Journal of
Computer & Electrical and Electronics Engineering Sciences
Original Article
DOI: 10.51271/JCEEES-0007

Cryptocurrency analysis using machine learning and deep learning approaches

Farah Muhammed Sakran

Department of Electronic and Computer Engineering, Faculty of Engineering, Çankırı Karatekin University, Çankırı, Turkey

Cite this article: Muhammed Sakran F. Cryptocurrency analysis using machine learning and deep learning approaches. *J Comp Electr Electron Eng Sci.* 2023;1(2):29-33.

Corresponding Author: Farah Muhammed Sakran, farahalameri338@gmail.com

Received: 18/07/2023 ◆ **Accepted**: 25/09/2023 ◆ **Published**: 31/10/2023

ABSTRACT

Since cryptocurrencies are becoming more widely used and accepted in the financial system, precise price forecasting is essential for optimizing bitcoin investments. In this research study, we evaluated various machine learning models, including linear regression (LR), decision tree regression (DT), random forest regression (RF), support vector regression (SVR), gradient boosting regression (GB), adaboost regression, extreme gradient boosting regression (XGR), light gradientboosting regression (LGBM), k-nearest neighbors regression (KNN), ridge, andlasso. Additionally, we incorporated two deep learning (DL) models, namely artificial neural networks (ANN) and convolutional neural networks (CNN), to forecast daily bitcoin prices (BP). The initial data was obtained from Kaggle, a well-known platform for data science projects, and we applied the min-max scaler technique for consistent scaling during preprocessing. To assess the predictive capabilities of the models, we utilized regression metrics such as root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R). Based on our findings, the CNN model demonstrated the highest effectiveness in predicting BPs among the DL models, with an RMSE of 0.0543, MAE of 0.0324, and an R value of 0.960. In the case of machine learning models, the RF model outperformed others, achieving an RMSE of 0.0246 and MAE of 0.0561.

Investors, scholars, and decision-makers may all gain from these findings' insightful revelations about BP forecasting. Developing these models further, investigating different preprocessing methods, and expanding the analysis to other cryptocurrencies might be the main goals of future research.

Keywords: Blockchain, cryptocurrency, machine learning

INTRODUCTION

Within the context of this study, previous research provides several insights that can be expanded upon (Sureshbhai, 2020). Due to the complexity of bitcoin systems, individuals often have misconceptions about their technological functioning, making them challenging to comprehend. Previous studies have extensively addressed the issue of key management as a significant problem for users (Sureshbhai, 2020). Custodial wallets, which eliminate the need for users to worry about key management, offer a means to interact with cryptocurrencies but require trust in the intermediary. However, limited information is available regarding the challenges faced by clients using custodial wallets (Mell, 2017). A study focusing on novice users' perceptions of Bitcoin's usability sheds light on the necessity for further research in this area (Begum, 2020). This study also examines the difficulties encountered by new bitcoin users and proposes potential solutions (Begum, 2020). The current article draws upon various findings from prior research. The complexity of cryptocurrencies often leads to discrepancies between users' mental models and the actual technological processes. Key management has primarily been addressed in previous research, highlighting its complexity for users (Wood, 2016). Custodial wallets, which alleviate key management concerns, are commonly used, yet there is limited research on the issues they pose for users. The first study (Financial Platform and News Website, 2008) investigating novice users' perceptions of Bitcoin's usability emphasizes the need for additional research. This work (Business Insider India, 2008) focuses on the challenges faced by new Bitcoin users and provides solutions to address them, addressing these unresolved issues.

