Journal of
Computer & Electrical and
Electronics Engineering Sciences

Original Article

DOI: 10.51271/JCEEES-0003

Investigation of financial applications with blockchain technology

Muhammed Ali Muhammed

Department of Electronic and Computer Engineering, Faculty of Engineering, Çankırı University, Çankırı, Türkiye

Cite this article: Muhammed MA. Investigation of financial applications with blockchain technology. *J Comp Electr Electron Eng Sci*. 2023; 1(1): 10-14.

Corresponding Author: Mohammed Ali Mohammed, Lm9997621@gmail.com

Received: 05/04/2023 ◆ **Accepted**: 11/04/2023 ◆ **Published**: 28/04/2023

ABSTRACT

The advent of blockchain technology and the rise of cryptocurrencies have had a profound impact on the financial industry and beyond. This article explores the recent advancements in machine learning (ML) and blockchain technology for the prediction of cryptocurrency prices. Our study presents a ML system that utilizes six different cryptocurrency datasets and applies various ML techniques such as LSTM, CNN, SVM, KNN, XGBoost, Astro ML, LASSO, RIDGE, linear regression, DT, and GP to predict prices. The research found that complex models may not always perform better than simpler models, and our proposed model had better performance as evaluated by the RMSE. The research evaluated various machine learning techniques for predicting cryptocurrency prices and reported the following RMSE values: Bitcoin prediction using Nadaraya-Watson kernel regression yielded an RMSE of 0.17, while Dogecoin prediction with linear regression resulted in an RMSE of 0.032. Ethereum price prediction using Gaussian regression achieved an RMSE of 0.02. For USD Coin, a combination of XGBoost, Gaussian regression, and Ridge techniques led to an RMSE of 0.014. Binance Coin price prediction using Gaussian regression had an RMSE of 0.032, and finally, Cardano Coin prediction employing LSTM reached an RMSE of 0.059.

This study demonstrated the effectiveness of various machine learning techniques in predicting cryptocurrency prices. The research contributes valuable insights to the field and can guide future work in cryptocurrency price prediction. The proposed model achieved promising results as evaluated by the RMSE metric.

Keywords: Blockchain, Cryptocurrency, Machine Learning.

INTRODUCTION

Over the past decade, the global economy and payment methods have undergone a significant transformation, primarily due to the emergence of digital currencies facilitated by advancements in information technology (IT). These digital currencies, such as Bitcoin and Ethereum, have gained popularity because of their potential for gains and expanding user base, and they are supported by blockchain technology. Blockchain enables decentralized digital currencies that offer increased autonomy, but this also creates significant challenges for governments and financial institutions (Gowda & Chakravorty, 2021). Bitcoin and Ethereum are currently the most widely used cryptocurrencies, with large trading volumes and market capitalizations. For instance, in April 2021, Bitcoin had a market capitalization of USD 1,304 billion and a trading volume of USD 64 billion, while Ethereum's market capitalization was USD 265 billion with a trading volume of USD 38 billion (Kim et al., 2021a). To manage potential risks for investors, academic researchers are currently examining the future value of cryptocurrencies and the market.

In spite of the difficulties encountered, ML and blockchain technologies have made remarkable strides in recent years, leading to the creation of novel and better products utilized by billions of people globally. The primary focus of researchers has been on implementing these methods in financial markets, including the forecasting of stock market trends and the detection of manipulation, which also have implications for cryptocurrencies as financial assets. In comparison to other fields, blockchain and ML research are comparatively young and less advanced, with more emphasis placed on non-technical aspects of Blockchain technology. Nevertheless, the rapid advancement of ML and the novelty of Blockchain technology provide an opportunity for future research and development in this field.

Literature Review

This related work covers various studies and experiments related to Bitcoin. (Gundlach et al., 2021) examined the correlation between the memory pool's availability and the dispersion of Bitcoin transaction confirmation times. (Koops, 2018) studied the probability distribution of the time it takes for Bitcoin transactions to be confirmed and developed

This work is licensed under a Creative Commons Attribution 4.0 International License.

stochastic fluid queueing algorithms to represent it. (Pabuçcu et al., 2020) utilized ML methods to forecast fluctuations in Bitcoin prices. (Rahouti et al., 2018) investigated the main security issues in the Bitcoin protocol, while (Hitam & Ismail, 2018) compared various ML algorithms for Bitcoin forecasting. (Ho et al., 2021)developed ANN models and ML algorithms to forecast the value of Bitcoin (Jagannath et al., 2021) presented a self-adaptive approach based on the jSO optimization algorithm to accurately predict the value of Bitcoin. (Yogeshwaran et al., 2019) explained how a trained machine model can predict the price of a cryptocurrency given enough information and computing power.

