

Bitcoin Price Prediction Using ML

SNEGA . S
Department of Computer
Science
and Engineering.
B.S.Abdur Rahman Crescent
Institute
of Science and Technology.
Chennai, India.
snegasen@gmail.com

NIVEDHA . B
Department of Computer
Science
and Engineering.
B.S.Abdur Rahman Crescent
Institute
of Science and Technology.
Chennai, India.
nivebalu26@gmail.com

RAMACHANDRAN. C.A.
Assistant Professor
Department of Computer
Science and Engineering.
B.S. Abdur Rahman Crescent
Institute
of Science and Technology.
Chennai, India.
msg2chandran@gmail.com

Abstract—Presenting Bitcoin Price Prediction Using ML, a machine learning model that is implemented with certain algorithms to deduce the price of bitcoin with the given other details of the factors that influence the bitcoin price directions. Recently, bitcoin has reached a new peak in its popularity due to the law enforcements for tax deductions from the profits gained from bitcoin investments and returns in India. Bitcoin is a digital asset and a payment system that is used as a form of Internet currency. It allows for anonymous payment from one person to another and is therefore a preferred payment method for criminal actions on the Internet. Recently Bitcoin has received a lot of attention from the media and the public due to its recent price hike. The objective of this paper is to determine the predictable price direction of Bitcoin price. Machine learning models can likely give us the insight we need to learn about the future of Cryptocurrency. It will not tell us the future but it might tell us the general trend and direction to expect the prices to move. The proposed model is to build a machine learning model where the data is used to learn about the pattern in the dataset and the machine learning algorithm is used to predict the bitcoin price based on the directly affecting biases. With the effective feedback from the domain experts, we were able to approach the better version of the bitcoin price prediction model.

I. INTRODUCTION

This project deals with the bitcoin price, which fluctuates on a daily basis. Bitcoin, if we need to explain about it in a simple manner we can consider it as a computer file that is basically stored in a digital wallet application or in the computer.

We can send a bitcoin as a whole or a part of it to other people, which gets recorded in a public transaction list that is referred to by the term ‘blockchain’.

Bitcoin does not have a centralized authority nor is it affiliated with any kind of currencies like USD or other kinds of fiat currencies, so based on the trade market and the time considering the commodities of supply and the demand.

II. DOMAIN OVERVIEW

The domain our project belongs to is Machine Learning which is a type of Artificial Intelligence that is a part of data science. Data science can be considered a field of study and development that revolves around information and data. Artificial Intelligence(AI) can be considered a part of data science and Machine Learning(ML), comes under AI. Data science deals with processing and analyzing data with a motive for producing useful stuff that can be used in many ways out of the data.

Artificial Intelligence can be considered as a part of Data Science, that deals with stimulating the machine, i.e. computer, with human knowledge. This helps the machines/ computers to behave in a very human-like manner. It is implemented using many ways of learning like deep learning , machine learning, etc. Here we have used machine learning.

A. Machine Learning

Machine Learning is more about prediction kind rather than the learn and behave kind. What machine learning does is, it makes the system take a good deal of past data and by using various algorithms and techniques it can make predictions corresponding to the data, and it learns how to do it automatically by itself from those algorithms. Simple example of ML is that we can take all of the cat family member's pictures and boost our ML model with algorithms and the names of the members which will be human fed. We train the model with the past images collected, and test it on new images, we can see that the machine has learned to identify the newer images as cats or not.

We people use ML in our daily lives, if we are able to use Google photos the well known photo gallery application from Google Llc gathers photos of

us and it asks us questions along with our various images , like "Is this the same person?", if we feed data on a regular basis for a week, we can see that almost 99% of our data is being classified and organized neatly, for each and every person in the pictures in our gallery.

Nowadays , ML is one of the fields that is soaring in the IT sector, as we can see so many automated services and devices in many other sectors too. Machine Learning is a part of Artificial Intelligence (AI), where the production of next generation humanoid- robots. These robots can be used in various fields from military to medicine. There are even robots that are being made to perform operations and other medical treatments. From all the research and developments that are taking place in a rapid manner, we can clearly see that our future has an irreplaceable place for machine learning and artificial intelligence.

III. Existing systems and Proposed Model

This section is provided with the information about the existing systems available related to bitcoins and along with their disadvantages. It also focuses on the proposed model and its advantages.

