J. Electrical Systems 20-10s (2024):1964-1971

[1]Sneha Cenoy
[2] Kaveri Pathak
[3] Maneesha

Predictive Analytics for Cryptocurrency Prices Using Machine Learning

Journal of Electrical Systems

Abstract: - This paper explores the use of big data in decision-making within the digital economy, with a specific focus on predictive analytics for Bitcoin price. Machine learning models are used as they are crucial for predicting Bitcoin prices due to their ability to analyze complex, volatile market data and automate trading decisions. These models enhance risk management and market efficiency by adapting to new trends and performing real-time analysis It compares various machine learning models including Support Vector Machine (SVM), Decision tree regression, Random Forest, Long Short-Term Memory (LSTM), ARIMA, Linear Regression, and Quadratic Discriminant Analysis (Quaddisc). Python with Scikit-learn module and Pandas package along with others are used to assess the predictive analysis of the dataset. The study finds that Support Vector Machine (SVM) outperforms other models with an accuracy of 87%, compared to 64.37% for quadratic discriminant analysis and 52.9% for Random Forest. These results underscore SVM's superior efficiency in predictive analytics of Bitcoin prices. The paper also emphasizes the importance of fostering a data-driven culture within organizations to leverage big data analytics effectively, enhancing decision-making processes through comparative studies of these models.

Keywords: Predictive analytics, Support Vector Machine, Decision tree regression, Random Forest, Python.

I. INTRODUCTION

Bitcoin

Bitcoin, a virtual credit that is created by a technology that is hard to copy named 'cryptocurrency', has gained popularity in recent years, from a hobby that only few can understand to a world innovation. The fact that the system is on a decentralized base built outside of the traditional financial system is attractive in the eyes of people that value a harsh line of autonomy over their financial decisions. On top of that, the high prospect of strong earnings has become the cause for plenty of capital shine and speculators alike to take an interest in the digital currency which therefore has led to widespread use and media attention. As Bitcoin got popular, there was more demand for data analysis tools in order to try and interpret its price and therefore make Bitcoin investment related decisions.

Dataset

The dataset used in this study provides comprehensive data on Bitcoin price movements over the past eight years, including opening, closing, highest, and lowest prices, as well as trading volume. It incorporates various technical indicators such as RSI, CCI, SMAs (e.g., 50-day, 100-day), MACD, Bollinger Bands, and True Range, aiding in trend analysis and volatility measurement. The primary objective is to predict Bitcoin's closing rate for the next day, enabling forecasts of future market movements.

Python Modules

The cryptocurrency prediction utilizes Python libraries like Pandas for data preprocessing and feature engineering, NumPy for numerical computations and array manipulations, and Statsmodels for statistical analysis. Matplotlib and Seaborn are employed for data visualization, while Scikit-learn provides a wide range of machine learning algorithms for model construction and optimization. These libraries enable the creation of predictive models based on Bitcoin prices, incorporating technical indicators and aiding in better understanding of the cryptocurrency market dynamics for informed decision-making by investors

II. MATERIALS AND METHOD

This study's methodology is using diverse machine learning techniques, to fulfill predicted objective tasks, especially the Bitcoin price prediction and classification in that domain. The predictive analysis of the dataset is

[1] Department of Computer Science, Birla Institute of Technology Dubai. f20220134@dubai.bits-pilani.ac.in
[2] Department of Computer Science, Birla Institute of Technology Dubai. f20220094@dubai.bits-pilani.ac.in
[3] Department of Computer Science, Birla Institute of Technology Dubai. maneesha@dubai.bits-pilani.ac.in

made possible with the use of Support Vector Machines (SVM), Autoregressive Integrated Moving Average (ARIMA) Long Short-Term Memory (LSTM) networks, Quadratic discriminant analysis, Decision Tree regression, Linear Regression, and Random Forest.

SVM

SVMs are a type of supervised machine learning algorithm that analyze raw data to identify patterns and make predictions, such as categorizing data into two groups or estimating prices. They use specific metrics, like Bitcoin price indicators (open, close, low, high, true range), to distinguish relevant factors. After feature selection, the dataset is split into training and testing sets for model development and evaluation. Regression metrics like recall and f1-score assess the model's performance. Optimized SVM models are then deployed to generate predictions or update based on monitoring systems, aiming to improve understanding of price movements and market dynamics in digital assets.

