

# Crypto-Currency price prediction using Decision Tree and Regression techniques

Karunya Rathan<sup>1</sup>, Somarouthu Venkat Sai<sup>2</sup>, Tubati Sai Manikanta<sup>3</sup>

<sup>1</sup> Dept. Of Information Technology, Sathyabama Institute Of Science And Technology, Chennai, TamilNadu

<sup>2</sup> Dept. Of CSE(3511537), Sathyabama Institute Of Science And Technology, Chennai, TamilNadu.

<sup>3</sup> Dept. of CSE(3511593), Sathyabama Institute Of Science And Technology, Chennai, TamilNadu.

**Abstract** - Crypto-currency such as Bitcoin is more popular these days among investors. In the proposed work, it is studied to forecast the Bitcoin price precisely considering different parameters that influence the Bitcoin price. This study first handles, it is identified the price trend on day by day changes in the Bitcoin price while it gives knowledge about Bitcoin price trends. The dataset till current date is taken with open, high, low and close price details of Bitcoin value. Exploiting the dataset machine learning module is introduced for prediction of price values. The aim of this work is to derive the accuracy of Bitcoin prediction using different machine learning algorithm and compare their accuracy. Experiment results are compared for decision tree and regression model.

**Key Words:** Bitcoin, Crypto-currency, Decision tree, Regression model, machine learning.

## 1. INTRODUCTION

Bitcoin is a crypto-currency which is utilized worldwide for computerized investment. Bitcoin is decentralized for example it isn't possessed by anybody. Trades made by Bitcoins are simple as they are not fixing to any nation. Investment is possible through different commercial centers known as bitcoin trades. These enable individuals to trade on Bitcoins utilizing various currencies. Mt. Gox is the biggest Bitcoin exchange, where bitcoin is stored as a virtual digital bank. The record of the considerable number of exchanges, the timestamp information handled in this market is called Blockchain. Each record of blockchain information is encrypted. Trades done by the client's name are made private only wallet ID is made open.

The Bitcoin's simply and similar like a stock but in different way. There are various algorithm using machine learning are utilized on price prediction on stock value. The features influencing Bitcoin are unique. For investors it is mandatory to predict Bitcoin prices. Bitcoin price do not affect by business announcements or government announcements and it is not at all like securities exchange. Thus, we exploit machine learning techniques to foresee the cost of Bitcoin.

Bitcoin is an effective crypto currency brought into the monetary market dependent on its one of a kind convention and Nakamoto's structure. Bitcoin try to achieve decentralization in currency market. Investors in Bitcoin market establish trust connections through the development of Blockchain using cryptography strategies. Bitcoin now-a-days gaining more interest due to innovations in Blockchain and machine learning.

Highly volatile significantly affects exchange methodologies for crypto-currency market. Bitcoin is the major crypto-currency is still facing the volatility problem. Thus, there should be a significant research efforts needed for an efficient way of price

prediction for crypto-currency through machine learning techniques is inevitable. The objectives of this study to the price prediction of bitcoin by feature selection of different machine learning techniques.

Volatility as a proportion of value fluctuations, it significantly affects exchange methodologies and investment choices just as on alternative estimating and proportions of fundamental risk. Bitcoin, as a pioneer in the blockchain money, it assumes a predominant job in an entire cryptographic money showcase capitalization. Subsequently, it is of extraordinary interest are growing these days on data mining and machine learning network to most likely predict Bitcoin value variances

While senders of conventional electronic installments are generally distinguished, investors of bitcoin work in anonymous. No central authority is available for bitcoin, investors don't have to recognize themselves when sending bitcoin to another client. Investors are identified by the address of wallet. It is mandatory for Exchanges to check identity of the clients handled by them, they are not permitted to make buy or sell trade without checking their identity. Bitcoin is more secure for investors.

The objective of the proposed study is for price prediction of bitcoin by feature selection of different machine learning techniques. Intuitively, idea is to first transform order book data into features over time, referred as feature series and then to develop prediction models to consume volatility and feature series simultaneously.