Our article seeks to determine the most accurate pricing prediction algorithm for selected cryptocurrencies. To achieve this, we plan to utilize several ML techniques, including linear regression, LASSO, decision tree, Ridge and XGboost, SVM, KNN, Kernel regression, CNN, and LSTM. The extensive literature on predicting cryptocurrency prices supports our research direction.

METHODS

Our model is a ML based system that utilizes six different cryptocurrency datasets for its operation. The first step in building the model is preprocessing of the data, which involves normalizing the data to make it suitable for the ML algorithms. After preprocessing, various ML techniques are applied to the data to create the final model. These techniques include LSTM, CNN, SVM, KNN, XGBoost, Astro ML, LASSO, RIDGE, linear regression, DT, and GP. The choice of technique is determined by the specific problem being solved and the characteristics of the data. The goal of this model is to predict prices of the cryptocurrency.

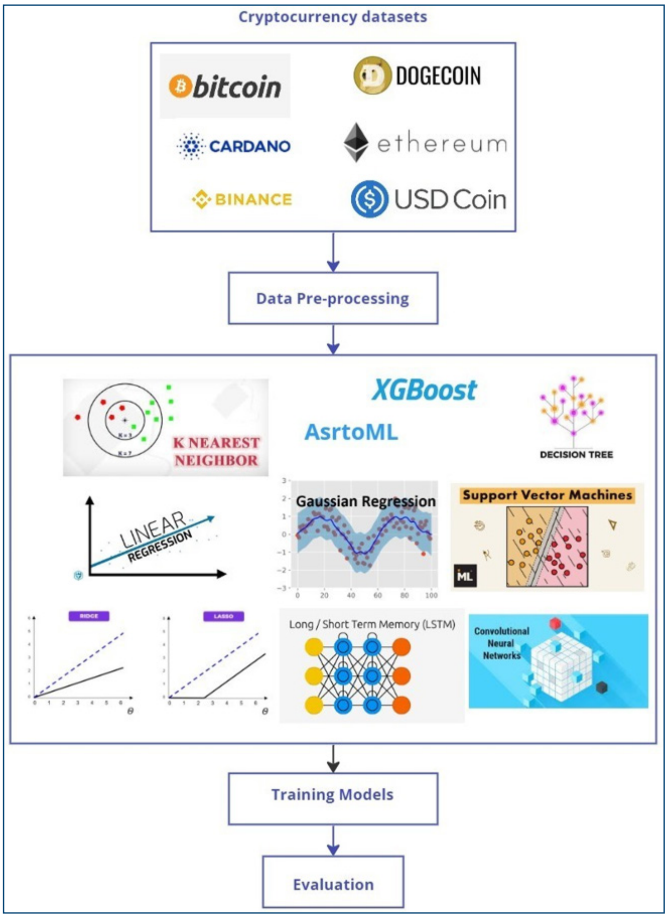

Figure 1. The methodology process

Cryptocurrency Datasets

Cryptocurrency datasets contain information about the various cryptocurrencies. The dataset contains 11 types of cryptocurrencies in USD and includes the prices of each type for the past 6 years.

In this study, we used six type cryptocurrencies datasets based on their availability of historic price data. These cryprocurrencies are Bitcoin, Dogecoin, Ethereum, USDCoin, Binance, and Cardano. **Table 1** lists the trade symbols for these cryptocurrencies.