A. Existing systems:

As far as bitcoin is concerned there are many research papers and other models that are built to deploy a machine learning model. Bitcoin and other cryptocurrencies are still a vaguely known concept from the public's eye. People get to know all kinds of negative criticisms and rumors faster than the good things. Majority of them know that even though bitcoin has high returns on investing, it has even more risk factors yet what people do is they do not fact check those informations and believe them as it is and they spread the false rumors to other people too. There are so many ridiculous rumors about bitcoin and other cryptocurrencies like they are only used for illegal and illicit purposes and that they are not secure. These kinds of false myths leave a really bad impression on these cryptocurrencies with people. So many researchers even hesitate at times to do projects and research surveys about these topics. As days go by

people get every news in their hand as it is, so their minds are getting cleared about such topics and they really started to focus on what this bitcoin is actually about. Even though there are many articles and projects that are greatly viewed and appreciated related to bitcoin, still the negative view will not fade that easily from the minds of people.

There are many systems that are currently existing for bitcoin related projects and even for the price prediction model, many people have used all kinds of algorithms where they still didn't get to cover some drawbacks.

Some commonly found drawbacks in the previously done projects are:

- Existing models are made for only data analysis and a predicting model was not built.
- The classification model was not discussed and performance metrics like accuracy are not calculated.

B. Proposed System:

Our analysis provides a comprehensive guide to sensitivity analysis of model parameters with regard to performance in the prediction. To propose a machine learning-based model to additionally compare and discuss the performance of various machine learning algorithms for the given dataset we have used python programming language to feature the backend and front end is deployed using flask and the page features are set up with the help of HTML and CSS.

Main advantages that stands out for the proposed model are listed below:

- The model can be used to predict the bitcoin future. Performance metrics like accuracy, recall and precision can be calculated.
- Bitcoin's future may be predicted and the investments can be made wisely.

The architecture that is depicted by Fig 1, clearly shows that the dataset that we use here is first being processed and then only it is sent to the testing

phase for analysis, to check the readiness of the dataset for further phases. Then the model will be trained by splitting the data. Then in the splitted data, it takes a part of data to train itself, a kind of self learning based on the patterns in the data. Then it tests itself by performing predictions by using the data that is left for testing.

Some of the advantages of the Proposed System when compared to the existing system are:

- The model can be used to predict the bitcoin future. Performance metrics like accuracy, recall and precision can be calculated.
- Bitcoin's future may be predicted and the investments can be made wisely.

IV. WORKING OF THE PROPOSED MODEL

Before you begin to dwell into creating a model and executing the project as per the designed plan, it is better to inform the involved members about the working of the model. So further going down, we have discussed the architecture of the model. By referring to the architecture (Ref. Figure 1), the initial step is to extract the data. Here, in our model we have used the dataset that is available in Kaggle, a leading platform for getting our hands on datasets and example coding solutions with an embedded compiler of its own. This data set that we have downloaded from Kaggle consists of 8 columns namely ,The price of bitcoin [USD], The total number of bitcoin confirmed transactions per day, Average transaction fees in USD per bitcoin transaction [USD], Google bitcoin trends search, Gold ounce price [USD], Oil WTI price [USD], M2 money supply in the USA, SP500 close index and The time period is between 12.2014 - 04.2020.

In the architecture chart given below (Fig 1), we can see that the taken data with the bitcoin currency details and with all the other factors that can influence the price of bitcoin, will be processed and this method by which we will remove the redundant and duplicate entries and also the null entries in the dataset to make it clean and available for further processing is called data preprocessing.

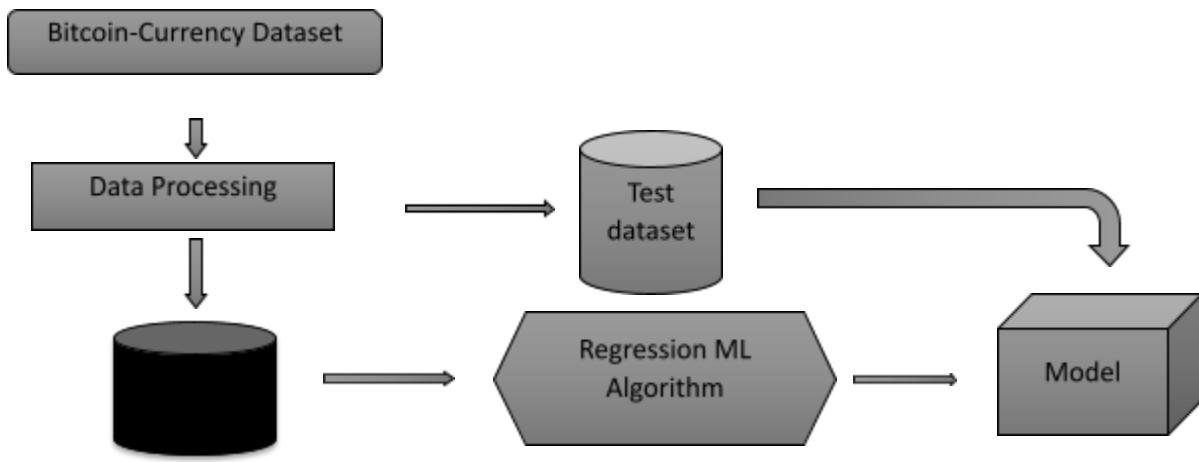


Figure 1: Architecture of the proposed model.