In the following chapter 2, we discuss about the price prediction and related work handled by different authors. In chapter 3, we described our proposed methodology and brief details about our implementation. Under chapter 4, we discuss the results and analysis. In chapter 5, this work is concluded with future enhancement options.

## 2. RELATED WORK

Many existing techniques have been studied by the researchers on crypto-currency market like fluctuations on its prices, social media sentiments, etc.

Fluctuations on crypto-currency prices attracted many researchers, Tian et.al [1] discussed the fluctuation on bitcoin prices by its execution orders such as buy or sell. They handled regression techniques and moving average values. They derived model for time series, which also applied Gaussian time model to predict the bitcoin values. However, they proved their model is efficient on time series data, in our proposed model, we considered dataset of various years and we evaluated based on close price of bitcoin.

Connor et.al [2] studied the bitcoin price through sentiments of various users provided on news columns and social media. Apart from bitcoin, they handled two more crypto-currency for prediction study. They applied feature selection and classification algorithm on the collected dataset along with token weights with positive and negative values. They used three

models namely, Naïve Bayes, regression models and SVM (Support Vector machine). For bitcoin, their experiments shown regression model outperform the others.

Young Bin Kim [3] studied the price fluctuation model in crypto-currency using comments given by users. Apart from bitcoin, they considered Litecoin and Ripple, which are the next two major currencies. User crawled data is taken for their study and they segregated sentiment opinions of five types such as very positive, positive, neutral, negative and very negative. HUISU JANG et.al [4] on his study discussed Bayesian neural networks considering block chain environment. They did machine learning with linear regression model. Their study identifies the bitcoin price and fluctuations.

Anshul et.al [5] used LSTM for bitcoin price forecast. LSTM, one of recurrent neural network algorithm will allow training bitcoin prices as time series data efficiently. In their study, though the time taken for compilation of LSTM model is high than existing ARIMA model, the accuracy is found to be high for LSTM.

Jethin et.al [6] handled Ethereum price and bitcoin price using Google Trends data and twitter texts. Twitter data considered main form of source for arriving decisions on price prediction. It is necessary to understand the effect of tweets on price forecast, which can give quick assessment on buy or sell suggestion to bitcoin traders. The author collected dataset from twitter using hashtag #btc and Ethereum dataset using hashtag #eth. They also extracted sales volume index data from Google. They pre-processed data and this data is given as input for linear regression model for bitcoin price prediction.

The author in [7], proposed a model to identify whether Bitcoin price is depends on volume of tweets or posts by various authors in social media. The author used Google price trend data and positive sentiment tweets as the study is related on price increase. The author collected around two months data of tweets, which is around one million data along with time line is extracted. Tweet Timeline is compared with price of bitcoin on the same time line, which has given correlations on this work. The work proposed by author [8] is finding the sentiment using multiple machine learning techniques. The author collected twitter dataset and performed the analysis, they handle various stages of pre-processing techniques such as removing URL, spelling corrections, and emoticons are replaced with their corresponding polarity values. They used classification techniques such as Naive Bayes algorithm and Support Vector Machine. Using NLTK package, they find the polarity of tweets, from which they arrived positive and negative sentiment.

The author [9] proposed finding anomalies on market. They observed that movement varies throughout the week, over the course of the year and after some time. Strikingly, request and supply increase/decline with the goal that costs are somewhat consistent after some time. They observed that costs are by and large lower on Sundays with the goal that imminent purchasers should move their interest right up 'til today of the week.

In the existing work [10], the authors considered daily price trend of crypto currency, particularly on bitcoin market, considering various features. They applied more than three normalization techniques on dataset, which is collected from quandl.com. Next, they proposed the feature selection problem, in which they considered five features fed into machine learning algorithm such as Bayesian regression, random forest.

### 3. PROPOSED WORK

We have collected the dataset for the document with following details from quandl.com and we applied machine learning algorithm such as decision tree and regression for prediction and price forecast.

