Navigating the Crypto Roller Coaster: Predictions for BITCOINS Path

[1]Singireddy Deekshitha, [2]N.Naveen Kumar
[1]*MCA Student, Department of Information Technology, Jawaharlal Nehru Technological University, India*
[2]*Associate Professor of CSE, Department of Information Technology, Jawaharlal Nehru Technological University, India*

Abstract: In this project we are using Subspace Learning Algorithm called Supervised Principal Coefficients Embedding (SPCE) to reduce high dimension data into low dimension. In machine learning all algorithms get trained on past datasets and this dataset often contains irrelevant (unimportant) features and this features will be removed by applying SPCE algorithm. SPCE algorithm will find coefficients or fitness or each attribute by calculating accuracy and whatever features gave high accuracy will be selected and remaining attributes will be remove out. In this project we are using Blockchain technology to store BITCOIN historic price data. This data will be retrieved by machine learning algorithms such as Logistic Regression, Support Vector Machine and ARIMA. This algorithms will get trained on this data and then predict future price of Bitcoin.

Index terms - Supervised Principal Coefficients Embedding (SPCE), machine learning algorithms, Logistic Regression, Support Vector Machine and ARIMA.

1. INTRODUCTION

Bitcoin [1] is the world's most esteemed cryptocurrency, exchanged on over 40 trades worldwide that acknowledge in excess of 30 different monetary forms. As per https://www.blockchain.info/, its ongoing business sector capitalisation is $9 billion USD, and it sees more than 250,000 exchanges consistently. Bitcoin, as a cash, gives an additional opportunity to cost forecast because of its extremely youthful age and going with instability, which is far higher than that of ordinary monetary forms [2]. It is additionally unmistakable from run of the mill government issued types of money regarding its straightforwardness; there is no complete information on cash exchanges or cash available for use for government issued types of money. The forecast of mature monetary business sectors, for example, the securities exchange, has been broadly investigated [3], [4]. Bitcoin gives a fascinating similarity to this since it is a period series expectation issue in a market that is still in its beginning phases. Conventional time series expectation draws near, for example, Holt-Winters remarkable smoothing models, depend on direct suspicions and require information that can be separated into pattern, occasional, and clamor classes to find success [5]. This strategy is more qualified to assignments like foreseeing deals, which have occasional impacts. These techniques are insufficient for this reason since the Bitcoin market has no irregularity and is very unpredictable. Given the undertaking's multifaceted nature, deep learning presents a charming specialized answer in light of its exhibition in related fields.

2. LITERATURE SURVEY

The advent of Bitcoin, introduced by Satoshi Nakamoto in 2008 [1], has sparked a significant interest in both academic and financial circles. Researchers have explored various aspects of Bitcoin, including its technological underpinnings, economic implications, and potential for financial applications. This literature survey aims to provide an overview of key studies related to Bitcoin price prediction, portfolio diversification with cryptocurrencies, and the application of machine learning techniques in forecasting financial time series.

Nakamoto's seminal paper, "Bitcoin: A Peer-to-Peer Electronic Cash System" [1], laid the foundation for the development of decentralized digital currencies.

This paper proposed a system for peer-to-peer transactions without the need for intermediaries, based on cryptographic principles. Since its publication, Nakamoto's work has inspired a plethora of research exploring the technical, economic, and social aspects of cryptocurrencies.

Briere, Oosterlinck, and Szafarz (2013) investigated the potential of Bitcoin as an asset class for portfolio diversification in their paper titled "Virtual currency, tangible return: Portfolio diversification with bitcoins" [2]. They found that incorporating Bitcoin into a traditional investment portfolio could enhance diversification benefits, particularly during periods of financial turmoil. This study highlights the growing recognition of Bitcoin as a legitimate investment vehicle and its potential role in mitigating portfolio risk.