ARIMA

ARIMA, a powerful tool for sequential data analysis and prediction, combines three key elements: Autoregressive (AR) models for trend, Integrated Moving Averages (I) for seasonality, and Moving Averages (MA) for irregularities. By considering lags, differencing, and assessing residual errors, ARIMA models offer flexibility in modeling and forecasting time-varying phenomena. Our ARIMA implementation involves parameterization, model training on historical data, and forecasting future time series points. This approach aims to gain insights into dataset behavior and provide informed forecasts to support decision-making processes.

LONG SHORT TERM MEMORY

LSTM, a type of RNN, is adept at capturing long-term dependencies in time series data. Unlike ARIMA, it doesn't require explicit feature engineering. With memory cells and gating mechanisms, LSTM retains crucial information for short and long-term predictions. We train the model on historical data to forecast future outcomes, using deep learning to uncover intricate temporal relationships and improve prediction accuracy.

The Mean Absolute Error (MAE) is an estimated measure employed to appraise the size of average errors between a prediction and the actual corresponding value. It is the mean of the absolute numbers describing the error of an estimate (the difference between the given and predicted values), it aims to express the model performance by the degree of prediction accuracy.

QUADRATIC DISCRIMINANT ANALYSIS

We're exploring quadratic classifiers, like Quadratic Discriminant Analysis Classifier (QDAC), for Bitcoin price forecasting. Using Python and Matplotlib, we'll process data, train classifiers, evaluate with metrics, and interpret results. Our goal is accurate data categorization and insights discovery, laying the foundation for further analysis.

They are very much alike just as the linear discriminant function although in the case of the Σk, which is not identical, the quadratic terms cannot be meted out. This function is called discriminant two-order and will contain terms of second order.

LINEAR REGRESSION

To analyze and predict Bitcoin prices using Linear Regression, we first gather historical data and relevant features, then preprocess them. Feature selection is conducted to identify key factors influencing price fluctuations. The data is split into training and test sets for model development and evaluation. A Linear Regression model is trained on the training data to establish a linear relationship between prices and features. Evaluation metrics like MSE, R-squared, and RMSE assess model accuracy.

DECISION TREE REGRESSION

Decision tree regression, widely appreciated for its simplicity, interpretability, and capacity to capture complex data relationships, partitions the feature space hierarchically to predict continuous outcomes, aiming to minimize variance within each partition. However, without proper constraints, it may overfit, a challenge addressed by techniques like pruning and ensembles. Additionally, it serves as a valuable tool in exploratory data analysis, offering insights into feature importance and interactions, facilitating a deeper understanding of the underlying data dynamics.

RANDOM FOREST

The method involves downloading Bitcoin price data and visually analyzing price movements. After preprocessing the data, including extracting price change approximations and target variables, the dataset is split into training and test sets. A Random Forest classifier is trained on the training data and evaluated using accuracy metrics and visualizations. The roadmap focuses on improving preprocessing, training, evaluating performance, and visualizing results to meet analysis objectives.

References

In recent years, the integration of machine learning (ML) and artificial intelligence (AI) in trading has gained significant traction. A multitude of studies have explored various ML algorithms for Bitcoin price prediction. Notably, random forest regression has shown superior performance in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE), effectively identifying influential variables and recommending a lag period model [1]. While LSTM models are widely used, opportunities exist for developing better algorithms [2]. Analyses focusing on daily trends and optimal feature selection have demonstrated their efficacy in predicting daily price changes accurately over a span of five years [3]. Classification-based ML approaches have achieved over 86% accuracy in predicting the direction of the cryptocurrency market *[4]*. Among regression-based algorithms, ordinary least squares (OLS) and logistic regression have shown high accuracy in Bitcoin price prediction *[5]*. Furthermore, sample dimensionality significantly impacts ML accuracy, with support vector machines (SVM) outperforming k-nearest neighbors (KNN) in forecasting Bitcoin prices *[6]*. Contrasting results between random forest regression and LSTM models underscore the importance of variable selection *[7]*. Research gaps remain in the comparative performance of various regression algorithms for predicting Bitcoin movement *[8]*. Recurrent neural networks (RNN) with LSTM architecture have demonstrated superiority over ARIMA models in cryptocurrency forecasting *[9]*, enhancing Bitcoin pricing forecasts by emphasizing the significance of stable ML tasks *[10]*. The need for diversified modeling techniques for accurate Bitcoin price prediction is well-documented *[11]*. A study compares model performance and favors the ARMA model, providing valuable insights for investors and researchers *[12]*. Another study utilizes an LSTM artificial neural network to forecast the next day's Bitcoin closing price using various feature combinations *[13]*. Additionally, a study predicts Bitcoin prices with 64.84% accuracy for daily prices and 59.4% for 5-minute intervals using ML *[14]*. Although predicting Bitcoin prices is inherently complex, deep neural networks and LSTM architectures show promise, with ongoing improvements through hyper parameter tuning *[15]*.

