

# BITCOIN PRICE ANALYZE AND PREDICTION USNG DATA SCIENCE PROCESS

B. Bairavan<sup>1</sup>, P. Gopalakannan<sup>2</sup>, Jibin George<sup>3</sup>  
Students, Dept. of Computer Science, Paavai Engineering College, India<sup>123</sup>

Mr. N. M. K. Ramalingam Sakthivelan<sup>4</sup>  
Asst. Professor, Dept. of Computer Science, Paavai Engineering College, India<sup>4</sup>

**ABSTRACT**— *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 made to learn about the pattern in the dataset and the machine learning algorithm is used to predict the bitcoin price.*

**Keywords**— **Machine learning, Bitcoin, Data science**

## 1, INTRODUCTION

AI is the ability to predict the future based on previous data. Computerize reasoning (AI) is a kind of computerized reasoning (AI) that allows PCs to learn without being explicitly updated. The core of artificial intelligence (AI) is the development of computer programmes that can adapt to new knowledge, as well as the nuts and bolts of machine learning. Preparation and forecasting need the use of certain calculations. As a result of using the preparation data, a computation may predict things about the results of another, examining the question of whether ornot the value retains or inverts. In this work, we focus on the Bitcoin value variance expectation issue. At the cost valuation prediction, this work introduces three different types of high points, including important components, typical specialized trading pointers, and elements developed by a Denoising auto encoder. An Attentive LSTM organization and an Embedding Network are used to evaluate these aspects (ALLEN). Specifically, a mindful LSTM organization can catch the time reliance portrayal of Bitcoin cost and an inserting organization can catch the concealed portrayals from related digital currencies. Test results show that ALLEN accomplishes predominant cutting edge execution among all baselines. Besides, we examine the effect of boundaries on the Bitcoin value change expectation issue, which can be additionally utilized in a genuine exchanging climate by financial backers.



## OBJECTIVE

The goal is to develop a machine learning model for Bitcoin Prediction, to potentially replace the updatable supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithms.

### 3. EXISTING SYSTEM

Bitcoin is a computerized resource and an installment framework that is utilized as a type of Internet cash. It takes into account mysterious installment starting with one individual then onto the next and is along these lines a favored installment technique for criminal activities on the Internet. As of late Bitcoin has gotten a great deal of consideration from the media and the general population because of its new value. The target of this paper is to decide the anticipated value heading of Bitcoin cost. AI models can almost certainly give us the knowledge we really want to find out with regards to the fate of Cryptocurrency. It won't let us know the future however it may let us know the overall pattern and course to anticipate that the costs should move. The model is to fabricate an AI model where the information is utilized to find out with regards to the example in the dataset and the AI calculation is utilized to foresee the bitcoin cost.

#### Drawbacks

- They had made only data analysis and they did not build a predicting model.
- The classification model was not discussed and performance metrics like accuracy are not calculated.

### 4. PROPOSED SYSTEM

Bitcoin's recent price increase has drawn a lot of attention from the media at the general public. It is the goal of this research to discover the Bitcoin price's predicted direction. The full dataset will be analysed using the supervised machine learning approach (SMLT) to identify variables, perform univariate, bivariate, and multivariate analysis, and examine missing value treatments, data validation, data cleaning, and data visualization. Based on our findings, model parameter sensitivity may be effectively assessed in terms of prediction accuracy. This paper proposes a machine learning-based approach and compares various machine learning methods against the provided dataset. The model can be used to predict the bitcoin future. Performance metrics like accuracy, recall and precision can be calculated. Bitcoin future may be predicted and the investments can be made wisely.



## ADVANTAGES

- We present and compare the results obtained with the three forecasting algorithms and the baseline method.
- We predict the price of the currencies at day for all included between Jan, 1st 2016 and Apr 24th, 2018.
- The analysis considers all currencies whose age is larger than 50 days since their first appearance and whose volume is larger than \$100000.
- To discount for the effect of the overall market movement (i.e., market growth, for most of the considered period), we consider cryptocurrencies prices expressed in Bitcoin.