In the realm of financial forecasting, researchers have explored various methodologies for predicting the prices of stocks, currencies, and other financial assets. Kaastra and Boyd (1996) proposed a neural network-based approach for forecasting financial and economic time series in their paper "Designing a neural network for forecasting financial and economic time series" [3]. They demonstrated the effectiveness of neural networks in capturing nonlinear relationships and patterns in financial data, offering advantages over traditional econometric models.

White (1988) explored the use of neural networks for economic prediction in his paper "Economic prediction using neural networks: The case of IBM daily stock returns" [4]. By training neural networks on historical stock returns data, White demonstrated the potential of neural networks in predicting future price movements. This study paved the way for further research into the application of machine learning techniques in financial forecasting.

Traditional time series forecasting methods, such as Holt-Winters exponential smoothing, have been widely used in financial analysis. Chatfield and Yar (1988) discussed practical issues related to Holt-Winters forecasting in their paper "Holt-Winters forecasting: some practical issues" [5]. While effective for capturing seasonal patterns in data, these methods may be less suitable for highly volatile markets like Bitcoin, where traditional assumptions may not hold.

Scott (2016) compiled a database of academic papers on Bitcoin in his blog post titled "Bitcoin academic paper database" [6]. This resource provides valuable insights into the diverse range of research topics related to Bitcoin, including cryptography, economics, and finance. Researchers can leverage this database to stay informed about the latest developments in the field and identify potential research gaps.

In recent years, there has been a growing interest in applying machine learning techniques to financial forecasting tasks. Rechenthin (2014) explored machine learning classification techniques for analyzing and predicting high-frequency stock direction in his paper "Machine-learning classification techniques for the analysis and prediction of high-frequency stock direction" [7]. By leveraging machine learning algorithms, Rechenthin demonstrated improvements in predicting stock price movements, particularly in high-frequency trading environments.

Shah and Zhang (2014) investigated Bayesian regression models for predicting Bitcoin prices in their paper "Bayesian regression and Bitcoin" [8]. By incorporating Bayesian techniques, they sought to capture the uncertainty inherent in financial markets and improve the accuracy of price forecasts. This study underscores the importance of probabilistic modeling in capturing the stochastic nature of Bitcoin price movements.

In summary, the literature on Bitcoin price prediction and financial forecasting encompasses a wide range of methodologies, including traditional econometric models, neural networks, and machine learning algorithms. Researchers continue to explore innovative approaches to address the challenges posed by the dynamic and volatile nature of cryptocurrency markets. By leveraging insights from diverse disciplines, including economics, finance, and computer science, scholars aim to enhance our understanding of Bitcoin's role in the global financial landscape and develop robust forecasting models for informed decision-making.

3. METHODOLOGY

i) Proposed Work:

The proposed system leverages Blockchain technology to store historical Bitcoin price data securely and transparently. The process begins by loading the Bitcoin data from the Blockchain, ensuring the integrity and reliability of the dataset. Next, the dataset undergoes preprocessing to handle missing values, normalize the data, and remove irrelevant features.

The core of the system lies in the implementation of the Supervised Principal Coefficients Embedding (SPCE) algorithm, a Subspace Learning technique. SPCE reduces the high-dimensional Bitcoin price data into a lower-dimensional space by identifying and selecting the most relevant features based on their accuracy and contribution to the prediction model.

Subsequently, the reduced dataset is fed into various machine learning algorithms, including Support Vector Machines (SVM), logistic regression, and Autoregressive Integrated Moving Average (ARIMA). These algorithms analyze the data to predict future Bitcoin price movements, leveraging the optimized feature set obtained through SPCE.

By integrating Blockchain technology for data storage and employing advanced subspace learning and machine learning techniques, the proposed system offers a robust framework for accurate and efficient Bitcoin price prediction. This system not only improves prediction accuracy but also enhances the scalability, security, and transparency of Bitcoin price forecasting in the dynamic cryptocurrency market.

ii) System Architecture:

The system architecture includes modules for dataset loading from Blockchain, preprocessing techniques such as normalization and feature selection via Supervised Principal Coefficients Embedding (SPCE). Machine learning models like Support Vector Machines (SVM), logistic regression, and Autoregressive Integrated Moving Average (ARIMA) are employed for prediction. Results are evaluated using quality metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and accuracy. This architecture enables efficient handling of Bitcoin price data, feature selection, and accurate prediction using diverse machine learning techniques while ensuring robust evaluation through quality metrics.

