



# BITCOIN PRICE PREDICTION USING SVM

Ajinkya Auti<sup>1</sup>, Dhananjay Patil<sup>1</sup>, Omkar Zagade<sup>1</sup>, Prasad Bhosale<sup>1</sup>  
<sup>1</sup>Student, Dept. of Computer Engineering, DIT, Pimpri, Pune, India.

Prashant Ahire<sup>2</sup>

<sup>2</sup>Associate Professor, Dept. of Computer Engineering, DIT, Pimpri, Pune, India.

**Abstract:** Bitcoin is one of the most valuable Crypto currency in the world with the prices as high as 68,078 United States Dollar (USD) in November of 2021. It made Bitcoin a very profitable market for investment but Bitcoin saw many ups and down as well. Prices of Bitcoin have highly fluctuated which make them very difficult to predict. Hence, this research aims to discover the most efficient and highest accuracy model to predict Bitcoin prices from various machine learning algorithms. By using 1-minute interval trading data on the Bitcoin exchange website named bitstamp some different regression models with Scikit-learn and Keras libraries had experimented. The dataset used contains minute by minute prices of Bitcoin of over 5 years and contains almost 3 Million entries. Since, the dataset used is a big data, evaluating the performance of algorithms over a large dataset will give accurate results.

**Keywords:** Bitcoin, SVM, Linear Regression, Crypto-Currency, Random Forest, LSTM, Decision Tree, KNN.

## I. INTRODUCTION

In October 2008, Bitcoin was first introduced by Satoshi Nakamoto through his white paper entitled "Bitcoin: peer-to-peer Electronic Cash System" [1]. Bitcoin is the first decentralized crypto currency while other digital currencies (aka Altcoin or alternative virtual currencies) are created by cloning or adjusting the mechanism of Bitcoin [2]. Currently, Bitcoin attracts the attention of government, business, and financial institutions. In the last few years, the number of transactions in Bitcoin has increased tremendously.

Bitcoin is currently a thriving open-source community and payment network, which is currently owned by more than 106 Million people. There are approximately 4 Lakh daily users of Bitcoin and 200 Million wallets are in use. As the value of Bitcoin fluctuates every day, it would be very interesting for investors to predict the Bitcoin value but it is not easy to predict. The proposed system needs to predict Bitcoin prices with more accurate results. Due to the open nature of the Bitcoin, the following paper describes the Support Vector Machine (SVM), Recurrent Neural Network (RNN), Long-Short Term Memory (LSTM) and Linear Regression (LR), to predict the price of Bitcoin.

## II. RELATEDWORK

### 1. Scikit-learn:

Scikit-learn is an open-source library for analyzing data mining. Scikit-learn can be used for preparing data in several ways: normalization, standardization, and cleaning outlier data or missing data[3].

### 2. Keras:

Keras is an open-source library used for high-level NN. It provides API for NN programming written in Python. It can also be used with Tensorflow, CNTK, and Theano libraries. Dividing codes into parts make Keras easily to build and understand. The parts of generating models normally consist of neural layers, cost functions, optimizer, and activation functions [4].

### 3. Tensor flow:

Tensorflow, created by Google, is an open-source deep learning framework. It can be used to train Neural Network (NN) models and to predict results by using much Graphical Processing Unit (GPU) to collaborate, therefore, powerful algorithms for deep learning and NN can be implemented. Tensorflow can generate data flow graphs for processing when graphs are composed of node groups.

## III. LITERATURE SURVEY

Connor Lamon et al[5] 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.

D. Shah and K. Zhang [6] in recent years, Bitcoin is the most valuable in the crypto currency market. However, prices of Bitcoin have highly fluctuated which make them very difficult to predict. Hence, this research aims to discover the most efficient and highest accuracy model to predict Bitcoin prices from various machine learning algorithms.

S. McNally et al [7] ascertained with what accuracy the price of Bitcoin can be predicted by the use of machine learning techniques. The price data were collected from the Bitcoin Price Index. They implemented Bayesian optimized



RNN along with the LSTM network with the highest accuracy of 52% for LSTM. The ARIMA model was also implemented for comparing it with the deep learning models which performed poorly.