## 5, OVERVIEW OF SYSTEM

### **Problem definition:**

Bitcoin is an internet-based medium of exchange in the form of digital assets which uses cryptographic functions to conduct financial transactions. Bitcoin leverage block chain technology to gain decentralization, transparency, and immutability. The main scope of the project is to finding the accuracy, Minimize the error rate and getting result from the flask framework deployment.

### **OVERVIEW OF PROPOSED SYSTEM:**

All historic open, high, low, close, trading volume and market cap info for all Bitcoin. I've had to go over the code with a fine tooth comb to get it compatible with CRAN so there have been significant enhancements to how some of the field conversions have been undertaken and the data being cleaned. This should eliminate a few issues around number formatting or unexpected handling of scientific notations. Project Goals Exploration data analysis of variable identification Loading the given dataset Import required libraries packages Analyze the general properties Find duplicate and missing values Checking unique and count values Uni-variate data analysis Rename, add data and drop the data To specify data type Exploration data analysis of bi-variate and multi-variate Plot diagram of pairplot, heatmap, bar chart and Histogram Method of Outlier detection with feature engineering Pre-processing the given dataset Splitting the test and training dataset



Comparing the Decision tree and Logistic regression model and random forest. Comparing algorithm to predict the result Based on the best accuracy

Preparing the Dataset :

This dataset was created in order to build models for bitcoin price prediction. It contains

- 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
- The time period is between 12.2014 - 04.2022

## **6, ALGORITHM EXPLANATION**

The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers uses this data to fine-tune the model hyper parameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model.

### **COMPARING ALGORITHM WITH PREDICTION IN THE FORM OF BEST ACCURACY RESULT**

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data

from different perspectives. The same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize. A way to do this is to use different visualization methods to show the average accuracy, variance and other properties of the distribution of model accuracies. In the next section you will discover exactly how you can do that in Python with scikit-learn. The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data and it can achieve this by forcing each algorithm to be evaluated on a consistent test harness.

## **REGRESSION**

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables). Regression helps investment and financial managers to value assets and understand the relationships between variables, such as commodity prices and the stocks of businesses dealing in those commodities. The two basic types of regression are simple linear regression and multiple linear regression, although there are non-linear regression methods for more complicated data and analysis. Simple linear regression uses one independent variable to explain or predict the outcome of the dependent variable Y, while multiple linear regression uses two or more independent variables to predict the outcome. Regression can help finance and investment professionals as well as professionals in other businesses. Regression can also help predict sales for a company based on weather, previous sales, GDP growth, or other types of conditions. The capital asset pricing model (CAPM) is an often-used regression model in finance for pricing assets and discovering costs of capital.