#### DATASET DETAILS

As Bitcoin is a kind of stock traded in stock market, dataset will be available in plenty with all time intervals. Live data from 2011 to till date is collected from quandl.com, which provided us the most comprehensive bitcoin price in date wise data. Dataset is extracted to CSV file.

Though there are many authorize websites are available for collecting bitcoin dataset for study, CoinMarketCap is one of the other authorized websites, which provides the transactions that bitcoin traded for the 24 hours of a day. These data are fed from various exchanges handling crypto currency.

Quandl.com has dataset related to finance, economic data from five hundred publishers. Data published in Quandl.com can be exploited for different development platforms and analysis tools. In this proposed work, we have collected the Quandl.com data with name mentioned as "BITSTAMPUSD".

The data collected with following features and stored as data.csv "Time\_stamp, Open, High, Low, Close, Volume\_btc, Volume\_currency, Weighted\_price". The figure, Figure1, represents the view of extracted dataset sample in comma separated file.

A	B	C	D	E	F	G	H
Date	Open	High	Low	Close	Volume (BTC)	Volume (Currency)	Weighted Price
9/13/2011	5.8	6	5.65	5.97	58.37138238	346.0973894	5.929230648
9/14/2011	5.58	5.72	5.52	5.53	61.14598362	341.8548132	5.590797514
9/15/2011	5.12	5.24	5	5.13	80.1407952	408.2590022	5.094271914
9/16/2011	4.82	4.87	4.8	4.85	39.9140068	193.7631466	4.854515047
9/17/2011	4.87	4.87	4.87	4.87	0.3	1.461	4.87
9/18/2011	4.87	4.92	4.81	4.92	119.8128	579.8431027	4.839575594
9/19/2011	4.9	4.9	4.9	4.9	20	98	4.9
9/20/2011	4.92	5.66	4.92	5.66	89.28071068	481.0492629	5.388053693
9/21/2011	5.7	5.79	5.66	5.66	17.62932238	100.5942336	5.706074879

**Figure1: Data set Sample View**

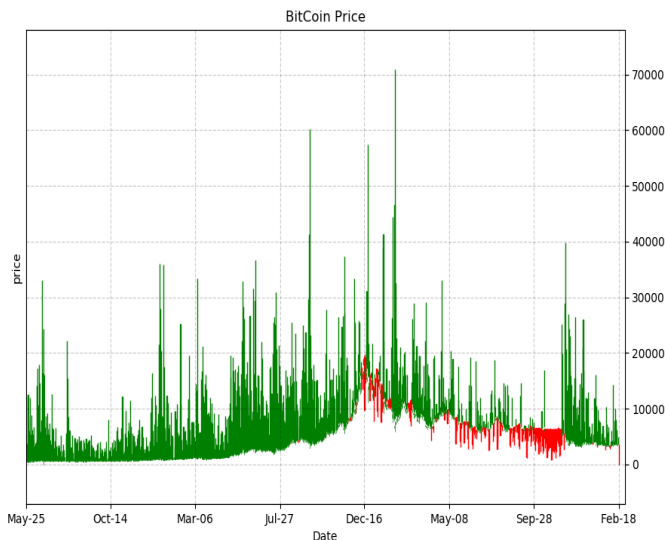
The dataset variable names are described below

Variable name	Attribute Description
Date	Trading Date
Open	Bitcoin Open price for particular time
High	Bitcoin High price achieved for particular time
Low	Bitcoin Low price achieved for particular time
Close	Bitcoin Close price for particular time
Volume (BTC)	Coin volume traded
Volume (Currency)	Coin value traded

Weighted price	Price per coin traded
----------------	-----------------------

**Table 1: Data set variable description**

The dataset collected from 2011 to till data is plotted on the below chart, Figure 2, which clearly depicts that bitcoin is a positive market.