After downloading the dataset, the model is made to preprocess the data and analyze the data until it is set for further processing. When the processing phase of data is considered as completed, we test the dataset again and then if everything is good to go for, the training of the model begins. We begin the training and testing phase. Here we are going to divide the data in a 7:3 ratio, so as to train the model with 70% of the data contained in the dataset and the remaining 30% for the testing phase.

A. Overview of the Model:

We cap the openly present metrics, closed metrics, highly considered metrics, lowly considered metrics and other metrics such as the trading volume and market information for all Bitcoins. We had to check the code with fine eyes to make it compatible with CRAN (CRAN can be considered a home for up to date Coding solutions that is spread all over the world) so that there are significant improvements in how some field changes are done and the data is cleaned. Then the cleaned data will be analyzed again and again so that we can make sure that there are no duplicate data.

B. Algorithms:

1. Linear Regression:

This algorithm is most frequently used where we need to find the linear relationship between two fields generally speaking. The dependency of the dependent variable (i.e) the predictable value, on the independent variable, along the slope with use of a constant value, can be referred to as linear regression.

$$Y_i = \beta_0 + \beta_1 X_i$$

Constant/Intercept
 ↓
 Independent Variable
 ↓
 Y_i = β₀ + β₁ X_i
 ↑
 Dependent Variable
 ↑
 Slope/Coefficient

Figure 2. Linear Regression Formula

As we can see in the image (Fig 2.), we can deduce the dependent variable by adding up the constant value with the product of the coefficient value along the slope and the value of the independent variable.

2. Gradient Boosting:

Boosting algorithms are well known for its use in strengthening the weaker algorithms by bagging it along with algorithms with stronger efficacy. This is one of the machine learning boosting models that is

used to increase the accuracy and efficiency of the predicted output. We know that the boosting technique is an ensemble learning technique that uses a weak algorithm with a stronger one, combined together to produce a model that works more effectively. The goal of this gradient boosting method is to find the function that can give the best approximation of the output variable from the input values. This is done by introducing a loss function that represents the error rate and then focuses on minimizing the loss function as much as possible to improve accuracy. The image given below (Fig 3.), is the output of the gradient boosting with its accuracy based on the error metrics. We can see that the accuracy percentage is 97.49%.

```
MEAN ABSOLUTE ERROR VALUE IS : 399.8999837293194
MEAN SQUARED ERROR VALUE IS : 373063.15121355886
MEDIAN ABSOLUTE ERROR VALUE IS : 205.18985302498913
ACCURACY RESULT OF GradientBoostingRegressor IS : 97.4%
R2_SCORE VALUE IS : 0.9745268635741131
```

Figure 3. Output of Gradient Boosting algorithm

3. Random Forest:

Random forest regression algorithm is one of the well known machine learning algorithms that can produce results with more accuracy rate with varied range of datasets. We know that random forest algorithm takes our dataset and breaks down the larger data clumps into smaller sub-datasets, and performs regression upon all the subsets and after calculating the results for each small data subsets it allows a major vote count to decide on the output value.

As shown in figure (Fig 4.), from the sci-kit learn package, by calling the ensemble module and importing the random forest algorithm library as RandomForestRegressor() as RF and fitting the training data into the random forest regression algorithm as x_train and y_train and then the algorithm gets familiar with the data by training the data patterns and learning them multiple times, within a fraction of minute and then finally tests itself by implementing predictions based on the learnt data patterns by using predict() method on the test data

The random forest algorithm is the best suited algorithm as per the training and the testing data as it has come out with an accuracy rate of 97.86 % .

```
MEAN ABSOLUTE ERROR VALUE IS : 316.6666740863787
MEAN SQUARED ERROR VALUE IS : 314311.31705083395
MEDIAN ABSOLUTE ERROR VALUE IS : 124.50569999999789
ACCURACY RESULT OF RANDOM FOREST REGRESSOR IS : 97.86
R2_SCORE VALUE IS : 0.9785411859984666
```

Figure 4. Random Forest algorithm Output

4. Decision Tree:

Decision tree regression that we have used here is used for both regression and classification purposes. It is formed in a tree-like structure. It is used to divide a dataset into smaller subsets while developing a coordinated decision tree simultaneously in an incremental manner. There are leaf nodes and decision nodes in a decision tree, where a leaf node is used for denoting the numerical data associated with the target and the decision node is used to represent the value for the currently testing attribute or the attribute that was tested at that instance for that particular value. A decision tree can also have two or more branches that allows it to have multiple leaf nodes.

