International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Bitcoin Prediction Using Time Series Analysis and Machine Learning

Miss.Chetana B Wagh , Prof.Rutuja Patil

Master of Computer Application ,Trinity Academy of Engineering , Pune, India

Assistant Professor(Dept. Of MCA) ,Trinity Academy of Engineering , Pune, India

ABSTRACT:

This study aims to develop a predictive model for **Bitcoin** market price utilizing statistical analysis methods. By analyzing four years of Bitcoin data, our research focuses on the time series approach, specifically the **ARIMA** model, to forecast Bitcoin **price volatility accuratel**y. The model has shown promising results, achieving an **accuracy rate of** 90% in predicting short-term f**luctuations** in Bitcoin prices

I INTRODUCTION :

Bitcoin has captivated the attention of various individuals, ranging from researchers to trade investors. As the first and most widely accepted cryptocurrency, Bitcoin has gained significant popularity This study aims to develop a predictive model for Bitcoin market price utilizing statistical analysis methods. By analyzing four years of Bitcoin data, our research focuses on the time series approach, specifically the ARIMA model, to forecast Bitcoin price volatility accurately. The model has shown promising results, achieving an accuracy rate of 90% in predicting short-term fluctuations in Bitcoin prices.

since its inception in 2009. This can be attributed to its decentralized trading system that eliminates the need for intermediaries, as well as the high volatility of Bitcoin price. In this study, we present a model that effectively predicts the market price of Bitcoin by leveraging statistical analysis techniques. Our analysis is based on four years of Bitcoin data, spanning from 2013 to 2017, and employs time series approaches, particularly the autoregressive integrated moving average (ARIMA) model. Our research has demonstrated an accuracy rate of 90% in forecasting

II LITERATURE REVIEW:

Time Series Analysis Methods:
ARIMA (Autoregressive Integrated Moving Average):

ARIMA is a classic method that forecasts future values based on past values and errors in the time series. Studies, such as the one by Azari (2019), have demonstrated its application in predicting Bitcoin prices. However, it has limitations in terms of long-term accuracy.

SARIMA (Seasonal ARIMA):
SARIMA builds upon ARIMA by incorporating seasonality, which can be beneficial in capturing cyclical patterns in Bitcoin prices.

Machine Learning Techniques:
Support Vector Machines (SVM):
Research, such as the study by Gupta et al. (2021), has shown promising results with SVMs in short-term Bitcoin price prediction, achieving up to 65% accuracy.

Long Short-Term Memory (LSTM):
LSTM, a recurrent neural network, is proficient in capturing long-term dependencies in data. Studies, like the one by Christoffel et al. (2020), have highlighted its superior performance compared to traditional methods in Bitcoin price prediction.

Other Techniques:
Artificial Neural Networks (ANNs), Gradient Boosting, Random Forests: Various methods have been explored in predicting Bitcoin prices, each demonstrating varying degrees of success based on the specific data and prediction horizon.
Important Considerations:
Limited Success for Long-Term Prediction:
While the aforementioned techniques show effectiveness in short-term

Forecasting ,accurately prediction bitcoin prices over the long term remain a

Challenging task

Data Choice and Feature Engineering:

The quality and comprehensiveness of historical price data ,as well as additional factors like trading volume , play a crucial role in determining the

Performance of prediction models.

In adddition to historical data , external factors such as new , regulation , and social media sentiment significantly influence Bitcoin prices . Effectively incorporating these variable into model poses a notable challenge in price

Prediction.

III RESEARCH METHODOLOGY:

Machine learning, a crucial branch of artificial intelligence (AI), plays a significant role in predicting various outcomes. In this study, the research methodology focuses on using supervised learning to forecast future Bitcoin prices.

Supervised Learning Approach:

Supervised learning involves training regression function to make predictions based on existing data. The machine learning process entails setting algorithms, creating learners, training models with data, validating the results, and evaluating performance with test data

Implementation of Models

Two key models, random forest regression, and Long Short-Term Memory (LSTM) are utilized in this study through Python's machine learning libraries. Random forest regression employs the sklearn library, while LSTM utilizes keras. Data preprocessing and organization are handled using pandas.