**Figure2: Data Visualization of Bitcoin price**

The below architecture, Figure 3, shows the overall flow of proposed work, which has the following advantages

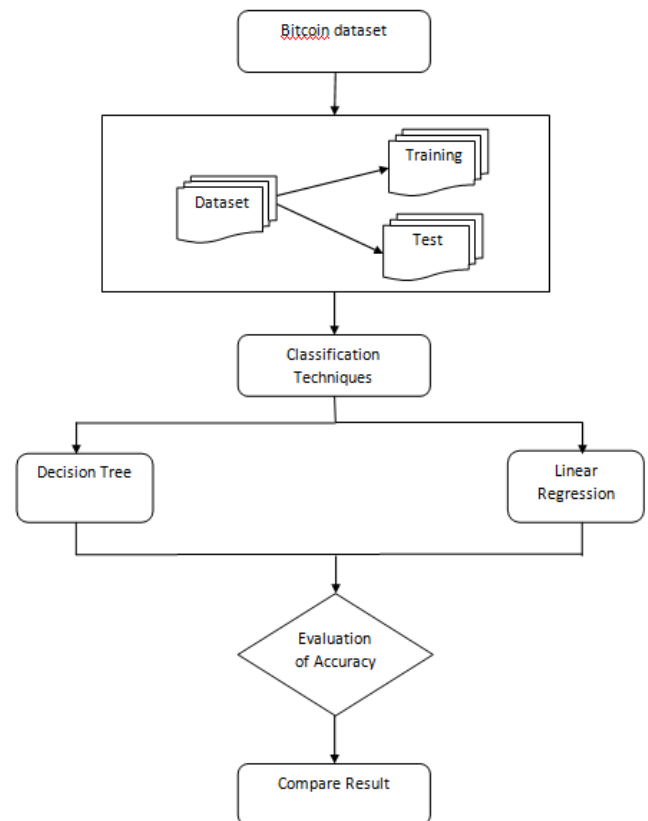
- Implement more than one machine learning to predict the value.
- Accuracy compared to show best algorithm for prediction.

The execution is carried out with following steps

- a. Dataset collection from quandl.com
- b. Dataset pre-processing
- c. Split dataset as train and test set
- d. Apply Machine learning Decision tree and linear regression
- e. Train the model
- f. Give test set and predict values
- g. Price forecast for five days

Bitcoin dataset collected live and stored as bitcoin.csv file is considered as dataset, which is split into train set and test set. We have considered 80% of data as training input for our machine learning algorithm model to train the model. The remaining 20% of data is considered as test for result prediction.

We exploited Lasso and Regression to predict the price trend for 20% test input and the predicted values to plotted and compared for accuracy.



**Figure 3: Overall architecture of crypto currency price prediction**

The machine learning algorithm models used here are decision tree and linear regression is briefly explained below.

#### a. Decision Tree

Decision tree is one of the learning models that are generally utilized in classifications. In this strategy, we split dataset into at least two sets. Internal nodes in Decision tree indicate a test on the features, branch portrays the result and leafs are decisions made after subsequent processing.

Decision Tree works as follows

- ❖ Place the best feature of the dataset as root of tree.
- ❖ Split the dataset into train and test set. Subsets ought to be made so that every subset contains information with a similar feature attribute.
- ❖ Above steps repeated on every subset until we get leaf in the tree.

The class label prediction for a record in decision tree is started from root of tree. The values are compared with the root attributes with next record attributes. This comparison arrives us the corresponding value for the next node to go. The decision tree applied on our dataset is depicted in below figure, Figure 4, in which the input column is considered for the tree is Open, High, Low, Close value of dataset. This is the sample decision taken by our model. The figure shows the three layers of the tree for our dataset.