The below image (Fig.5.) that depicts the output ran from executing the Decision tree regressor from the python sklearn library, with the testing data after we have trained the model with the training set of data.

```
MEAN ABSOLUTE ERROR VALUE IS : 414.6785049833887
MEAN SQUARED ERROR VALUE IS : 589688.8246813953
MEDIAN ABSOLUTE ERROR VALUE IS : 154.3699999999999
ACCURACY RESULT OF DECISION TREE REGRESSOR IS : 95.98
R2_SCORE VALUE IS : 0.959740479832693
```

Figure 5. Output of Decision Tree Algorithm

5. LASSO Algorithm:

Least Absolute Shrinkage and Selection Operator, known as LASSO algorithm, is well known for its discrete prediction among the ML algorithms that can be used for both simpler and more interpretable values of output. Generally speaking the regression case is basically a classification problem where the model actually returns the values of output in a continuous manner.

In the below given figure(Fig.6.), the output of the LASSO shows the output of the LASSO algorithm with its accuracy based on the error metrics. We can see that the accuracy percentage is 87.96%.

```
MEAN ABSOLUTE ERROR VALUE IS : 1046.693884664911:  
MEAN SQUARED ERROR VALUE IS : 1765440.8407453676  
MEDIAN ABSOLUTE ERROR VALUE IS : 874.961350390691  
ACCURACY RESULT OF LASSO REGRESSION IS : 88.08071  
R2_SCORE VALUE IS : 0.8796192205745419
```

Figure 6. LASSO algorithm Output.

We have deployed the above explained five algorithms to test which was the most suited algorithm for the taken dataset.

At the end of the execution of the model, we were aware of the fact that even though it was a small data set with around 1500 entries, we got the random forest algorithm as the most accurate one with an accuracy rate of 97.86%.

V. OTHER MODULES

We have completed the modules by dividing it into four modules. Each of the modules have their own essential purposes and they are divided in such a way that there is no glitch in the model. The modules are:

- Data analysis and Preprocessing.
- Data visualization and Correlations.
- Comparing Algorithms for accurate results.
- Deployment using flask.

A. Data analysis and Preprocessing

In this module, we will have to load the data and begin with cleaning the data after analyzing the dataset. Preprocessing is the most basic and most important step in a dataset related operation, since the raw data must be converted into organized data. The raw data that we gain or extract will have all kinds of null values and missing entries in them. It is not possible for us to complete anytime soon if we do this analysis and the preprocessing of the dataset manually. Thus , by using the functions that are inbuilt in the pandas python library, we are going to clean the data and make it ready for actual use in further analysis and visualization. The below shown (Fig. 7.) is an example output that displays the description of the database. It shows the max values, least values, mean, std, average and the overall percentage variances of each and every field in the dataset.

By importing the pandas library we can use all the predefined methods for dataframes, after converting our CSV dataset into a dataframe using the `read_csv()` command from python to read the CSV file directly as a data frame. We can further analyze the data by using basic commands of the pandas library such as `shape()`, `info()`, `unique()`, `describe()` and `columns()`.

```
#To describe the dataframe
df.describe()
```

	BTC_price_USD	n_transactions	fee_USD
count	1002.000000	1002.000000	1002.000000
mean	5352.851926	254421.630739	2.230409
std	3892.224190	88589.954049	5.568724
min	172.000000	54142.000000	0.030000
25%	884.052500	207520.500000	0.230000
50%	5895.335000	267637.500000	0.700000
75%	8171.610000	322157.500000	1.767500
max	18911.790000	490644.000000	54.790000

Figure 7. Describe() method in Data analysis and pre-processing Module.

This module imports packages and reads data. The data that is read will be undergoing the process of pre-processing. The used input is data and the expected output is the data after removing noisy data. An example of another output is given below. Here the duplicated() method is used to remove duplicates after checking them and other preprocessing like changing the column names to our own preferred ones are done.

This figure (Fig 8.), shows the datatype of the values stored in each column of the taken dataset. This is the afterimage that shows the details of the data frame after preprocessing.

```
#Checking datatype and information about dataset
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1002 entries, 0 to 2309
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
0    Date              1002 non-null   object  
1    BTC_price_USD     1002 non-null   float64 
2    n_transactions    1002 non-null   float64 
3    fee_USD           1002 non-null   float64 
4    btc_search_trends 1002 non-null   int64  
5    Gold_price_USD    1002 non-null   float64 
6    SP500_close_index 1002 non-null   float64 
7    Oil_WTI_price_USD 1002 non-null   float64 
8    M2_money_supply_USA 1002 non-null   float64 
dtypes: float64(7), int64(1), object(1)
memory usage: 78.3+ KB
```

Figure 8. Info() method in Data Analysis and pre-processing module.