Table 1: Symbol list of Cryptocurrencies datasets

Cryptocurrency	Symbole
Bitcoin	BTC
Dogecoin	D
Ethereum	ETH
USDCoin	USDC
Binance Coin	BNB
Cardano Coin	ADA

The dataset of this study was obtained from Dropbox. To read the dataset we use Pandas framework for Python programming language. These dataset converts a comma-separated values file into a DataFrame, which is a tabular structure that can handle different types of data and allows for iteration or chunking of the file. It contain seven features present as follows:

- date: represents the calendar day, month, and year of cryptocurrency prices.
- open: refers to the initial price at the start of the day.
- high: refers to the highest price reached during the day.
- low: refers to the lowest price reached during the day.
- close: refers to the final price at the end of the day.
- adj_close: refers to the average price of the day.
- volume: refers to the amount of cryptocurrency traded during the day

Data preprocessing is the process of preparing data for analysis and modeling. It involves a series of steps to clean, transform, and format the data so that it can be effectively used by ML algorithms. Some common steps in data preprocessing include, Data cleaning, Data transformation, Data formatting, and data reduction. In this paper, we use only data cleaning and data transformation. Then, the dataset was split into training (70%) and testing (30%) sets. We identified eleven models that are commonly used for prediction price. Each model has been built from scratch for each cryptocurrency. In order to enable a comparison between our models, we used the same model parameters across different cryptocurrency. In the following sections, we will discuss each model individually.

Predictive Methods

In the study, the dataset was divided into a training set and a testing set, with 70% and 30% of the data, respectively. The aim of the study was to predict cryptocurrency prices using eleven commonly used models. Each model was developed separately for each cryptocurrency, but the same parameters were used for each model to facilitate comparison between them. The study will discuss each model individually in subsequent sections.

Long Short-Term Memory Architecture

LSTM (Lahmiri & Bekiros, 2019) models are popular for cryptocurrency price prediction because they can memorize previous inputs and identify long-term dependencies in time series data. These models are trained on historical price data and other relevant variables to make predictions on future prices (Hu et al., 2015) The architecture of the LSTM model involves an input layer for historical price data and other variables, an LSTM layer to process the input data and learn dependencies, and an output layer to produce future price predictions. A loss function is used to evaluate the model's performance, and an optimization algorithm updates the model parameters to minimize the loss function. By following these steps, LSTM models can effectively predict cryptocurrency prices.

Convolutional Neural Network Architecture

A CNN (Hu et al., 2015) can be used to analyze cryptocurrency data to make informed trading decisions, as it is well-suited for image and time-series data. A CNN for cryptocurrency price prediction typically consists of an input layer for historical price data and relevant variables, convolutional layers to extract features and detect patterns, pooling layers to down-sample the output, fully connected layers to make predictions based on the features, and an output layer to produce the final prediction of the cryptocurrency price. The input data is preprocessed and normalized before being fed into the model.

Support Vector Machines Architecture

SVMs (Noble, 2006) use historical price data to predict future price movements by finding the best function that fits the data and predicts continuous values. This is done by transforming the data into a high-dimensional feature space, where a hyperplane can represent the prediction function. The optimization problem seeks to minimize the error between predicted and actual values while maximizing the margin, which is the distance between the hyperplane and the closest data points. This ensures a robust prediction function that is not overfit to the training data. To make predictions on new data, the new point is transformed into the feature space and plugged into the prediction function. The output of the function is the predicted value for the new data point.

K-nearest Neighbors Architecture

K-Nearest Neighbors (KNN) (Peterson, 2009) is a simple ML algorithm that can be used for cryptocurrency price prediction. It stores all available data points and corresponding continuous target values. When a new data point is encountered, the algorithm finds the K nearest neighbors based on a distance metric, and uses their target values to predict the new data point. The prediction is calculated as the average of the target values of the K nearest neighbors. While KNN regression is fast and simple, it can be sensitive to the choice of distance metric and number of neighbors used, and may not perform well for high-dimensional data. Careful parameter selection is necessary for good results.

XGBoost Architecture

XGBoost (Chen & Guestrin, 2016) is a ML algorithm that uses a series of weak models, such as decision trees, to improve prediction accuracy. It uses gradient descent to optimize the parameters of the models and iteratively improve their performance. The tree building process involves initializing residuals, splitting the data into subsets based on feature values, and selecting the split that results in the minimum loss. Tree pruning is used to remove branches that do not significantly contribute to prediction accuracy. Boosting is used to combine the predictions from multiple trees, with each tree being weighted based on its accuracy. The final prediction is obtained by summing the weighted predictions from all the trees, resulting in a more accurate overall prediction.

Astro ML Architecture

AstroML (VanderPlas et al., 2012) is a Python library designed for ML applications in astrophysics that can also be applied to cryptocurrency data. In this context, regression techniques from AstroML can be utilized to model the relationship between various features and the price of a cryptocurrency. The choice of regression model depends on factors such as the complexity of the data, the nature of the relationship between features and the target variable, and the desired level of accuracy. Regression techniques such as linear regression, polynomial regression, decision trees, and RF can be employed to make predictions about the future price of a cryptocurrency based on its past performance and other relevant factors.