Jang and Lee [8] analyzed a time series of the Bitcoin process and also performed Bayesian Neural Networks(BNN) algorithm and SVM on it. The dataset was extracted from the Bitcoin charts and blockchain.info. The work showed that BNN performed well for prediction of the Bitcoin price and was also able to explain the reason for high volatility in its price.

Madan et al [9] are also use the datasets from Ok coin, but separating the data into series of 30, 60, and 120 minutes. Binomial Logistic Regression, Support Vector Machine (SVM), and Random Forest are used to predict Bitcoin's prices with the accuracy at 97% and 55% for the next 10 minute's prices. However, there is no cross-validate in this research which might cause the obtained models to be over fitting.

PANT, Dibakar Raj et al[10] used the sentiment on Twitter as input data for RNN for predicting the price of Bitcoin, and they got the overall price prediction accuracy using RNN is found to be 77.62%.

#### IV. PREDICTION TECHNIQUES

**A. Linear regression model:** 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.[11]

**B. K-means Clustering:** K-means creates k groups from a set of objects so that the members of a group are more similar and based on this data is clustered as normal, stressed or highly stressed.[12]

**C. Naive Bayes:** Naive Bayes techniques are a great deal of coordinated learning figurings reliant on applying Bayes' speculation with the "honest" supposition of opportunity

between each pair of features. Overlooking their plainly over-improved suppositions, guiltless Bayes classifiers have worked very well in some genuine conditions. They require a limited measure of preparing information to survey the critical parameters. Honest Bayes understudies and classifiers can be unbelievably speedy appeared differently in relation to progressively present day systems. The decoupling of the class prohibitive component dispersals suggests that each movement can be uninhibitedly evaluated as a one dimensional scattering. This along these lines decreases issues originating from the scourge of dimensionality.[12]

**D. Random Forest:** Random Forests get the outfit learning framework where distinctive weak understudies are merged to make a strong understudy. It is a meta estimator that fits various decision tree classifiers on various sub-primer of the enlightening assortment and use averaging to improve the farsighted accuracy and authority over fitting. The sub-test size is reliably proportional to the rule data test.

**E. KNN(K-Nearest Neighbors):** It very well may be utilized for both order and relapse issues. Be that as it may, it is all the more generally utilized in characterization issues in the business [10]. K nearest neighbors is a straight forward calculation that stores every single accessible case and arranges new cases by a lion's share vote of its k neighbors. The case being allotted to the class is generally normal among its K closest neighbors estimated by a separation work. These separation capacities can be Euclidean, Manhattan, Minowski and Hamming separation. Initial three capacities are utilized for constant capacity and fourth one (Hamming) for clear cut factors. On the off chance that  $K = 1$ , at that point the case is basically relegated to the class of its closest neighbor. Now and again, picking K ends up being a test while performing KNN displaying.

**F. 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.[11]

#### V. COMPARISON OF RELATED WORKS

Title of Paper	Algorithms	Advantages	Limitations
<ul style="list-style-type: none"> <li>Bitcoin Price Prediction using Deep Learning</li> </ul>	Bayesian Regression and GLM/Random forest:	<ul style="list-style-type: none"> <li>It works the prediction by taking the coinMarkup cap.</li> <li>Quandl is used to filter the dataset by using the MATLab properties.</li> </ul>	<ul style="list-style-type: none"> <li>It is a long process for filter the data.</li> <li>Low redundancy to perform the prediction.</li> </ul>
<ul style="list-style-type: none"> <li>Bitcoin Volatility Forecasting with a Glimpse into Buy</li> </ul>	LSTM(Long Short Term Memory ) and ARIMA(Autoregressive	<ul style="list-style-type: none"> <li>It is easy way to buy and sell the Bitocins.</li> <li>The process of buying</li> </ul>	<ul style="list-style-type: none"> <li>One drawback is there is no proof for transaction.</li> </ul>