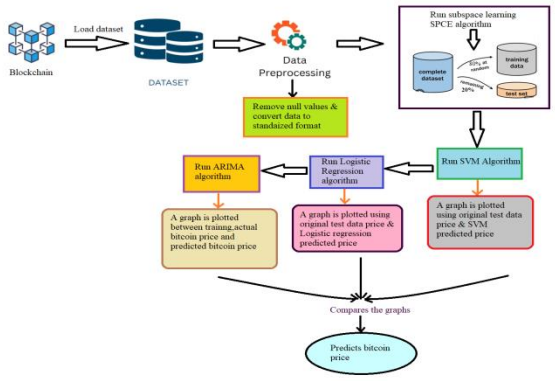


Fig 1 Proposed Architecture

iii) Implementation:

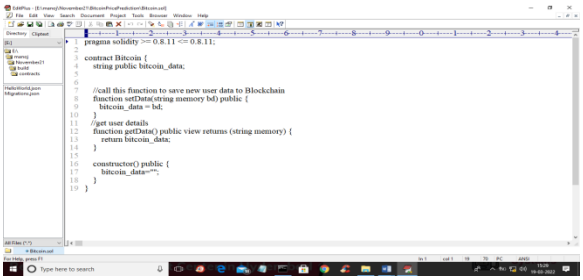
In this project we are using Subspace Learning Algorithm called Supervised Principal Coefficients Embedding (SPCE) to reduce high dimension data into low dimension. In machine learning all algorithms get trained on past datasets and this dataset often contains irrelevant (unimportant) features and this features will be removed by applying SPCE algorithm. SPCE algorithm will find coefficients or fitness or each attribute by calculating accuracy and whatever features gave high accuracy will be selected and remaining attributes will be remove out.

In this project we are using Blockchain technology to store BITCOIN historic price data. This data will be retrieved by machine learning algorithms such as Logistic Regression, Support Vector Machine and ARIMA. This algorithms will get trained on this data and then predict future price of Bitcoin.

In all algorithms Logistic Regression is giving accurate prediction prices. To calculate performance of each algorithm we have divided dataset into train and test where application used 80% dataset for training and 20% dataset for testing. After training model we will apply test data on trained model to predict prices and this prices will be check with original prices and if there is not much difference

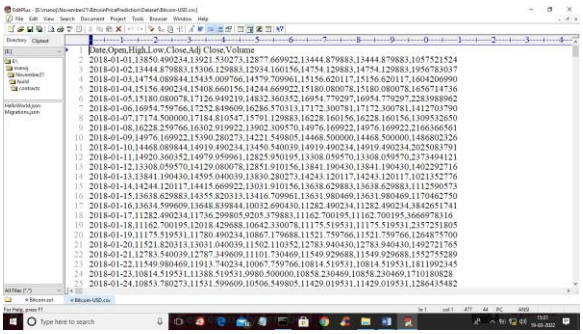
between original and predicted prices then algorithm will be consider as accurate.

All Bitcoin data will be saved inside Blockchain Ethereum tool and to store this data in Blockchain Ethereum we have designed SOLIDITY code which will allow us to read Bitcoin historic data.



In above screen we will call SetData function to store Bitcoin prices in Blockchain and will call GetData function to read data back.