Digital currency's roots trace back to Chaum's untraceable payment mechanism from the 1980s, along with blind signature technology (Mikhaylov, 2019). The 1990s saw various developments in digital money payments, such as fair offline e-cash and untraceable offline cash (Mikhaylov, 2019). Still, these methods depended on trusted parties to prevent double-spending attacks. Strategies like B-Money (Chuen, 2017) and Bit Gold (Chuen, 2017) later aimed to remove these intermediaries, but the adoption of decentralized consensus faced significant hurdles. A breakthrough came with Hal Finney's "Reusable Proofs of Work" in 2004 (Corbet, 2021), using trusted computing as a backend. More recent efforts involve forecasting cryptocurrency price movements, like

Bitcoin prices (Aggarwal, 2019). Rather than handling all cryptocurrency data collectively, the proposed approach analyzes each cryptocurrency's data separately. It employs ANN, CNN, and other ML models to leverage their unique advantages, aiming to create a comprehensive, accurate method for Bitcoin price prediction.

Literature Review

Numerous studies have applied diverse machine learning (ML) and deep learning (DL) algorithms to predict Bitcoin prices and identify influencing factors. Notably, LSTM models have shown strong performance (Jiang, 2019; Chen, 2020). For example, a 2019 study (Jiang, 2019) reported an RMSE of 47.91, while a 2021 study (Carbó and Gorjón, 2022) proposed an ensemble LSTM model yielding an RMSE of 37.24. However, LSTM models have shown some limitations, such as failing to identify a positive link between gold prices and Bitcoin prices (Jiang, 2019). Other research has utilized advanced hybrid models such as WT-CATCN (Saadah, 2020), achieving a 25% improvement in accuracy with an RMSE of 19.020, and a combined GRU and 1DCNN model (Wardak and Rasheed, 2022) that outperformed alternatives with an RMSE of 43.933, 3.511, and 0.00128. Twitter sentiment analysis was also found to correlate positively with Bitcoin price (Jiang, 2019). Another study (Chen, 2020) achieved 95.7% accuracy and a 0.05 RMSE using LSTM. Furthermore, research has shown that various ML algorithms including decision tree and regression models (Rathan, 2019), XGBoost and SDA (Borges and Neves, 2020), as well as ensemble learning (Mallqui and Fernandes, 2018), can be effective in predicting Bitcoin prices. These studies underline the potential of ML and DL models in cryptocurrency forecasting, with room for further exploration and optimization.

METHODS

In this research study, we aimed to optimize Bitcoin investments by developing a precise price forecasting model with existing artificial intelligence algorithms. We evaluated a range of ML models. To incorporate DL techniques, we also utilized ANN and CNN. The initial dataset was obtained from Kaggle, a renowned platform for data science projects, and we applied the min-max scaler technique for consistent scaling during the data preprocessing stage. Throughout the study, we assessed the predictive capabilities of these models using regression metrics. The model used a multi-stage methodology, which included preparing the data to standardize it for use with ML/DL algorithms. It made use of a variety of ML/DL approaches to build a forecasting framework that was precise and tuned for the dataset's specific problems and characteristics.

Datasets Description

The dataset utilized in this research study provides valuable information on the price movements and trading activity of various cryptocurrencies. This rich dataset allows us to create reliable prediction models by examining the potential relationships or correlations between different digital assets. By combining multiple cryptocurrencies, we gain a deeper understanding of the Bitcoin market, adding complexity and generating a more comprehensive picture. To ensure the dataset is suitable for analysis, appropriate preprocessing procedures are implemented. One crucial