and Sell Orders.	integrated moving average)	and selling the Bitcoins are done in online. <ul style="list-style-type: none"> <li>It is comfortable place to done the transactions</li> </ul>	<ul style="list-style-type: none"> <li>Conversion will be late.</li> </ul>
<ul style="list-style-type: none"> <li>Bayesian regression and Bitcoin.</li> </ul>	Bayesian regression	<ul style="list-style-type: none"> <li>The Advantage of Bayesian regression in Bitcoin price prediction results has been showed in binary values.</li> <li>It helps to understand the results very neatly.</li> </ul>	<ul style="list-style-type: none"> <li>It takes long time to solve the data set.</li> </ul>
<ul style="list-style-type: none"> <li>Project Based Learning: Predicting Bitcoin Prices using Deep Learning</li> </ul>	CNN(Convolutional Neural Networks) and RNN(Recurrent neural networks)	<ul style="list-style-type: none"> <li>The main Advantage of CNN is Weight Sharing.</li> <li>It is easily calculate the large data set prices.</li> </ul>	<ul style="list-style-type: none"> <li>The Convolution is a significantly slower operation then, say maxpool, both forward and backward.</li> </ul>
<ul style="list-style-type: none"> <li>Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data</li> </ul>	Sentiment Analysis, RNN, LSTM	<ul style="list-style-type: none"> <li>Uses real time Tweets and social media trends to detect short term fluctuation.</li> </ul>	<ul style="list-style-type: none"> <li>Low accuracy.</li> <li>Trends may affect the result and price may go sideways.</li> </ul>
<ul style="list-style-type: none"> <li>Performance Comparison of Bitcoin Prediction in Big Data Environment</li> </ul>	Linear Regression	<ul style="list-style-type: none"> <li>Uses Apache Spark Cluster and GPU.</li> <li>Apache Spark runs 2x faster than GPU.</li> </ul>	<ul style="list-style-type: none"> <li>Can't Predict the price but the environment can be used to predict the price faster.</li> </ul>

VI. ADVANTAGES OF USING SVM

Support Vector Machine is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. While a noisy data may affect the result.

- It is convincing in high dimensional spaces.
- It works well with clear margin of separation.
- It is effective in cases where number of dimensions is greater than the number of samples.

VII. CONCLUSION

In this survey report after comparing several algorithms i.e. Linear Regression, LSTM, Bayesian Regression, RNN, Random Forest used in previous research with SVM we can conclude that SVM shows better results. The machine learning algorithms will improves that feature idea of crypto currencies. That will improves the market price of global investments. In this paper we proposed the new algorithm to find the future price with enhanced accuracy. That helps the customer increment the profits.

VIII. REFERENCES

- S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
- G. Heleman, M. Rauchs, "Global cryptocurrency benchmarking study," Cambridge Centre for Alternative Finance (2017).
- L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," arXiv preprint arXiv:1309.0238, 2013.
- C. François, "Keras: Deep learning library for theano and tensorflow," URL: <https://keras.io/k>, 2015.
- Connor Lamon, Eric Nielsen, Eric Redondo, "Cryptocurrency Price Prediction Using News and Social Media Sentiment" semantic scholar, 2017.
- D. Shah and K. Zhang, "Bayesian regression and Bitcoin," in 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton),2015, pp. 409-415.
- S. McNally, J. Roche, and S. Caton, "Predicting the price of Bitcoin using machine learning," in Parallel, Distributed and Network-based Processing (PDP), 2018 26th Euromicro International Conference on. IEEE, 2018, pp. 339-343.
- H. Jang and J. Lee, "An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on blockchain information," IEEE



Access, vol. 6, pp. 5427–5437, 2018.

- [9]. I. Madan, S. Saluja and A. Zhao. "Automated Bitcoin Trading via Machine Learning Algorithms," 2015.
- [10]. PANT, Dibakar Raj, et al. Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. In: 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS). IEEE, 2018. p. 128-132.
- [11]. KarunyaRathan, Somarouthu V, et al. Crypto-Currency price prediction using Decision Tree and Regression techniques.
- [12]. LekkalaSreekanth Reddy, Dr.P. Sriramya. A Research OnBitcoin Price Prediction Using Machine Learning Algorithms.