LASSO Architecture

Lasso (Park & Casella, 2008) is a ML technique used to model the relationship between features and the price of a cryptocurrency. Its goal is to improve the interpretability and accuracy of the model by adding a penalty term to the loss function during optimization. This encourages the model to have sparse coefficients, reducing the importance of many features and resulting in a more parsimonious and interpretable model. Lasso regression selects a subset of relevant features to use in the prediction, making it easier to understand the relationship between the features and the target variable and potentially leading to more accurate predictions.

RIDGE Architecture

Ridge Regression (Hoerl & Kennard, 1970) is a regularization method used in regression analysis to model the relationship between features and the target variable, such as the price of a cryptocurrency. Its main goal is to improve the stability and interpretability of the model by adding a penalty term to the loss function that encourages smaller coefficients for all features.

The Ridge method achieves this by reducing the magnitude of the coefficients, making the model less sensitive to small variations in the data, resulting in a more stable and interpretable model. As a result, Ridge Regression can be useful for avoiding overfitting, reducing multicollinearity, and improving the accuracy of predictions by creating a more stable and interpretable model

Linear Regression Architecture

Linear Regression (LR) (Weisberg, 2005) is a statistical method used to model the relationship between a dependent variable and one or more independent variables. This method can be applied in the context of cryptocurrency to model the relationship between various features and the price of a cryptocurrency. The LR model assumes that the relationship between the dependent variable and independent variables can be modeled as a linear equation. The model estimates the coefficients for each feature by using historical data for each feature and the cryptocurrency price. The coefficients are then used to predict the future price of a cryptocurrency by multiplying the values of the independent variables by their corresponding coefficients and summing the results.

Decision Tree Architecture

A DT (Quinlan, 1986) algorithm is a ML tool that uses a tree-like structure to weigh the costs, probabilities, and benefits of different potential outcomes. Starting with a single node, the algorithm branches into possible outcomes, with each leading to additional nodes that branch off into other possibilities. Decision tree algorithms provide a systematic and mathematical approach to predicting the best choice by mapping out potential outcomes and the factors that influence them. These algorithms can be used for various purposes, from informal discussions to automating complex decision-making processes.

Gaussian Process Architecture

A Gaussian Process (GP) (Quiñonero et al., 2005) for cryptocurrency prediction models the relationship between predictors and the cryptocurrency price as a Gaussian distribution. This enables probabilistic predictions and quantification of uncertainty. A Gaussian prior distribution is assigned to the unknown function that maps predictors to the cryptocurrency price, and it is updated to a posterior distribution using observed data. The posterior distribution reflects the observed relationships and can be used to make predictions and estimate their uncertainty.

RESULTS

The article discusses the evaluation of different models for predicting cryptocurrency prices using RMSE values. XGBoost is the best performer with the lowest RMSE values for most cryptocurrencies. Ridge and GR models also demonstrate relatively low RMSE values and are good performers. However, other factors such as data availability, computational efficiency, and specific application need to be considered while selecting the best model.

The article discusses the evaluation of different models for predicting cryptocurrency prices. The performance of these models is evaluated using RMSE values. Lower RMSE values indicate better predictive accuracy of the model.

The XGBoost model is the best performer with the lowest RMSE values for most cryptocurrencies. For instance, its RMSE value of 0.014 for USD Coin is the lowest among all models. The Ridge and GR models also demonstrate relatively low RMSE values for most cryptocurrencies, indicating that they have good predictive accuracy as well.

Table 2:RMSE values of different models for predicting cryptocurrency prices

Model	Bitcoin	Dogecoin	Etherum	USD coin	Binance	Cardano coin
CNN	0.68	0.90	0.78	0.59	0.75	0.75
SVM	0.35	0.22	0.26	0.03	0.39	0.35
KNN	0.31	0.22	0.21	0.015	0.38	0.31
XGBoost	0.07	0.21	0.17	0.014	0.37	0.19
LASSO	0.42	0.22	0.33	0.032	0.39	0.36
Ridge	0.03	0.19	0.03	0.014	0.15	0.049
LR	0.02	0.03	0.02	0.015	0.032	0.03
Kernel regression	0.17	0.23	0.33	0.024	0.4	0.3
DT	0.32	0.22	0.2	0.02	0.37	0.22
GR	0.02	0.03	0.02	0.014	0.032	0.03
LSTM	0.26	0.22	0.08	0.01	0.25	0.059

However, other factors such as data availability, computational efficiency, and specific application need to be considered while selecting the best model. Therefore, based on the RMSE values provided in the table, the XGBoost and Ridge models seem to have better predictive accuracy than the other models.

