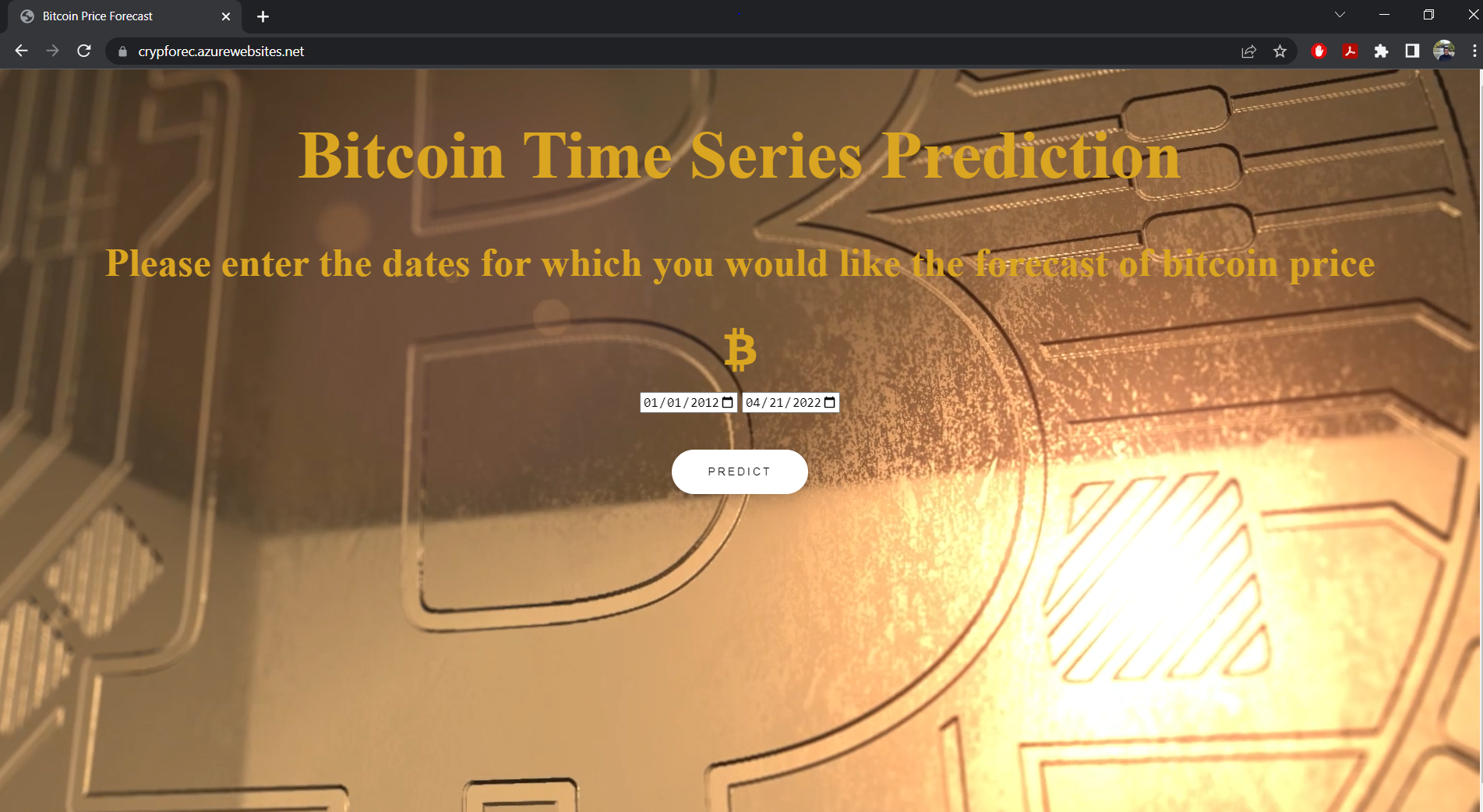
**Step 5 – Web Deployment:**

Using the flask framework, an application has been built using the python code.