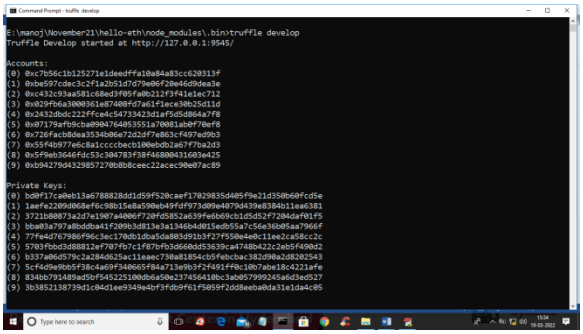
Below is the Bitcoin dataset used in this project



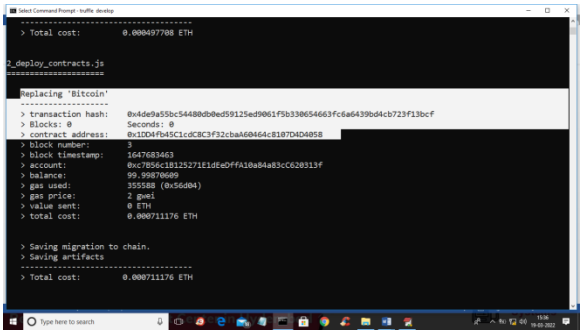
In above dataset we have Bitcoin open and close prices with dates and this data will be stored in Blockchain and then read back and trained by machine learning algorithms to predict future prices.

In below screen we are showing concept to deploy SOLIDITY code on Blockchain and this SOLIDITY can be called from Python program to store and read data.

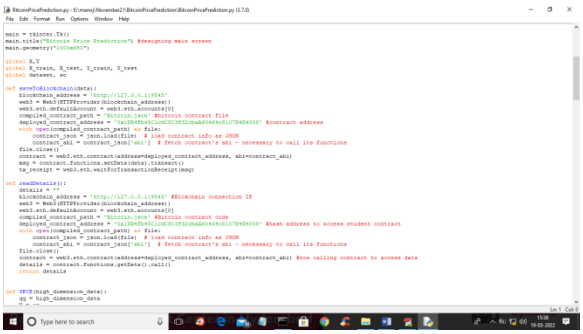
First go inside 'hello-eth/node_modules/.bin' folder and then look for file called 'runBlockchain.bat' and then double click on that file to start Ethereum tool like below screen



In above screen private keys are generated and in same window type 'truffle migrate' command and press enter key to deploy Bitcoin SOLIDITY contract in Blockchain and will get below output



In above screen we can see Bitcoin contract deployed and we can see its contract address and this address we will specify in python program to access data from that Blockchain Solidity code and now let above Blockchain console running. In below screen you can see that contract calling code in python



In above screen read red colour comments to know about Blockchain Bitcoin contract calling.

iv) Modules:

To implement this project we used the following modules is given below:

- Load bitcoin data from blockchain
- Preprocess dataset
- Run subspace learning SPACE algorithm
- Run SVM algorithm
- Run logistic regression algorithm
- Run ARIMA algorithm

Module Description:

1. Load Bitcoin Data from Blockchain:

This module retrieves historical Bitcoin price data from the Blockchain, ensuring data integrity and transparency.

2. Preprocess Dataset:

In this stage, the dataset undergoes preprocessing steps such as handling missing values, normalizing the data, and removing irrelevant features to ensure data quality and suitability for analysis.

3. Run Subspace Learning SPACE Algorithm:

The Subspace Learning algorithm, specifically Supervised Principal Coefficients Embedding (SPCE), is executed to reduce the dimensionality of the dataset. SPCE identifies and selects the most relevant features based on their accuracy and contribution to the prediction model.

4. Run SVM Algorithm:

Support Vector Machines (SVM) algorithm is applied to the reduced dataset to predict future Bitcoin price movements. SVM utilizes a hyperplane to separate data points into different classes, making it suitable for classification and regression tasks.

5. Run Logistic Regression Algorithm:

Logistic Regression algorithm is employed to model the relationship between the dependent variable (Bitcoin price) and independent variables (selected features). It estimates the probabilities of different outcomes and is well-suited for binary classification tasks.