The passage compares the RMSE results of various authors who applied different techniques to predict the price of different cryptocurrencies. Pabuçcu et al. experimented with various methods and found that the Random Forest (RF) technique had the best RMSE of 0.293, followed by ANN with 0.341, SVM with 0.438, and Naive Bayes with 0.461. Meanwhile, Jagannath et al. utilized LSTM to forecast the price of Bitcoin, but their RMSE of 1.9 was relatively higher than the other models. Awoke et al. improved on this method by applying both LSTM and GRU techniques, achieving lower RMSE values of 0.045 for LSTM and 0.051 for GRU. Kim et al. focused on predicting the price of Ethereum using ANN and SVM. They obtained an RMSE of 0.068 and 0.048, respectively, indicating that SVM performed better in forecasting Ethereum prices.

Authors	Cryptocurrency	Techniques	RMSE
Pabuçcu et al	Bitcoin	ANN	0.341
Pabuçcu et al	Bitcoin	SVM	0.438
Pabuçcu et al	Bitcoin	Naïve Bayes	0.461
Pabuçcu et al	Bitcoin	RF	0.293
Jagannath et al	Bitcoin	LSTM	1.9
Awoke et al	Bitcoin	LSTM	0.045
Awoke et al	Bitcoin	GRU	0.051
(Kim et al., 2021b)	Ethereum	ANN	0.068
(Kim et al., 2021b)	Ethereum	SVM	0.048
This research	Bitcoin	Nadaraya–Watson kernel regression	0.17
This research	Dogecoin	Linear regression	0.032
This research	Ethereum	Gaussian regression	0.02
This research	USD coin	XGboost, Gaussian regression, Ridge	0.014
This research	Binance coin	Gaussian regression	0.032
This research	Cardano coin	LSTM	0.059

The proposed model achieved lower RMSE values for different cryptocurrencies using various techniques. The Nadaraya-Watson kernel regression technique provided an RMSE of 0.17 for Bitcoin, Linear Regression provided an RMSE of 0.032 for Dogecoin, Gaussian regression provided an RMSE of 0.02 for Ethereum, XGboost, Gaussian regression, and Ridge achieved an RMSE of 0.014 for USD coin, Gaussian regression provided an RMSE of 0.032 for Binance coin, and LSTM achieved an RMSE of 0.05 for Cardano coin.

DISCUSSION

The study results showcased the efficacy of various machine learning techniques in predicting cryptocurrency prices, with some simpler models outperforming complex ones. These findings align with existing literature, emphasizing the importance of model selection and the trade-off between complexity and performance. The proposed model demonstrated promising results based on the RMSE metric, highlighting its potential for practical applications in the financial industry.

Study Limitations: One limitation of this study is the reliance on historical price data, which may not fully account for unforeseen events or market fluctuations.

Additionally, the study focused on a limited number of cryptocurrencies and did not consider other potential factors such as market sentiment or regulatory changes. The choice of machine learning techniques could also be expanded to include more recent advancements or other ensemble methods.

CONCLUSION

This study examines the effectiveness of time-series prediction models for cryptocurrency prices, as machine learning and blockchain technology have gained attention in recent years. The research found that complex models for anticipating cryptocurrency prices may not perform better than a simple persistence model. The proposed model in the study had better performance on average, as evaluated by the RMSE. Future work could involve testing the models on different time frames or comparing their performance on different subsets of the data, and exploring other prediction models such as deep learning, recurrent neural networks, or reinforcement learning. There are many opportunities for future research in the field of cryptocurrency price prediction using ML.