This paper addresses the research gap by systematically comparing various ML models, including SVM, Bayesian regression, random forest, LSTM, ARIMA, linear regression, and quadratic discriminant analysis (QDA), for Bitcoin price prediction. While previous studies have explored individual models, this paper fills the gap by evaluating their comparative performance in terms of predictive accuracy. Additionally, the paper emphasizes the importance of fostering a data-driven culture within organizations to enhance predictive analysis capabilities, highlighting the critical role of organizational culture in leveraging big data analytics for decision-making processes. Thus, this paper contributes to the literature by providing insights into the most efficient ML models for Bitcoin price prediction and underscores the importance of organizational culture in data-driven decision-making.

III. RESULTS

Table 1

ALGORITHM	PRECISION	F1-SCORE	RECALL	SUPPORT	ACCURACY
Quadratic-discriminant analysis(class 0)	0.83	0.39	0.25	311	64.37
Random Forest(class 1)	0.56	0.56	0.57	360	52.90
LSTM	0.54	0.70	1.00	360	53.65
ARIMA(class 1)	0.49	0.66	1.00	332	49.47
SVM(class 1)	1.00	0.84	0.72	311	87.034

ALGORITHM	MEAN-SQUARED ERROR	R-SQUARED	ROOT MEAN SQUARED ERROR
Linear Regression	587365.9181	0.99791977	766.3980154
Decision Tree	1290074.8788	0.99543	1135.8146

IV. DISCUSSION

In classification tasks, precision, recall, F1-score, and support are essential metrics for evaluating model performance, each providing unique insights: precision checks the extent to which positive picked instances are actually positive, recall tests the ability of the model to identify all positive instances, the F1 score balances between precision and recall and support provides an easy way to weight the classes in the data set. These metrics all work together in discerning the right and wrong model for specific ends as well as in deciphering the efficiency of the model across classes in multiple classification. For example, while using the experiment of the prediction of the Bitcoin prices, the accuracy of the Model is shown to be excellent, a precision of 1.00 and F1-score of 0.84 for class 1 were obtained with the total accuracy of 87%. 034%. However, the precision recorded in the current study by applying ARIMA model was relatively low, with a value of 0. 49 and an accuracy of 49/ The mean score reached 49 and the accuracy degree was ended up being 49. 47% to 57 % respectively indicating moderate performance by the models.. 90% to 76. Otherwise, the number of correctly recognized messages for the LSTM model was rather low being equal to 53% while the number of perceived incorrect messages was as high as 40%, that is 0%. 65%. From these analyses, it can be concluded that while some disadvantages and advantages are inherent to each of the analyzing algorithms, the best results and high accuracy of the model is obtained using the SVM. Additionally, when it comes to regression problems, Linear Regression provided a better solution in this case of predicting Bitcoin prices compared to Decision Tree Regression that yielded lesser Mean Squared Error equal to 587,365. 9181 and a higher R-squared of 0. 99791977 and a lesser Root Mean Squared Error equal to 766. 3980154 making it more accurate and suitable for this particular use of predicting Bitcoin prices.