B. Data visualization and correlations

This module focuses on the analysis of data and visualizing the data, based on the correlations of the fields in the dataset. Visualization is an important phase that actually explains based on which driving factors does this project/ model achieve its goals. The used input is pre-processed data and the expected output is visualized data. Sometimes data does not make sense until it is visible in the form of charts and bars. By visualizing the data we can understand the data better than when we didn't. Visualized data gives us a deep view of the data. The unnoticed insights and patterns in data are made clear while using visualized

data. We have used matplotlib and seaborn packages from python.

In this Module we have used histogram plots, box plots, scatter plots, pie charts and a heatmap. We all know that histogram, unlike a simple bar chart that shows the properties like quantity and such, histogram plot is used to represent how the considered fields are grouped. Putting it in simpler terms, histogram plot is a visual representation of the frequency distribution.

The five variables are the max value, the least value, the median value, the first quartile value and the third quartile value. We can see how the median is positioned between the first and the third quartile values. We have also used scatter plots, one of the commonly seen and most frequently used plots, as we all know this is used to reveal the correlation between two fields. Pie chart is used here to show proportionality among the search trends and heat maps(Fig 9.) are used to define the correlations between the columns of the dataset in a pictorial manner so that it is easier to understand them with one look..Finally, a heatmap was used to display the correlation between all the columns in our dataset.Below (Fig 10.) given is an example image that represents the histogram plot of the number of transactions and the bitcoin price.

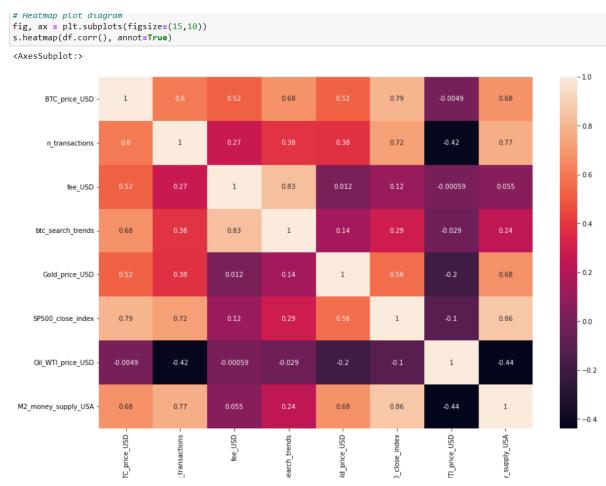


Figure 9. Heat Map to represent the correlations.

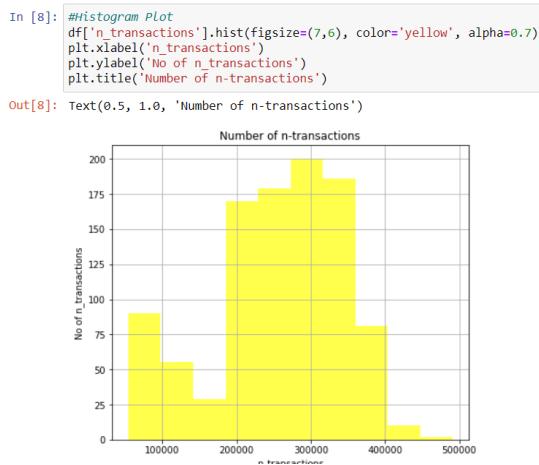


Figure 10. Histogram of the number of transactions for bitcoin price fluctuations.

C. Comparison the algorithms for accurate results :

In this phase , the comparison of the algorithms will take place. Executing several algorithms that are suitable in dealing with training the machine learning model.Training the model and testing the model by using error metrics to conclude with the most accurate algorithm for the used dataset.

D. Deployment Using Flask:

In this module the entire backend model that is built using the python language is deployed in a web page format with the help of flask, a powerful web framework, along with the pickle and joblib packages.

VI. IMPLEMENTATION

A. Comparison of Algorithms:

Executing several algorithms that are suitable in dealing with training the machine learning model.Training the model and testing the model by using error metrics to conclude with the most accurate algorithm for the used dataset.

Even though we have used around 5 different algorithms, while checking for the algorithm that gives the best accuracy it is the random forest

algorithm, followed by the gradient boosting algorithm.