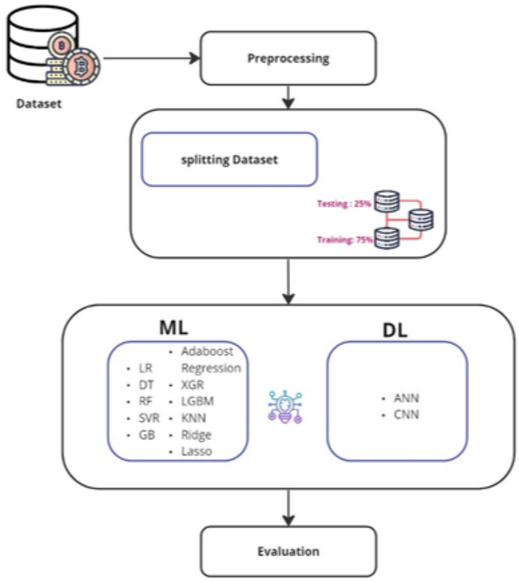

Figure 1. Proposed model

preprocessing technique employed is the min-max scaler. This technique scales the data to a specific range, typically between 0 and 1. By normalizing the data, it eliminates biases caused by variations in magnitudes between different features, enabling fair comparisons. The application of the min-max scaler ensures that the dataset is uniformized and ready for the subsequent modeling phase. The dataset used in this study presents a diverse and intriguing collection of historical price and volume data for four significant cryptocurrencies. By incorporating multiple digital assets and employing preprocessing techniques, we lay the foundation for developing precise and trustworthy prediction models for BP forecasting. This approach allows us to capture the complexities of the cryptocurrency market and enhance the ACC of our forecasting models.

	Date	Adj Close (BNB)	Volume (BNB)	Adj Close (BTC)	Volume (BTC)	Adj Close (USDT)	Volume (USDT)	Adj Close (ETH)	Volume (ETH)
0	09/11/2017	1.99077	19192200	7143.580078	3226249984	1.00818	358188000	320.884003	893249984
1	10/11/2017	1.79684	11155000	6618.140137	5208249856	1.00601	756446016	299.252991	885985984
2	11/11/2017	1.67047	8178150	6357.600098	4908680192	1.00899	746227968	314.681000	842300992
3	12/11/2017	1.51969	15298700	5950.069824	8957349888	1.01247	1466060032	307.907990	1613479936
4	13/11/2017	1.68662	12238800	6559.490234	6263249920	1.00935	767884032	316.716003	1041889984
...	...	...	...	...	...	...	...	...	...
1748	23/08/2022	299.03000	1034959891	21528.090000	31878280659	1.00000	52978066914	1662.770000	18322041914
1749	24/08/2022	296.45000	935405757	21395.020000	31962253368	1.00000	48708536003	1657.060000	16780932907
1750	25/08/2022	301.58000	973529233	21600.900000	31028679593	1.00010	45854093190	1696.460000	14818795695
1751	26/08/2022	279.60000	1210077085	20260.020000	42326789564	1.00000	62406193046	1507.780000	26713710143
1752	27/08/2022	277.30000	1023169984	20029.790000	32739299328	1.00000	48439021568	1470.760000	20430931968

Figure 2. BitCoin dataset

The next step in the data preparation process is to split the dataset into training and testing sets. This division is crucial to evaluate the effectiveness and generalization capabilities of the prediction models. In this study, a 25:75 split ratio is employed, where 25% of the dataset is allocated to testing and the remaining 75% is used for training the models. By splitting the data in this manner, a significant portion is dedicated to training the models, allowing them to learn patterns and relationships from a substantial amount of historical data. This enhances the models' ability to make accurate predictions. The testing set serves as an independent dataset to assess how well the models perform when applied to new data. It provides an unbiased evaluation of the models' prediction capabilities and their adaptability to novel situations. The 25:75 split ratio strikes a balance between having sufficient training data to adequately train the models and reserving a reasonable amount for testing. This ensures that the models can generalize effectively to new scenarios without being overfitted to the training data.

During the training phase, the models learn from the training set and adjust their internal parameters to minimize prediction errors. Their performance is then evaluated by comparing their predictions to the actual values using the testing set. This assessment process helps determine the precision, robustness, and applicability of the models in real-world scenarios. Moving on to preprocessing, it is the subsequent step in the data preparation procedure. Preprocessing involves transforming the data into a standardized format suitable for prediction models. One commonly used preprocessing technique is the min-max scaler, which normalizes the data. The min-max scaler rescales the values within a specific range, typically between 0 and 1. This normalization ensures that all features are on an equal scale, preventing any single feature from dominating the model's learning process. By applying the min-max scaler, the values of the dataset are proportionally adjusted to fit within the defined range. This process ensures consistent scaling across all features while preserving the relative relationships between data points. Normalized data enables the models to effectively learn from the data and generate accurate predictions. Additionally, the min-max scaler can handle outliers and extreme values in the dataset. By compressing the data into a narrow range, the influence of outliers is reduced, preventing them from negatively impacting the model's performance. The preprocessing phase involving the min-max scaler is crucial for preparing the data before feeding it into the prediction models. It allows for efficient normalization of the dataset, enabling the models to learn from patterns and correlations in the data, leading to more accurate and reliable predictions.