## **FEATURE EXTRACTION**

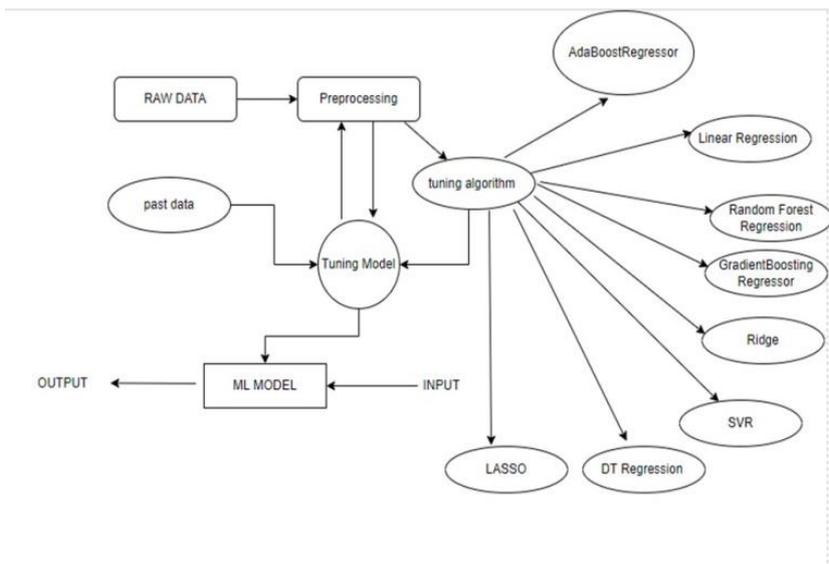
### **Data flow diagram**

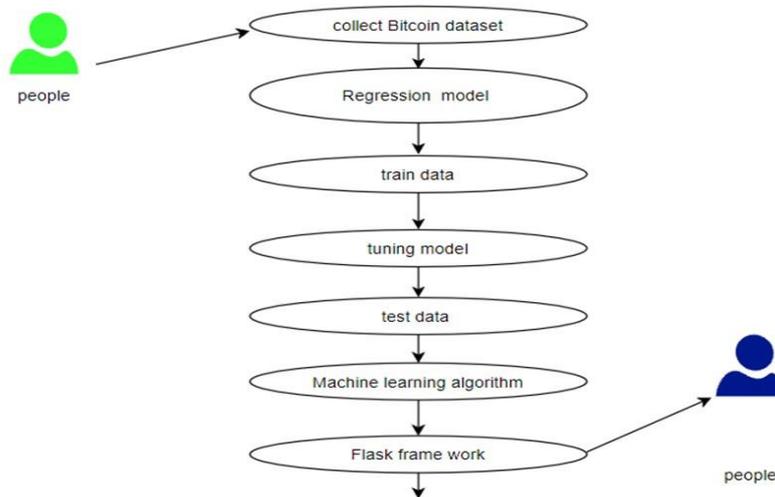
A two-dimensional diagram explains how data is processed and transferred in a system. The graphical depiction identifies each source of data and how it interacts with other data sources to reach a common output. Individuals seeking to draft a data flow diagram must identify external inputs and outputs, determine how the inputs and outputs relate to each other, and explain with graphics how these connections relate and what they result in. This type of diagram helps business development and design teams visualize how data is processed and identify or improve certain aspects.

Symbol	Description
	An <b>entity</b> . A source of data or a destination for data.
	A <b>process</b> or task that is performed by the system.
	A <b>data store</b> , a place where data is held between processes.
	A <b>data flow</b> .

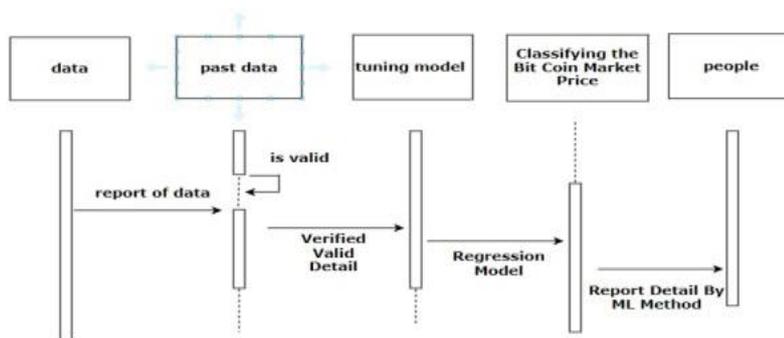
**LEVEL1**

An entity relationship diagram (ERD), also known as an entity relationship model, is a graphical representation of an information system that depicts the relationships among people, objects, places, concepts or events within that system. An ERD is a data modeling technique that can help define business processes and be used as the foundation for a relational database.





**LEVEL2;**



Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e., models with fewer parameters) The word “LASSO” stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. It is a statistical formula for the regularisation of data models and feature selection. Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicol linearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination Lasso Regression uses

L1 regularization technique (will be discussed later in this article). It is used when we have more number of features because it automatically performs feature selection. The L1 regularization performed by Lasso, cause regression coefficient of the less contributing variable to shrink to zero or near zero. Lasso regression uses shrinkage, where the data values are shrunk towards a central point such as the mean value.