We can refer to the accuracy rate table, which displays the accuracy rate of all the algorithms used (Table 1.).To define the accuracy rate of each algorithm used, some of the mathematical error metrics that are computationally available in the Sci-Kit learn library of python programming language.

a. Mean Absolute Error (MAE):

It is the average value calculated of the absolute errors. Absolute error is the difference between the expected and the predicted values. It matches the unit values with the input unit values.

b. Mean Squared Error (MSE):

This error score is the mean value of the squared differences between the expected and predicted values. MSE gives the output with the squared units of the input given.

c. Median Absolute Error(MedAE):

In this method the loss is considered by calculating the median value that comes from the whole absolute differences between the expected and predicted values. It returns the predicted loss value with the units same as the input.

d. Score Variance:

This is the variance in the output prediction pattern that occurs when the same algorithm or model that we use, is used on different sets of training data.

c. R2 Score:

This is the most predominant metric that determines how much the predicted outcome varies from the actual target that is expected of the model / algorithm which will be deduced from 1 to give the variance coefficient.

The above mentioned five error metrics along

with their definitions are the ones we have used to define the accuracy approximates.

In all of these error metrics R2 Score is the most important metric that we are using to determine the accuracy of each algorithms. 1 is the highest accuracy we could get and 0 is the least accuracy we can get.

SERIAL NUMBER	NAME OF THE ALGORITHM	ACCURACY RATE
1	Linear Regression	88.39%
2	Gradient Boosting	97.69%
3	Random Forest	97.87%
4	Decision Tree	95.98%
5	LASSO Regression	87.96%

Table 1. Accuracy Rate of Algorithms.

B. Flask Deployment :

Flask is an excellent web development framework for REST API creation. It is built on top of Python which makes it powerful to use all the python features. Flask is used for the backend, but it makes use of a templating language called Jinja2 which is used to create HTML, XML or other markup formats that are returned to the user via an HTTP request. Flask provides you with tools, libraries and technologies that allow you to build a web application.

HTML, when expanded it stands as HyperText Markup Language. The basic structure of HTML encloses the head and then the body sections within the HTML tags(<html>...</html>). Here we have used image tags to set backgrounds and then we have used many other commonly used tags. Inline scripting is used to call all the files respective of the model that will be deployed as a website. The styling of this is done using CSS scripting . Here CSS scripting is done

using the external scripting method where we will be storing all formatings and positionings of each and every element in the separate file that will be stored with the extension of css.

The following figure(Fig 11.)is the result of executing the flask deployment model, where the whole ML model is run in the background to provide us with the front end that gets input from the user to display the predicted bitcoin price as output. All we need to do is input the details of all the input boxes and then click on the predict button to get the price of bitcoin corresponding to the given input values.

We have used appropriate images and used a gif type media file as a background to make the UI appear more pleasant on the eyes. The entire webpage is featured with text boxes and a submit button. Once the text boxes are filled out and then if we hit on the submit button we can get the bitcoin price that the model has predicted at the bottom of the page.

The output from flask deployment will appear in the form of a warning that says that there is a need for redirection, which we must view in a new page . We can just go through the warning until we come across the URL that is being displayed with the selected host and port. When clicked on that URL it takes us to the web page in a new window, where we can view the Bitcoin price prediction website.



Figure 11. Web Page Design.

C. Back-end Implementation:

The whole implementation of the back end of the model is done by using the jupyter notebook that is a part of anaconda software. As said in the descriptions of each module, compromises of precise content that are only relative to the model.

As far as the backend is concerned we have made use of libraries like pandas, numpy, seaborn and scikit learn. All the modules have been explained in the previous chapters in a detailed manner. The whole implementation takes place by using pickle, a powerful package of python that is used to convert hierarchical object files into byte streams.

The joblib package in python is used to provide a simple pipeline network kind of connection which makes a lightweight connection between the python machine learning model and the HTML webpage that is boosted with CSS styling.

D. Front-end Implementation:

The front end here is created using the notepad application, where the HTML file is saved with an extension of .html and we have used external CSS to gear up the styling features of the website even more. The use of an external CSS makes the website appear even more dynamic and expressive. The website goes with a blue-green theme, where a big bitcoin symbol appears in the middle of the page. The background of the website is featured with a gif file that makes the website more awake.

After entering the values that are needed all the values are obtained from html and are applied in the ML model for the algorithms to work on and then the algorithm that has the most accuracy will be chosen to display the result on the web page. This is all done by using flask where we have deployed pickle and joblib.

VII. LITERATURE SURVEY

In this section, the brief description of the research and literature articles that were of utmost help for the project completion are included. Every paper had its own resources and ideas that gave the right thing for us to proceed with the project.