Predictive Methods

The next stage in our methodology involves applying various ML and DL models to forecast daily Bitcoin prices. We evaluate a range of ML models, including LR, DT, RF, SVR, GB, Adaboost Regression, XGR, LGBM, KNN, Ridge, and Lasso. Additionally, we incorporate two DL models, namely ANN and CNN. Linear Regression is a simple regression method used to model the relationship between a dependent variable and one or more independent variables through a linear equation. Decision tree regression represents data using a tree structure and makes predictions of the dependent variable's value by traversing the tree based on input features. Random forest regression improves prediction accuracy by using multiple decision trees and reducing overfitting. Support vector regression uses support vector machines to model linear or nonlinear relationships between data points. Gradient boosting regression enhances prediction accuracy by combining weak learners (usually decision trees) to make more accurate predictions. Adaboost regression is an ensemble method that boosts the performance of weak regression models by combining them. XGBoost is an optimized version of gradient boosting that offers faster and more efficient learning. LightGBM is a gradient boosting method optimized for speed and distributed learning. K-Nearest neighbors regression predicts values based on the average of the nearest data points in the feature space. Ridge regression is a linear regression method that adds L2 regularization to resist overfitting. Lasso Regression adds L1 regularization to linear regression, reducing unnecessary features and increasing model simplicity. Artificial neural networks are deep learning methods that model complex relationships inspired by biological neural systems. Convolutional neural networks are a type of artificial neural network primarily used for image processing and pattern recognition, employing convolution operations on input data. These models are selected based on their specific capabilities and suitability for addressing the challenges and characteristics of the dataset. By leveraging a diverse range of ML and DL techniques, we aim to optimize the ACC and predictive capabilities of our forecasting model for Bitcoin prices. This set of metrics evaluates the performance of encryption algorithms. These metrics include MSE, which measures the square root of the average "error" between the original and encrypted pixel values in images. The MSE equation (1) calculates the discrepancy for each pixel and averages them across the entire image. Another metric is RMSE, which estimates the size of the error between the expected and actual values. The RMSE equation (2) takes the square root of the MSE to provide a more interpretable measure of the error.

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\ [f(i.j)-f^{`}(i.j)]^2}{MN} \quad (1)$$

$$RMSE(\hat{\theta}) = \sqrt{MSE\ (\hat{\theta})} \quad (2)$$

RESULTS

Utilizing the approach, we addressed the study goals mentioned in the first chapter in this work. After providing a general overview of the Cryptocurrency Prices dataset, the needs of the study were thoroughly explained. The experiment's findings were then carefully examined and evaluated. After comparing several ML regressors, it appears that the RF provides the best performance with an MAE of 0.024609, RMSE of 0.056137, and R2 of 0.958287. The Random Forest Regressor therefore seems to be the most accurate model for this task, offering the lowest loss. In addition, it's worth mentioning that DT and XGB also achieved commendable results with relatively low MAE and RMSE, and high R2 scores. This indicates a strong fit to the data and an ability to predict Bitcoin prices reliably. On the other end of the spectrum, the Lasso regression model did not perform well in comparison to other models, with a significantly higher MAE and RMSE, and a negative R2 score, which indicates a poor fit to the data. These results demonstrate the power of ensemble learning methods like RF and XGBoost in predicting volatile cryptocurrency prices such as Bitcoin, and underscores the importance of choosing the right model and tuning it correctly for predictive tasks in financial contexts.