ETHICAL DECLARATIONS

Referee Evaluation Process: Externally peer-reviewed.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

Author Contributions: All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
2. Gowda, N., & Chakravorty, C. (2021). Comparative study on cryptocurrency transaction and banking transaction. Global Transitions Proceedings, 2(2), 530–534. https://doi.org/10.1016/j.gltp.2021.08.064
3. Gundlach, R., Gijsbers, M., Koops, D., & Resing, J. (2021). Predicting confirmation times of Bitcoin transactions. ACM SIGMETRICS Performance Evaluation Review, 48(4), 16–19. https://doi.org/10.1145/3466826.3466833
4. Hitam, N. A., & Ismail, A. R. (2018). Comparative Performance of Machine Learning Algorithms for Cryptocurrency Forecasting. Indonesian Journal of Electrical Engineering and Computer Science, 11(3), 1121. https://doi.org/10.11591/ijeecs.v11.i3.pp1121-1128
5. Ho, A., Vatambeti, R., & Ravichandran, S. K. (2021). Bitcoin Price Prediction Using Machine Learning and Artificial Neural Network Model. Indian Journal of Science and Technology, 14(27), 2300–2308. https://doi.org/10.17485/IJST/v14i27.878
6. Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, 12(1), 69–82. https://doi.org/10.1080/00401706.1970.10488635
7. Hu, B., Lu, Z., Li, H., & Chen, Q. (2015). Convolutional Neural Network Architectures for Matching Natural Language Sentences. http://arxiv.org/abs/1503.03244
8. Jagannath, N., Barbulescu, T., Sallam, K. M., Elgendi, I., Okon, A. A., Mcgrath, B., Jamalipour, A., & Munasinghe, K. (2021). A Self-Adaptive Deep Learning-Based Algorithm for Predictive Analysis of Bitcoin Price. IEEE Access, 9, 34054–34066. https://doi.org/10.1109/ACCESS.2021.3061002
9. Kim, H.-M., Bock, G.-W., & Lee, G. (2021a). Predicting Ethereum prices with machine learning based on Blockchain information. Expert Systems with Applications, 184, 115480. https://doi.org/10.1016/j.eswa.2021.115480
10. Kim, H.-M., Bock, G.-W., & Lee, G. (2021b). Predicting Ethereum prices with machine learning based on Blockchain information. Expert Systems with Applications, 184, 115480. https://doi.org/10.1016/j.eswa.2021.115480
11. Koops, D. (2018). Predicting the confirmation time of Bitcoin transactions. http://arxiv.org/abs/1809.10596
12. Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons & Fractals, 118, 35–40. https://doi.org/10.1016/j.chaos.2018.11.014
13. Noble, W. S. (2006). What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565
14. Pabuçcu, H., Ongan, S., & Ongan, A. (2020). Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. Quantitative Finance and Economics, 4(4), 679–692. https://doi.org/10.3934/QFE.2020031
15. Park, T., & Casella, G. (2008). The Bayesian Lasso. Journal of the American Statistical Association, 103(482), 681–686. https://doi.org/10.1198/016214508000000337
16. Peterson, L. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883. https://doi.org/10.4249/scholarpedia.1883
17. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251
18. Quiñonero, J., Quiñonero-Candela, Q., Rasmussen, C. E., & De, C. M. (2005). A Unifying View of Sparse Approximate Gaussian Process Regression. In Journal of Machine Learning Research (Vol. 6).
19. Rahouti, M., Xiong, K., & Ghani, N. (2018). Bitcoin Concepts, Threats, and Machine-Learning Security Solutions. IEEE Access, 6, 67189–67205. https://doi.org/10.1109/ACCESS.2018.2874539
20. VanderPlas, J., Connolly, A. J., Ivezic, Z., & Gray, A. (2012). Introduction to astroML: Machine learning for astrophysics. 2012 Conference on Intelligent Data Understanding, 47–54. https://doi.org/10.1109/CIDU.2012.6382200
21. Weisberg, S. (2005). Applied Linear Regression. John Wiley & Sons, Inc. https://doi.org/10.1002/0471704091
22. Yogeshwaran, S., Kaur, M. J., & Maheshwari, P. (2019). Project Based Learning: Predicting Bitcoin Prices using Deep Learning. 2019 IEEE Global Engineering Education Conference (EDUCON), 1449–1454. https://doi.org/10.1109/EDUCON.2019.8725091

Muhammad Ali Mohammed

My name is Muhammad Ali Mohammed. I am doing my master's degree in Çankırı Karatekin University Computer Engineering Department. I am working on machine learning and cryptocurrencies.