1. Yang Li, Zibin Zheng and Hong-Ning Dai (2019) [1] Investigates the Bitcoin price fluctuation prediction problem, which can be described as whether Bitcoin price keeps or reverses after a large fluctuation by using three kinds of features including basic features, traditional technical trading indicators, and features generated by a Denoising autoencoder. The evaluation of these features is done by using an Attentive LSTM network and an Embedding Network (ALEN). Also investigates the impact of parameters on the Bitcoin price fluctuation prediction problem, which can be further used in a real trading environment by investors.

2. Lekkala Sreekanth Reddy, Dr.P. Sriramya (2020) [2] Nowadays many papers have accurate prices but some others don't, but the time complexity is higher in those predictions, so to reduce the time complexity here in this paper they used an algorithm linked to artificial intelligence named LASSO (least absolute shrinkage selection operator).

3. Shivam Pandey, Mr.Anil Chavan, [3] (2021) In the first phase, this paper aims to understand and identify daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. For the second phase of our survey, using the available information, it focuses on prediction of the sign of the daily price change. In this project they have attempted to predict the prices of Bitcoins using two deep learning methodologies.

4. Muhammad Ali Nasir1, Toan Luu Duc Huynh (2021) [4] Considering all the debate on the role of cryptocurrencies in the current economy as well as their dynamics and forecasting, this brief study analyzes the predictability of Bitcoin volume and returns using Google search values.

5. David Garcia and Frank Schweitzer (2015) [5] This paper illustrates the increases in opinion

polarization and exchange volume precede rising Bitcoin prices, and that emotional valence precedes opinion polarization and rising exchange volumes. Applying these insights to design algorithmic trading strategies for Bitcoin, reached very high profits in less than a year. Verified using robust statistical methods that take into account risk and trading costs, confirming the longstanding hypothesis that trading-based social media sentiment has the potential to yield positive returns on investment.

6. Ruchi Mital ,Shefali Arora and M.P.S. Bhatia (2018) [6] This price prediction helps the number of users who are using cryptocurrencies for multiple types of transactions. They used deep learning models like LSTM to analyze various cryptocurrencies. The main drawback is that they might not have enough information for long-term analysis.

7. Alireza Ashayer (2019) [7] Two performance metrics for each model were measured: mean squared error and the accuracy of the model in predicting the direction of price movement for the next day. To enable a meaningful comparison between the mean squared error value among different cryptocurrencies, daily closing prices were normalized before being used by the models.

8. Neha Mangla, Akshay Bhat, Ganesh Avabratha, Narayana Bhat. (2019) [8] In this paper they used the previous timestamps to develop a predictive model, by using the ARIMA and RNN methods, yet this kind of predictive model does not meet the requirement of fulfilling the long term efficiency condition.

9. Edwin Sin , Lipo Wang. (2017) [9] They concluded that the ensemble method, GASEN, is able to perform well for the classification task with utmost consistency of accuracy of around 58% to 63%. With a simple trading strategy, the ensemble was able to obtain promising results in making profit of close to 85% in return.

10. Franco Valencio, Alfonso Gomez Espinosa, Benjamin Valdes Aguirre (2019) [10] Comparing the utilization of (NN)neural networks, (SVM)support vector machines and (RF) random forest were using elements from Twitter and market data as input data. Concludes that it is possible to predict cryptocurrency

markets using machine learning and sentiment analysis, while Twitter data by itself could be used to predict specific cryptocurrencies and also that NN outperform the other present models.

11. Mohammed Khalid Salman and Abdullahi Abdu Ibrahim [11] , a paper based on price prediction of different cryptocurrencies by conducting technical trade indicators and machine learning processes, using price mining of data and building a neural network.

12. Forecasting and trading cryptocurrencies with machine learning under changing market conditions [12] ,by Helder Sebastiao and Pedro Godinho considers several classes of machine learning models, namely, linear models, random forests (RFs), and support vector machines (SVMs).

13. Article that talks about how to compare the machine learning models and algorithms [13], by Samadhritha Gosh, which gave in depth knowledge and information related to comparison of various machine learning algorithms and the methods of error reduction while prediction. They also provide profound knowledge about the basis of which we will have to decide the most suitable algorithm for a particular set of data.

14. Azim Muhammad Fahmi, Noor Azha Samsudin, Aida Mustapha (2017), [14] conducted an analysis based on regression which was for bitcoin's price prediction, also this whole prediction process was done with the help of the cloud computing platform of the Microsoft organisation , Microsoft Azure..They used the very well known algorithms and methods like LR, BLR, RNN and BTDR in the microsoft azure platform.

15. The paper published by Ladislav Kristoufek(2014), [15] discusses the main driving factors of bitcoin price and deals with the evidence from the coherence analysis of wavelets. By using the wavelets framework, it examines all the possible drivers of bitcoin price fluctuations considering both time and frequency. It concludes with the utmost details that affirms the short term and the long term factors.