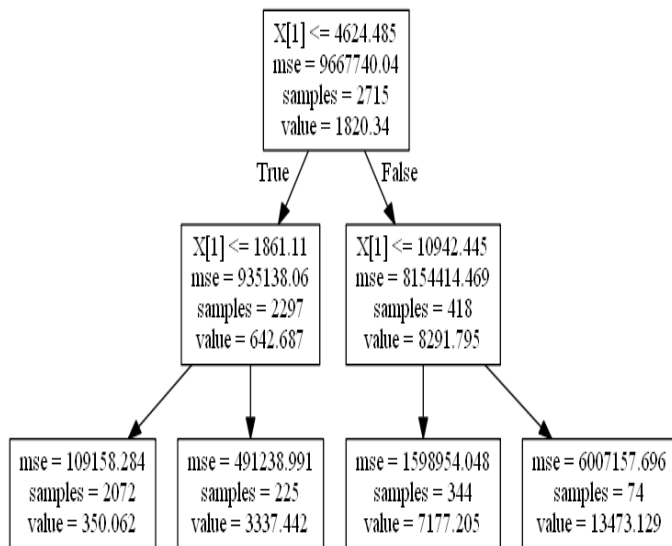


Figure 4: Decision Tree model on Bitcoin dataset

#### b. Linear Regression

In basic, regression, anticipate scores on one attribute from the scores on a second attribute. The attribute that anticipated is known as the model variable and is named as Y. The attribute base for forecasts is known as the prediction attribute and is named as X. At the point when there is just a single prediction attribute, the prediction strategy is called linear regression. In regression model, the subject of prediction of Y and plotted as an element of X frame is a straight line.

The proposed work is implemented in Python 3.6.4 with libraries scikit-learn, pandas, matplotlib and other mandatory libraries. We downloaded dataset from quandl.com with necessary authentication keys. The data downloaded contains up-to date data. The dataset is 80% considered as train set and 20% considered as test set. Machine learning algorithm is applied such as decision tree and regression. Five days forecast price prediction is done using decision tree and regression. The values are compared.

The implementation of regression on pre-processed dataset is done, the predicted price for the given test dataset is plotted against the id. The following figure, Figure 5 shows the predicted versus the original value of bitcoin price for the test set.

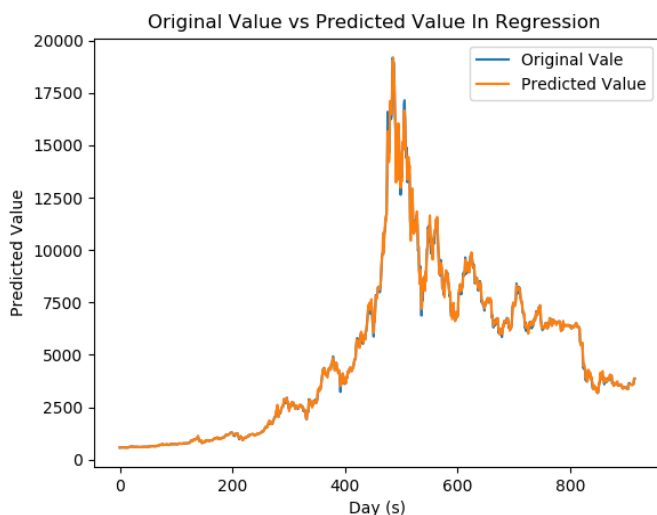


Figure 5: Predicted price Vs Original price

#### 4. RESULTS AND DISCUSSIONS

The result shows that bitcoin price prediction is efficient using liner regression algorithm. Linear regression achieves around 97.5% accuracy in price prediction, whereas decision tree achieves 95.8% accuracy. Our proposed work outperforms existing works accuracy.

The following table shows the accuracy arrived in our experimental study.

Algorithm	Accuracy
Decision Tree	95.88013
Regression	97.59812

Table 2: Experimental Analysis

The below figure, Figure 6, shows the accuracy comparison of our proposed work, in which regression method outperforms decision tree.