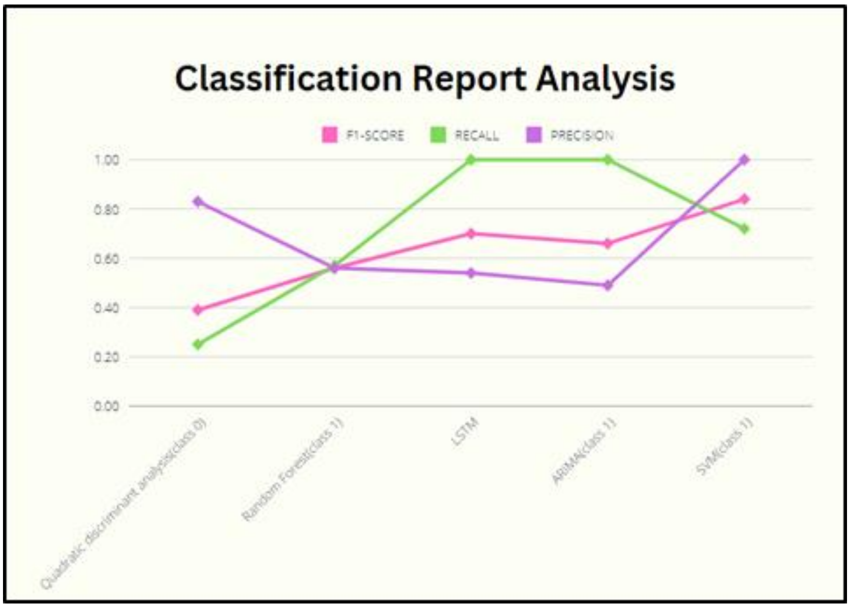

Figure 1. Classification report analysis(precision, recall, f1-score)

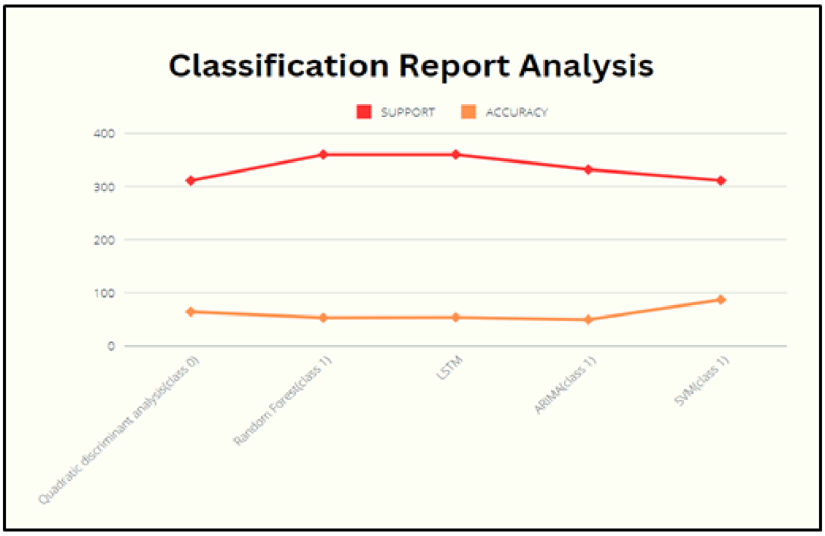

Figure 2. Classification report analysis(support , accuracy)