The Lasso penalty shrinks or reduces the coefficient value towards zero. The less contributing variable is therefore allowed to have a zero or near-zero coefficient.

	Date	BTC price [USD]	n- transactions	fee [USD]	btc search trends	Gold price[USD]	SP500 close index	Oil WTI price[USD]	M2(Not seasonally adjusted)[1e+09 USD]
0	2014-01-02	749.27	54770.0	0.18	8	1225.0	1831.98	95.14	11089.3
1	2014-01-03	781.23	60980.0	NaN	8	1238.4	1831.37	93.66	11089.3
2	2014-01-04	807.39	52052.0	NaN	8	NaN	NaN	NaN	11089.3
3	2014-01-05	828.74	58662.0	0.38	8	NaN	NaN	NaN	11089.3
4	2014-01-06	904.98	67358.0	NaN	8	1237.8	1826.77	93.12	11089.3

Module of project;

- Installing the Python anaconda platform.
- Loading the dataset.
- Summarizing the dataset.
- Visualizing the dataset.
- Evaluating some algorithms

**Compare with Algorithm;**

In the example below 6 different algorithms are compared:

- Logistic Regression
- Linear Regression
- Decision Tree Regression
- Random Forest Regression
- Support Vector Regression
- Lasso Regression

## 7. RESULT AND DISCUSSION

### Datasets

Machine learning needs data gathering have lot of past data's. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can't be used directly. It's used to pre-process then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

### CONCLUSION

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. The analysis of dataset by supervised machine learning technique (SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset. 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 and additionally, to compare and discuss the performance of various machine learning algorithms for the given dataset. The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This application can help to find the bitcoin Market Price.

### REFERENCES

- [1]. Gulani Senthuran, Malka N. Halgamuge, " Prediction of Crptocurrency Market Price Using Deep Learning and Blockchain Information: Bitcoin and Ethereum".
- [2]. Yang Li 1,2 , Zibin Zheng 1,2,\* and Hong-Ning Dai, " Enhancing Bitcoin Price Fluctuation Prediction Using Attentive LSTM and Embedding Network".
- [3]. Lekkala Sreekanth Reddy, Dr.P. Sriramy, " A Research On Bitcoin Price Prediction Using Machine Learning Algorithms".
- [4]. Mr. Shivam Pandey<sup>1</sup>, Mr.Anil Chavan<sup>2</sup>, Miss. Dhanashree Paraskar<sup>3</sup>, Prof. Sareen Deore, " Bitcoin Price Prediction using Machine Learning".
- [5]. Muhammad Ali Nasir<sup>1\*</sup> , Toan Luu Duc Huynh<sup>2</sup> , Sang Phu Nguyen<sup>3</sup> and Duy Duong<sup>3</sup>, " Forecasting cryptocurrency returns and volume using search engines"



- [6] —Bitcoin Volatility Forecasting with a Glimpse into Buy and Sell Orders| Tian Guo ; Albert Bifet ; Nino AntulovFantulin ; IEEE 2018
- [7] F. Andrade de Oliveira, L. Enrique ZÃ¡rate and M. de Azevedo Reis; C. NeriNobre, —The use of artificial neural networks in the analysis and prediction of stock prices,| in IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2151-2155.
- [8] M. Daniela and A. BUTOI, —Data mining on Romanian stock market using neural networks for price prediction|. informaticaEconomica, 17,2013.
- [9] —Bitcoin Cost Prediction using Deep Neural Network Technique|Kalpanasonika ,Sayasri S , Vinothini , SugaPriya , IEEE 2018
- [10] —An improved K-Means clustering algorithm| ;Juntao Wang, Xiaolong Su, IEEE.