Table 1. Loss values of different models for predicting cryptocurrency prices

Regressor	MAE	RMSE	R2
Linear Regression	0.065311	0.107985	0.845651
Decision Tree Regressor	0.026540	0.069409	0.936232
Random Forest Regressor	**0.024609**	**0.056137**	**0.958287**
SVR	0.062093	0.083781	0.907090
Gradient Boosting Regressor	0.035017	0.066422	0.941602
AdaBoost Regressor	0.055780	0.080166	0.914934
XGB	0.026591	0.058331	0.954962
LGBM	0.029611	0.060188	0.952049
KNN	0.028625	0.059169	0.953658
Ridge	0.065359	0.107970	0.845694
Lasso	0.233261	0.274884	-0.000177

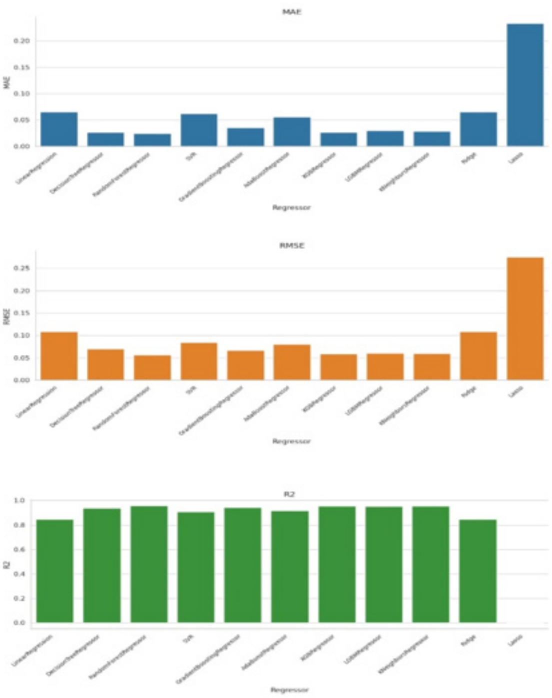

Figure 3. ML comparison

The results provided are evaluation metrics of two different models, ANN and CNN which were presumably used for predicting Bitcoin prices. Starting with the ANN model, it obtained an RMSE of 0.09349, an MAE of 0.0612, and a correlation coefficient of 0.8842. These figures suggest a decent model performance, but there seems to be room for improvement. In comparison, the CNN model outperforms the ANN model on all three metrics. It achieved an RMSE of 0.05437, which is considerably lower than the ANN model's RMSE, suggesting a more accurate prediction. Similarly, the MAE for the CNN model is 0.03247, also significantly lower than the ANN model's MAE, suggesting less absolute error in the CNN model's predictions. The correlation coefficient of the CNN model is 0.960, closer to 1 than the ANN model's correlation coefficient. This suggests that the CNN model's predictions are more strongly correlated with the actual values.

Table 2. Comparative results ML/ DL

Model (Type)	MAE	RMSE	R
Random Forest (ML)	**0.024609**	**0.056137**	**0.958287**
ANN (DL)	0.061200	0.093494	0.884297
CNN (DL)	0.032474	0.054376	0.960863

The graph depicted in Figure 4 compares the true values versus the predicted values of Bitcoin prices as predicted by the DL models, namely ANN and CNN. The figure shows two distinct curves, the red curve representing the predicted values from the models and the blue curve indicating the actual or true Bitcoin prices. From the visualization, it is evident that the red curve, which signifies the predicted values, consistently lies above the blue curve, indicating the actual values. This pattern reveals that both the ANN and CNN models, on average, predicted a higher Bitcoin price than the actual observed price. However, the fact that the predicted curve closely follows the trend of the actual curve indicates that these models have done a commendable job