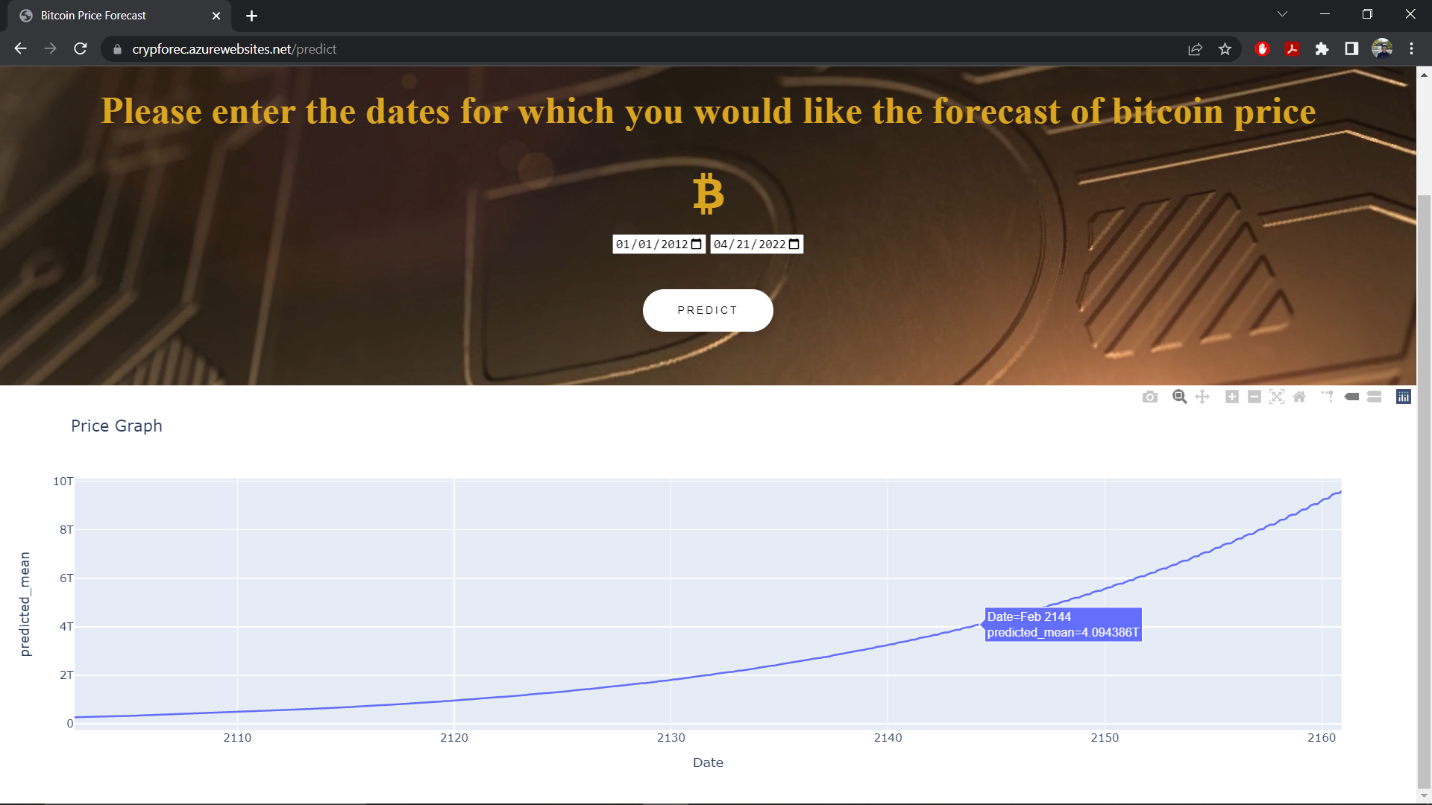
We have designed the web layout of the application using HTML and CSS.

The entire application has been uploaded to the GitHub repository and is used to deploy online through Microsoft Azure Cloud Services.

Here is the front-end application interface:



Predictions are shown over an interactive graph displayed on the screen where user can hover the mouse and check the individual values for desired dates.



**AGILE Methodology:**

Step 1 – Getting all the group members on board

Step 2 – Discuss and choose the concept and idea for the project

Step 3 – Team planning and task division, scheduling, setting targets

Step 4 – Timely Connects to discuss progress and challenges

Step 5 – Using GitHub for version control and submitting work

Step 6 – Share mutual feedback and work on it in timely manner

Step 7 – Repeat the process in multiple stages and keep track of project progress

Step 8 – Complete necessary quality checks and deploy the application

**References:**

1. D. Shah and K. Zhang, "Bayesian regression and Bitcoin", 52nd Annual Allerton Conference on Communication Control and Computing (Allerton), pp. 409-415, 2015.
2. Huisu Jang and Jaewook Lee, "An Empirical Study on Modelling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information", IEEE Early Access Articles, vol. 99, pp. 1-1, 2017.
3. F. Andrade, de Oliveira, L. Enrique ZÃ!rate, M. de Azevedo Reis and C. Neri Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices", IEEE International Conference on Systems Man and Cybernetics, pp. 2151-2155, 2011.
4. M. Daniela and A. BUTOI, "Data mining on Romanian stock market using neural networks for price prediction", informatica Economica, vol. 17, 2013.