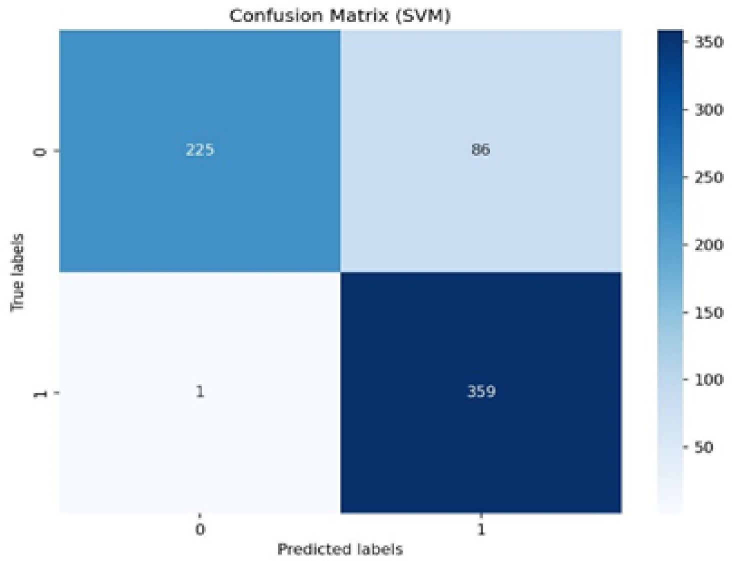

Figure 3. Confusion matrix for support vector machine algorithm

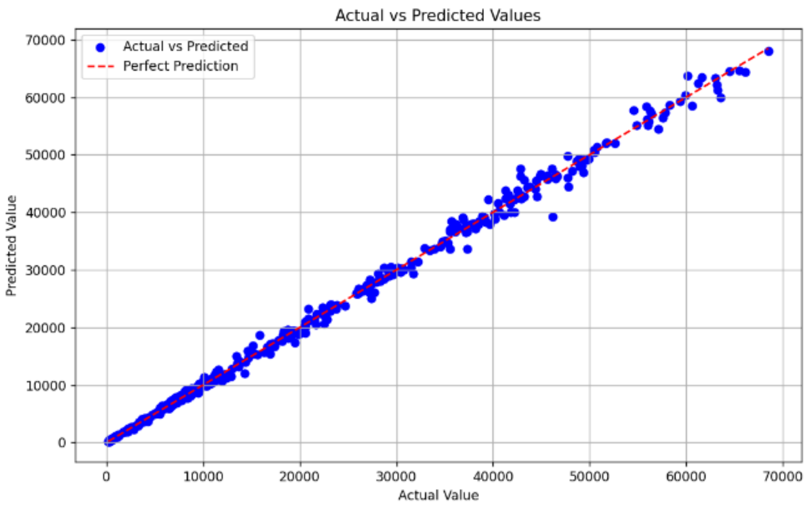

Figure 4 . Scatter plot for Linear Regression Algorithm

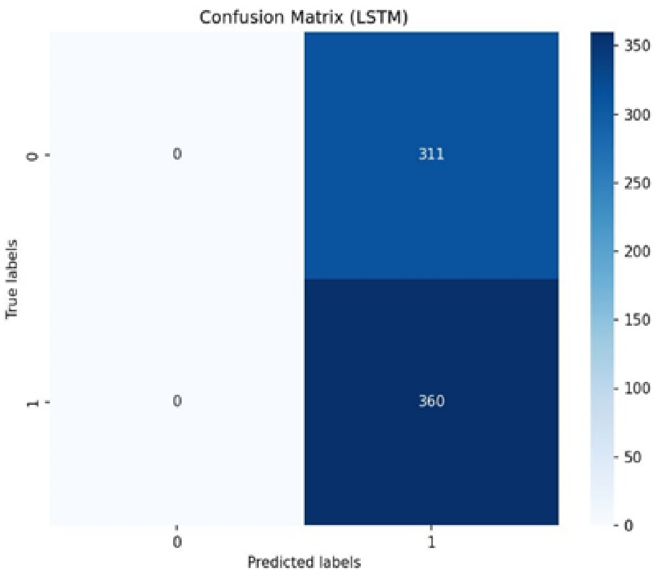

Figure 5. Confusion Matrix for long short term memory

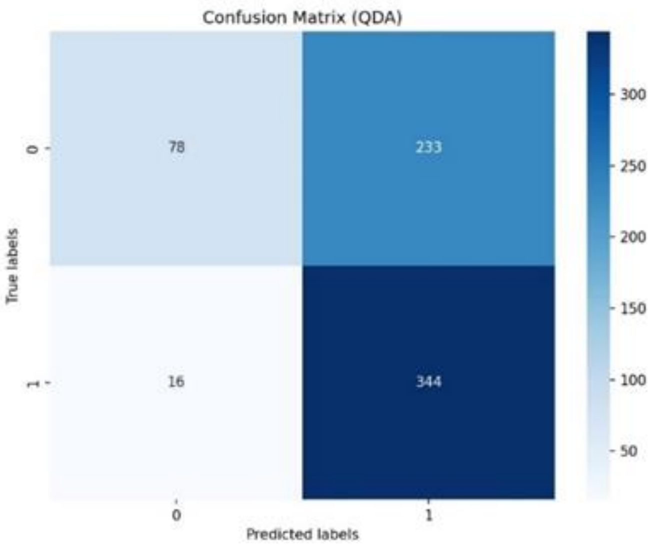

Figure 6. Confusion Matrix for Quadratic Discriminant Analysis

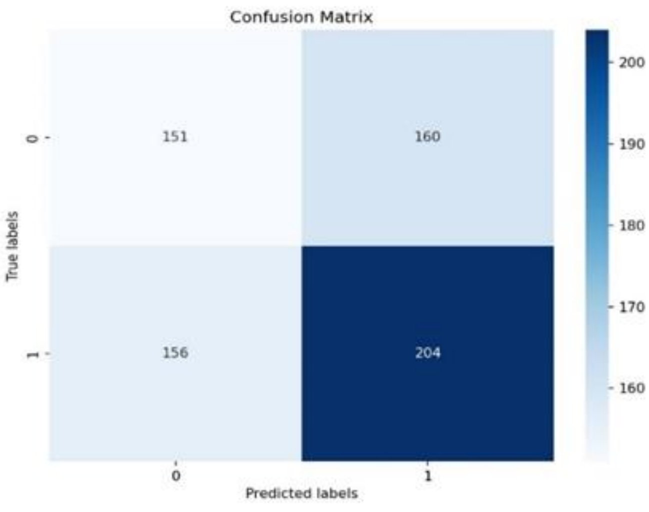

Figure 7. Confusion Matrix for Random Forest Algorithm

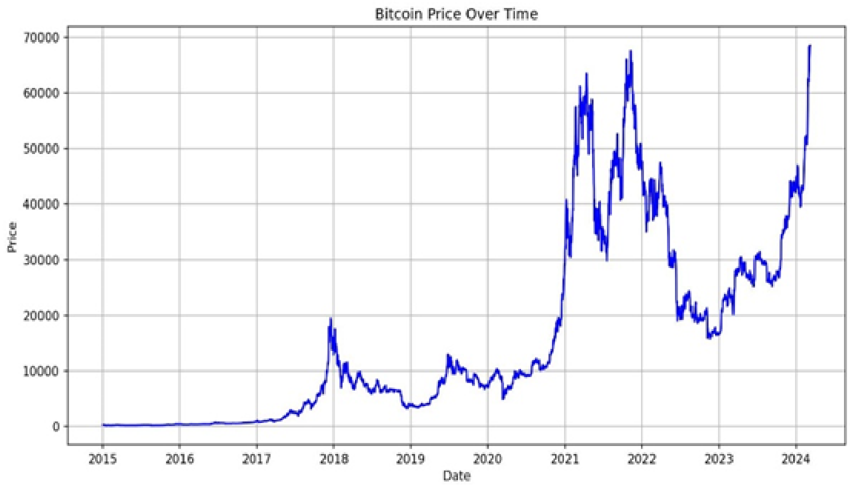

Figure 8. Bitcoin Price Variations over Time (2015-2024)

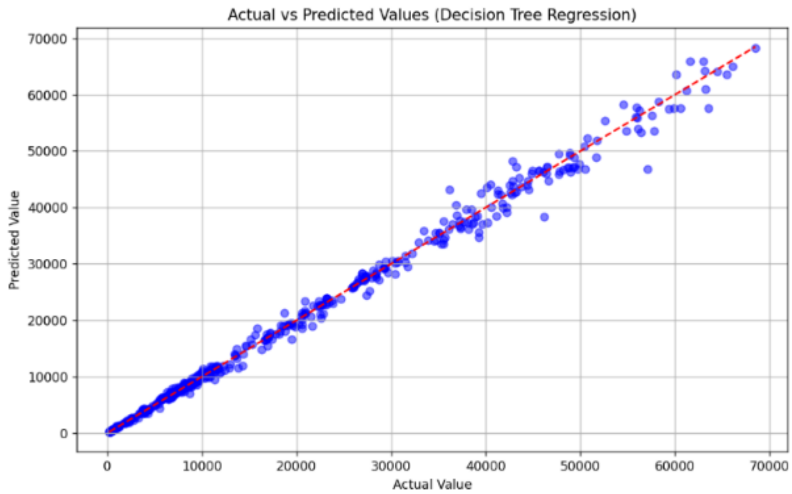

Figure 9. Scatter Plot for Decision Tree Regression Algorithm

V. CONCLUSION

Our analysis of various machine learning algorithms for predicting Bitcoin prices highlights Support Vector Machine (SVM) and Linear regression as the most effective options. SVM demonstrates high accuracy in classification tasks, while Linear regression outperforms in regression tasks, showing superior predictive accuracy and precision. Combining these methods could further enhance predictive capabilities. However, algorithm choice should consider factors like interpretability and scalability. Overall, leveraging SVM and Linear regression, alongside fostering a data-driven culture, enhances decision-making in the digital economy.