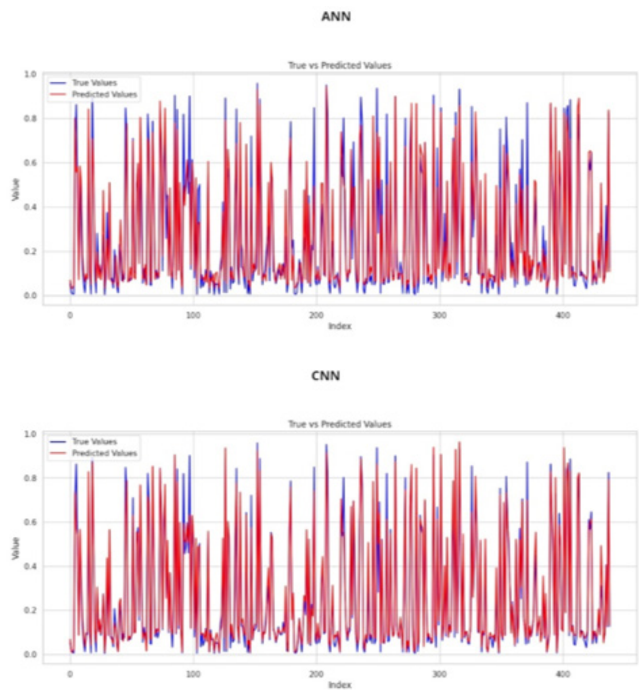

Figure 4. DL true vs prediction

in capturing the general pattern and movements of Bitcoin prices. The models were able to predict the direction of price changes accurately, even if the predicted prices were slightly higher than the actual prices.

DISCUSSION

In this study, traditional ML and DL models were used to predict Bitcoin prices. The Random Forest (RF) emerged as the best ML model, exhibiting robust performance and accuracy. However, CNN, a DL model, outperformed all with its ability to capture complex time-series patterns. Despite these promising results, there were limitations. The models were restrained by the historic data, which might not reflect future events or trends. Also, important external factors affecting Bitcoin prices weren't considered. The performance of DL models could be enhanced with extensive hyperparameter tuning, not fully explored in this study due to computational limits. Future work should focus on model optimization, exploring diverse neural network architectures, and employing larger datasets for better prediction accuracy.

CONCLUSION

In summary, this study highlighted the effectiveness of various machine learning and DL models for Bitcoin price prediction. Among traditional machine learning models, random forest regressor stood out, while in the realm of DL, the CNN model outperformed the ANN model. The results underscore the potential of advanced analytics and predictive modeling techniques in navigating the complex and volatile landscape of cryptocurrency markets. However, a degree of caution is needed when interpreting the results due to the limitations of the study. Future research should focus on addressing these limitations, incorporating more external factors that affect cryptocurrency prices, and refining the models to better capture the nuances of cryptocurrency price fluctuations. The pursuit of more accurate and reliable predictive models for Bitcoin prices remains a challenging yet important task for researchers, investors, and policy makers alike.