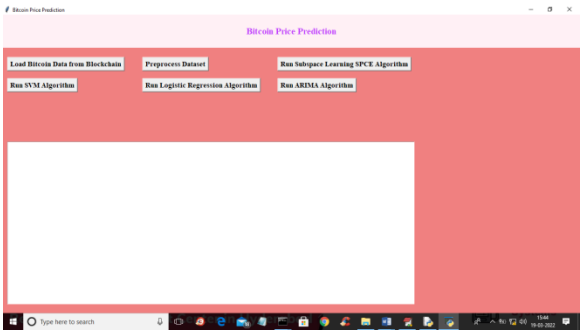
6. Run ARIMA Algorithm:

Autoregressive Integrated Moving Average (ARIMA) algorithm is utilized to forecast future Bitcoin prices based on historical price data and trends. ARIMA models capture the linear dependencies and temporal patterns in time series data, making them suitable for predicting Bitcoin price movements.

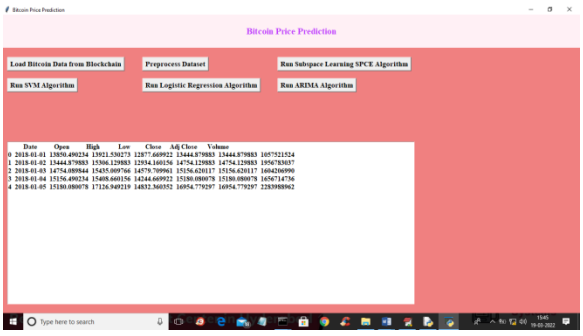
These modules collectively form the pipeline for analyzing Bitcoin price data, reducing dimensionality, and predicting future price trends using a combination of machine learning and statistical techniques.

4. EXPERIMENTAL RESULTS

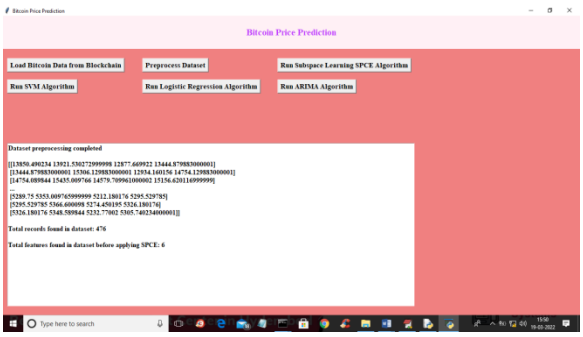
To run project double click on 'run.bat' file to get below screen



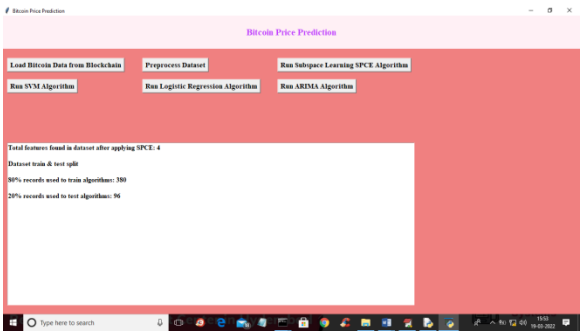
In above screen click on 'Load Bitcoin Data from Blockchain' button to load data from Blockchain and get below output



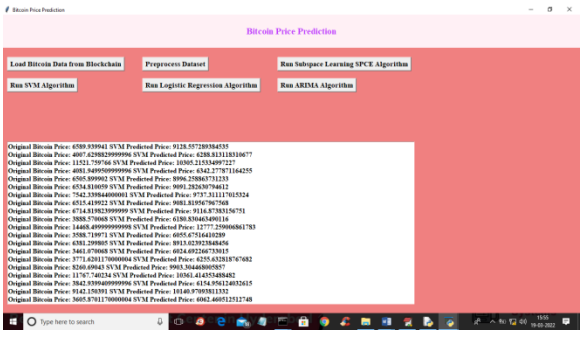
In above screen dataset loaded from Blockchain and now click on 'Preprocess Dataset' button to preprocess dataset such as replacing missing values with 0 and then find total attributes before apply SPCE