ACKNOWLEDGMENT

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] Chen, J. (2023) 'Analysis of bitcoin price prediction using machine learning', Journal of Risk and Financial Management, 16(1), p. 51. doi:10.3390/jrfm16010051.

[2] M, Vaidehi & Pandit, Alivia & Jindal, Bhaskar & Kumari, Minu & Singh, Rupali. (2021). BITCOIN PRICE PREDICTION USING MACHINE LEARNING. International Journal of Engineering Technologies and Management Research. 8. 20-28. 10.29121/ijetmr.v8.i5.2021.953.

[3] Velankar, S., Valecha, S. and Maji, S. (2018) 'Bitcoin price prediction using machine learning', 2018 20th International Conference on Advanced Communication Technology (ICACT) [Preprint]. doi:10.23919/icact.2018.8323675.

[4] Hafid, A., Senhaji Hafid, A. and Makrakis, D. (2022) Bitcoin price prediction using machine learning and technical indicators [Preprint]. doi:10.20944/preprints202212.0188.v1.

[5] Gupta, M. and Srivastava, R. (2021) 'A comparative analysis and prediction over Bitcoin Price using machine learning technique', Fusion: Practice and Applications, pp. 31–41. doi:10.54216/fpa.050103.

[6] Nagamani, P. et al. (2023) 'Bitcoin price prediction using machine learning algorithms', Advances in Engineering Research, pp. 389–396. doi:10.2991/978-94-6463-252-1_43.

[7] Rajua, S. M., & Tarif, A. M. (2022). Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis. Computer Science. International Islamic University Malaysia, Gombak, Malaysia.

[8] Dimitriadou, A.; Gregoriou, A. Predicting Bitcoin Prices Using Machine Learning. Entropy 2023, 25, 777. https://doi.org/10.3390/e25050777

[9] KERVANCI, I. sibel and AKAY, F. (2020c) 'Review on bitcoin price prediction using machine learning and Statistical Methods', Sakarya University Journal of Computer and Information Sciences, 3(3), pp. 272–282. doi:10.35377/saucis.03.03.774276.

[10] Nithiya, S., & Parimala, G. (2023). Bitcoin Price Prediction Using Machine Learning. In Proceedings of the 2023 International Conference on Computer Communication and Informatics (ICCCI) (pp. xxx-xxx). IEEE. DOI: 10.1109/ICCCI56745.
2023.10128307

[11] Chen, J. (2023). Analysis of Bitcoin Price Prediction Using Machine Learning. Journal of Risk and Financial Management, 16(1), 51. https://doi.org/10.3390/jrfm16010051

[12] Hao, M., Su, F., Wang, K., & Zheng, X. (2022). Bitcoin Price Prediction Based on Machine Learning and Granger Causality Test. DOI: 10.2991/978-94-6463-036-7_51

[13] Singh S, Pise A, Yoon B. Prediction of bitcoin stock price using feature subset optimization. Heliyon. 2024 Mar 19;10(7):e28415. doi: 10.1016/j.heliyon.2024.e28415. PMID: 38560114; PMCID: PMC10981097.

[14] Ranjan, S., Kayal, P. and Saraf, M. (2022) (PDF) bitcoin price prediction: A machine learning sample dimension approach.Available at: https://www.researchgate.net/publication/360096794_Bitcoin_Price_Prediction_A_Machine_Learning_Sample_Dimension_Approach (Accessed: 14 May 2024).

[15] M Vaidehi,. *et al.* (2021) *View of bitcoin price prediction using Machine Learning: International Journal of Engineering Technologies and Management Research*, *View of BITCOIN PRICE PREDICTION USING MACHINE LEARNING | International Journal of Engineering Technologies and Management Research*. Available at: https://www.granthaalayahpublication.org/ijetmr-ojms/index.php/ijetmr/article/view/IJETMR21_A05_2586/771 (Accessed: 01 June 2024).