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. Sureshbhai, P. N., Bhattacharya, P., and Tanwar, S. (2020 June). 'KaRuNa: A blockchainbased sentiment analysis framework for fraud cryptocurrency schemes. In 2020 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1-6). IEEE.
2. Mell, P., Kelsey, J., & Shook, J. (2017). Cryptocurrency smart contracts for distributed consensus of public randomness. In Stabilization, Safety, and Security of Distributed Systems: 19th International Symposium, SSS 2017, Boston, MA, USA, November 5-8, 2017, Proceedings 19 (pp. 410-425). Springer International Publishing.
3. Begum, A., Tareq, A., Sultana, M., Sohel, M., Rahman, T., & Sarwar, A. (2020). Blockchain attacks analysis and a model to solve double spending attack. *International Journal of Machine Learning and Computing*, 10(2),352-357.
4. G. Wood. Ethereum: A secure Decentralised Generalised Transaction Ledger. Accessed: 2016. [Online]. Available: https://scholar.google.Com/scholar?hl=en&as_sdt=0%2C5&q=ETHEREUM%
5. Financial Platform and News Website. Accessed: 2008. [Online]. Available: https://www.investing.com/
6. Business Insider India. Accessed: 2008. [Online]. Available: https://www. businessinsider.in/
7. A. Mikhaylov. Asset Allocation in Equity, Fixed-Income and Cryptocurrency on the Base of Individual Risk Sentiment. Accessed: 2019. [Online]. Available: https://pdfs.semanticscholar.org/df78/f60a84c2a17f47bce27578746c6313251%a88.pdf
8. D. L. K. Chuen, L. Guo, and Y. Wang, "Cryptocurrency: A new investment opportunity?" *J. Alternative Investments,* vol. 20, no. 3, pp. 16-40, 2017.
9. S. Corbet, Y. Hou, Y. Hu, C. Larkin, B. Lucey, and L. Oxley, "Cryptocurrency liquidity and volatility interrelationships during the COVID-19 pandemic," Finance Res. Lett., early access, May 2021, Art. no. 102137.
10. A. Aggarwal, I. Gupta, N. Garg, and A. Goel, "Deep learning approach to determine the impact of socio-economic factors on BP prediction," in 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. 1-5, 2019.
11. X. Jiang, "Bitcoin price prediction based on deep learning methods," *J. Math. Finance.*, Vol. 10, pp. 132-139, 2019.
12. Saadah, S.; Whafa, A.A. Monitoring Financial Stability Based on Prediction of Cryptocurrencies Price Using Intelligent Algorithm. In Proceedings of the 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 5-6 August 2020; pp. 1-10.
13. Jay, P., Kalariya, V., Parmar, P., Tanwar, S., Kumar, N., Alazab, M. Stochastic neural networks for cryptocurrency price prediction. IEEE Access 2020, 8, 82804–82818.
14. A. Politis, K. Doka, and N. Koziris, "Ether price prediction using advanced deep learning models,". In Proceedings of the 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Sydney, Australia, 3-6 May 2021; pp. 1-3.
15. S. MyungJae, D. Mohaisen, and J. Kim, "Bitcoin Price Forecasting via Ensemble-based LSTM Deep Learning Networks". Paper presented at 2021 International Conference on Information Networking (ICOIN), Jeju Island, Republic of Korea, January 13-16; pp. 603-608, 2021.
16. J. Carbó and S. Gorjón, "Application of Machine Learning Models and Interpretability Techniques to Identify the Determinants of the Price of Bitcoin," Banco de Espana Working Paper No. 2215, 2022.
17. C. Kang, C. Lee, K. Lim, "Cryptocurrency Price Prediction with Convolutional Neural Network and Stacked Gated Recurrent Unit," Data, Vol. 7, No. 149, 2022.
18. A. Wardak and J. Rasheed. (2022). "Bitcoin Cryptocurrency Price Prediction Using Long Short-Term Memory Recurrent Neural Network," *European Journal of Science and Technology,* No: 38, pp. 47-53.
19. Chen, Z., Li, C., and Sun. W. (2022). "Journal of Computational and Applied Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, p. 112395, doi: 10.1016/j.cam.2019.112395.
20. Rathan, K., Sai, S.V., and Manikanta, T.S. (2019). "CryptoCurrency price prediction using Decision Tree and Regression techniques," 3rd *Int. Conf. Trends Electron. Informatics*, no. Icoei, pp. 190-194.
21. Borges, T. A., and Neves, R.F. (2020). "Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods," *Appl. Soft Comput. J.,* vol. 90, p. 106187, doi:10.1016/j.asoc.2020.106187.
22. Mallqui, D.C.A., and Fernandes, R.A.S. (2018). "Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques," *Appl. Soft Comput. J.* doi: 10.1016/j.asoc.2018.11.038.

Farah Muhammed Sakran

Farah Sakran is doing her master's degree in Electronic-Computer Engineering Department of Çankırı Karatekin University, Institute of Science.