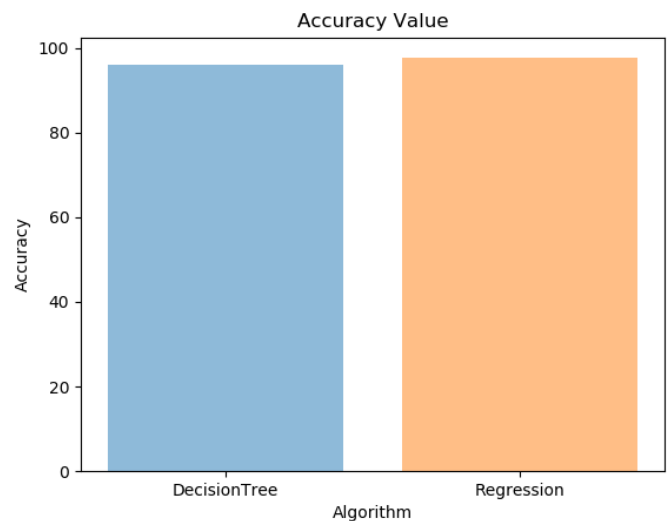


Figure 6: Accuracy Comparison

Price forecast is done for 5 days using machine learning techniques such as Decision tree and regression. The result is compared with the score value to identify the accuracy value and plotted. Predicted price forecast using our proposed method is shown in the below figure, Figure 7.

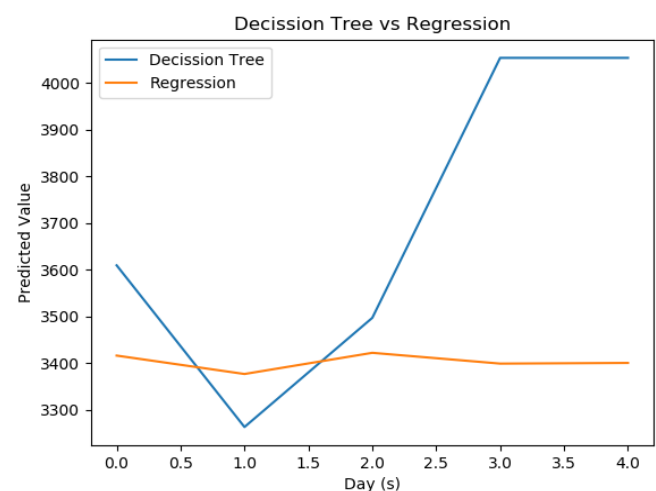


Figure 7: Price Forecast for five days

## 5. CONCLUSIONS

Bitcoin is a booming crypto-currency market, and various researches have been studied in fields of economics and price prediction. In our proposed work, Bitcoin dataset is considered from 2011 to till date price and applied machine learning models such as Decision Tree and Linear regression models. Also the price forecast for five days is done using Decision Tree and Linear regression models. The proposed learning method suggest the best algorithm to choose and adopt for crypto currency prediction problem. The experimental study results show that linear regression outperforms the other by high accuracy on price prediction.

## REFERENCES

- [1] Tian Guo, Nino Antulov-Fantulin, "Predicting short-term Bitcoin price fluctuations from buy and sell orders" Machine Learning , arXiv:1802.04065, Feb 2018
- [2] Connor Lamon, Eric Nielsen, Eric Redondo, "Cryptocurrency Price Prediction Using News and Social Media Sentiment" semantic scholar, 2017
- [3] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, Chang Hun Kim, "Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies" Interdisciplinary Program in Visual Information Processing, Korea University, Seoul, Korea Aug 2016
- [4] HUISU JANG AND JAEWOOK LEE "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information," IEEE access, December 4, 2017,
- [5] Anshul Saxena, T.R. Sukumar, "Predicting bitcoin price using lstm And Compare its predictability with arima model", International Journal of Pure and Applied Mathematics Volume 119 No. 17 2018
- [6] Jethin Abraham, Daniel Higdon, John Nelson, Juan Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis", SMU Data Science Review: Vol. 1 : No. 3 , Article 1.
- [7] Matta, Martina & Lunesu, Maria Ilaria & Marchesi, Michele, "Bitcoin Spread Prediction Using Social And Web Search Media, 2015.
- [8] A, Vishal & Sonawane, Sheetal, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, 2016, 139. 5-15. 10.5120/ijca2016908625.
- [9] Nofer, M, "The value of social media for predicting stock returns: Preconditions, instruments and performance analysis", 2015, 10.1007/978-3-658-09508-6.
- [10] Siddhi Velankar, Sakshi Valecha, Shreya Maji, "Bitcoin Price Prediction using Machine Learning", International Conference on Advanced Communications Technology (ICACT), 2018