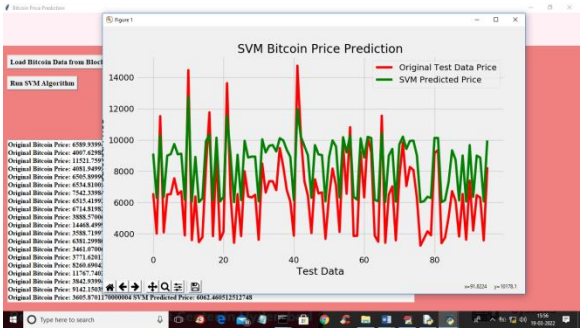
In above screen dataset is normalized and then we can see dataset contains total 476 records and before applying SPCE dataset contains 6 attributes. Now click on 'Run Subspace Learning SPCE Algorithm' button to apply SPCE algorithm to reduce high dimension 6 attributes to low dimension and then will get below output



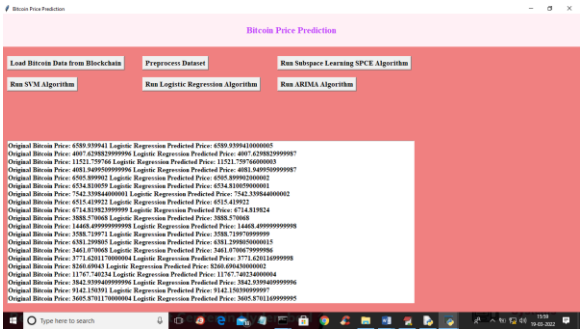
In above screen after applying SPCE attributes/features size reduced to 4 from 6 and we can see application using 380 records for training and 96 records for testing. Now dataset is ready and now click on 'Run SVM Algorithm' button to train SVM and get below output



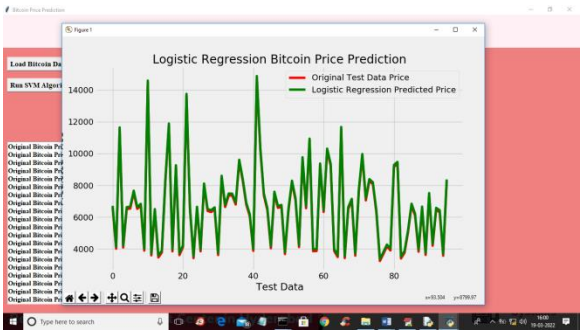
In above screen we can see original Bitcoin prices and then we can see Bitcoin predicted prices and below graph showing difference between original and predicted prices



In above graph x-axis represents DAYS/DATE and y-axis represents Bitcoin prices and red line represents original prices and green line represents predicted prices. In above graph we can see there is huge difference between predicted and original prices so we can say SVM is not accurate and now close above graph and then click on 'Run Logistic Regression Algorithm' button to train LR and get below output

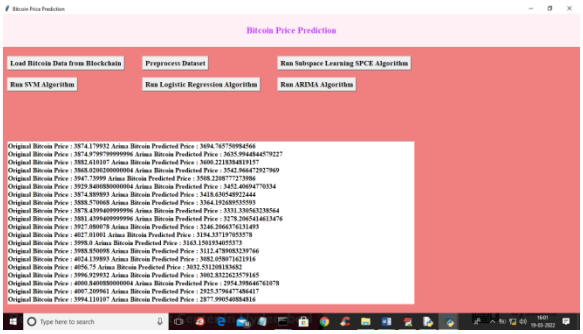


In above screen we can see logistic prediction prices and original prices and below is the LR graph of price prediction