As mentioned previously, all the above mentioned works from all the different authors were of utmost use for the successful completion of our project.

VIII. CONCLUSION

The analytical process starts from data cleaning and processing of missing values, then exploratory analysis and finally model building and evaluation. The best accuracy on a public test set with a higher accuracy score will be found out. This application can be of help to find the bitcoin market price.

The model can be further connected to an AI by optimizing it to an Artificial Intelligence environment, we can also automate the model by creating a web application or a desktop application. The thing that we have in our minds for the future enhancement is another kind even though it is related to AI.

IX. FUTURE ENHANCEMENT

Having created a model based on the dataset that is already recorded in the past, we wish to enhance the system in the upcoming future by collecting the data directly from the search engines to predict the bitcoin price of that day. At the same time studying the patterns of the data that is required to predict the bitcoin price, on a day to day basis so that one day in future the model will be able to predict bitcoin prices by predicting all the data beforehand. Thus, predicting the tomorrow, by today.

ACKNOWLEDGMENT

We cannot express enough thanks to our team supervisor Dr. A. Ramachandran, for his guidance and every valuable minute that he spent to mold us for the successful completion of our project and paper works. We also thank our university staff and faculty members of the CSE department who continued their support, even when we cannot meet through, they held their parts neatly on time, thus have contributed a whole lot for this project. Last but not least we thank our parents and friends out there who are always at our

backs smiling and cheering for us so that we can move forward in the directions we want without much fear.

We also thank the authors of all the mentioned literatures, as they did contribute much to add knowledge of bitcoin and the error metrics for us.

REFERENCES

1. BitConduite[1](exploratory visual analysis of entity activity on the Bitcoin Network)by Christoph Kinkeldey,Jean-Daniel Fekete,Tanja Blascheck,Petra Isenberg .[Issue10.1109/MCG.2021.3070303]
2. Paper on enhancing the price of bitcoin's fluctuation prediction using attentive LSTM and embedding network by Yang Li, Zibin Zheng and Hong-Ning Dai (2020) [2] published by applied journal as an open access article.
3. Lekkala Sreekanth Reddy, Dr.P. Sriramya[3] (2020) [2] Published a research paper on Bitcoin price prediction by making use of Machine learning algorithms in the International journal of scientific and technology research(IJSTR).ISSN 2277-8616.
4. Paper about the prediction of bitcoin price using ML algorithms by Shivam Pandey, Mr.Anil Chavan, [4] (2021), published by IRJET, in volume8-issue5.
5. The forecasting of crypto currency returns and the volume using search engines paper that was published by Muhammad Ali Nasir1, Toan Luu Duc Huynh Sang Phu Nguyen and Duy Duong(2019) [5], in the springer journals.
6. David Garcia's and Frank Schweitzer's paper on "Social signals and algorithmic trading of bitcoin", [6] during 2015 6.
7. Research paper published, 2018, on automated cryptocurrency price prediction using ML[7] , by Ruchi Mital ,Shefali Arora and M.P.S. Bhatia (2018) .
8. Alireza Ashayer's , "Modeling and prediction of cryptocurrency prices using machine learning techniques,"[8] June (2019) .

9. Neha Mangla, Akshay Bhat, Ganesh Avabratha, Narayana Bhat. Cryptocurrencies Market price prediction using Data Science process [9] 2018;ISSN 2321-9653..
10. Edwin Sin , Lipo Wang.Bitcoin price prediction using ensembles of neural networks[10] .2017.publised 978-1-5386-2165-3/17 by IEEE.
- 11 Franco Valencio, Alfonso Gomez Espinosa, Benjamin Valdes Aguirre [11] paper on price prediction movements of cryptocurrencies using sentiment analysis and ML,published on june 2019, by entropy e21060589.
12. “Price prediction of different crypto currencies using technical trade indicators and ML”, by Mohammed Khalid Salman and Abdullahi Abdu Ibrahim [12], nov 2020.,by IOP conference series.032007.
13. Forecasting and trading cryptocurrencies with machine learning under changing market conditions [13] ,by Helder Sebastiao and Pedro Godinho published by Ideas. s40854-020-00217-x.
14. How to compare the machine learning models and algorithms [14], by Samadhritha Gosh,published in neptune blogs in 2022.
15. Azim Muhammad Fahmi, Noor Azha Samsudin, Aida Mustapha ‘s paper on ,“Regression based analysis of bitcoin prediction”, [15] ,published in the year 2017.
16. The paper published by Ladislav Kristoufek, [16] on,”What are the main drivers for bitcoin price”, during 2015,published by research gate.280044258-rs.