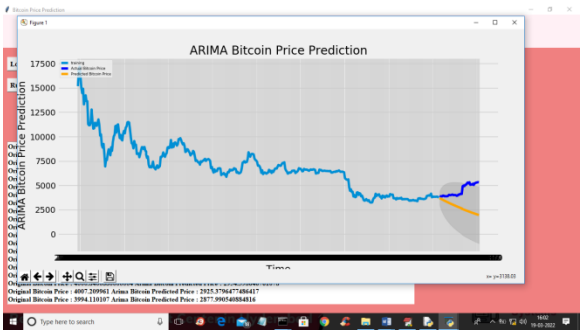


In above graph we can see LR is giving accurate price prediction as both lines are fully overlap and we can

see there less difference between both lines so LR is accurate and now click on 'Run ARIMA Algorithm' button to train ARIMA and get below output



In above screen we can see original and ARIMA predicted prices and below is the ARIMA prediction graph



In above ARIMA graph light blue line is the training prices and dark blue line is the test data and yellow line is the predicted data and both test and predicted data is shadowed so we can say ARIMA prediction is also accurate.

In this project we have evaluated the performance of SVM, Logistic Regression and ARIMA and in all algorithm Logistic Regression and ARIMA is giving better prediction prices.

5. CONCLUSION

After establishing the learning framework and completing the normalization, we intend to use the methods mentioned above and choose the best method to solve the Bitcoin prediction problem. In above ARIMA graph light blue line is the training prices and dark blue line is the test data and yellow line is the predicted data and both test and predicted data is shadowed so we can say ARIMA prediction is also accurate. In this project we have evaluated the performance of SVM, Logistic Regression and ARIMA and in all algorithm Logistic Regression and ARIMA is giving better prediction prices.

6. FUTURE SCOPE

In future iterations, we aim to enhance the predictive accuracy and robustness of Bitcoin price forecasting by incorporating advanced deep learning techniques such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks. Additionally, exploring ensemble methods and hybrid models combining multiple algorithms could further improve prediction accuracy. Furthermore, integrating sentiment analysis from social media and news sources to capture market sentiment could provide valuable insights for more accurate predictions. Continual refinement of algorithms and the incorporation of additional data sources will enable us to stay ahead in the dynamic cryptocurrency market and achieve even better forecasting results.

REFERENCE

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
[2] M. Briere, K. Oosterlinck, and A. Szafarz, "Virtual currency, tangible ` return: Portfolio diversification with bitcoins," Tangible Return: Portfolio Diversification with Bitcoins (September 12, 2013), 2013.
[3] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10, no. 3, pp. 215–236, 1996.
[4] H. White, "Economic prediction using neural networks: The case of ibm daily stock returns," in Neural Networks, 1988., IEEE International Conference on. IEEE, 1988, pp. 451–458.
[5] C. Chatfield and M. Yar, "Holt-winters forecasting: some practical issues," The Statistician, pp. 129–140, 1988.
[6] B. Scott, "Bitcoin academic paper database," suitpossum blog, 2016.
[7] M. D. Rechenthin, "Machine-learning classification techniques for the analysis and prediction of high-frequency stock direction," 2014.
[8] D. Shah and K. Zhang, "Bayesian regression and bitcoin," in Communication, Control, and Computing

(Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014, pp. 409–414.
[9] G. H. Chen, S. Nikolov, and D. Shah, "A latent source model for nonparametric time series classification," in Advances in Neural Information Processing Systems, 2013, pp. 1088–1096.
[10] I. Georgoula, D. Pournarakis, C. Bilanakos, D. N. Sotiropoulos, and G. M. Giaglis, "Using time-series and sentiment analysis to detect the determinants of bitcoin prices," Available at SSRN 2607167, 2015.
[11] M. Matta, I. Lunesu, and M. Marchesi, "Bitcoin spread prediction using social and web search media," Proceedings of DeCAT, 2015.
[12] "The predictor impact of web search media on bitcoin trading volumes."
[13] B. Gu, P. Konana, A. Liu, B. Rajagopalan, and J. Ghosh, "Identifying information in stock message boards and its implications for stock market efficiency," in Workshop on Information Systems and Economics, Los Angeles, CA, 2006.
[14] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," 2015.
[15] I. Madan, S. Saluja, and A. Zhao, "Automated bitcoin trading via machine learning algorithms," 2015.