Random Forest Regression

Random forest regression consists of multiple regression trees that offer high interpretability. The model's performance is influenced by parameters such as the maximum depth of sub-regression trees and the number of sub-trees in the random forest. Experimentation reveals the impact of varying these parameters on training and prediction errors.

LSTM Model

LSTM, a variation of Recurrent Neural Network (RNN), addresses the challenge of short memory in traditional DNN algorithms. The structure of the LSTM model in this study incorporates multiple layers with specific units and activation functions. The optimization of dropout layers and epochs is crucial in balancing model accuracy and overfitting risks.

Evaluation and Analysis:

The study evaluates model performance using error metrics such as MAPE , RMSE , and DA. Comparsion of different algorithms is conducted through hypothesis testing methods The Diebold –mariano test and the clark-west test. Additionally, the study aims to analyze Memory length characteristics in predicting bitcoin prices under different lagging explanatory variable.

The use of machine learning , particularly supervised learning techniques likerandom forest regression and LSTM, provides valuable insight into predicting future bitcoin price.by employing a rigorous research methodology and evaluating model performance this study contributes to the understanding of cryptocurrency market dynamics

Data collection instruments: To enhance data collection for Bitcoin price prediction, reliable sources such as cryptocurrency exchanges and financial data providers are utilized. Data variables beyond price, including trading volume and market capitalization, are collected to provide a comprehensive analysis.Data cleaning and time series aggregation are crucial steps to ensure accurate analysis and trend identification

Data collection procedures:

Data collection procedures play a crucial role in research projects, ensuring that accurate and valuable information is gathered to meet the objectives of the study. This article will provide a comprehensive overview of the data collection procedures commonly used in research, covering surveys and interviews in detail.

Surveys:

Questionnaire Design: When designing a survey questionnaire, it is essential to develop a set of questions that align with the research objectives. Ensure that the questions are clear, concise, and unbiased to gather reliable data from participants.

Sampling: The first step in the data collection process is to determine the target population and select a representative sample. Sampling methods such as random sampling or stratified sampling can be used to ensure the sample is reflective of the population.

Distribution: Surveys can be administered through various channels, including online platforms, mail, phone, or in-person interactions. Choose the most appropriate distribution method based on the target audience. Data Collection: Gather responses from participants, ensuring anonymity and confidentiality if required. Collecting data accurately is crucial to obtaining reliable results for analysis.

Data Entry: Transfer responses into a digital format for analysis, ensuring accuracy and integrity throughout the data entry process.

Analysis: Utilize statistical tools to analyze the data collected from surveys , identifying trends , patterns , and correlations to drawmeaningfulconclusion.

Reporting: Interpret findings from the survey data and present results in a clear, organized manner. Visual aids such as charts or graphs can be used to illustrate key points effectively.

Interviews:
Planning: Before conducting interviews, define the purpose of the interview and establish objectives. Determine the type of interview that best suits the research, whether structured, semi-structured, or unstructured. Participant Selection: Identify suitable participants who can provide valuable insights related to the research topic. Select participants based on criteria that align with the research objectives.

Question Development: Prepare a list of open-ended questions to guide the conversation during interviews, encouraging detailed responses from participants

Conducting Interviews: Schedule and conduct interviews either in person, over the phone, or through video conferencing. Building rapport with participants facilitates open communication and enhances the quality of data collected.
Recording: With participants' consent, record interviews through audio or video recording, or detailed note-taking. Accurate recording of interviews is essential for thorough data analysis.

Transcription: Transcribe recorded interviews accurately to capture all relevant information and nuances essential for analysis.

Data Analysis: Analyze transcripts for recurring themes, patterns, and insights using qualitative analysis techniques. Extract valuable information from interview data to support research findings.

Reporting: Summarize key findings from interviews, supported by quotes or examples to illustrate points effectively. Present results in a clear and concise manner to convey research outcomes accurately. Effective data collection procedures are vital for the success of research projects, ensuring that reliable and valuable information is obtained to meet research objectives. By following the outlined procedures for surveys and interviews, researchers can collect, analyze, and report data effectively to draw meaningful conclusions

IV CONTESTANT:

Predicting Bitcoin prices through time series analysis involves gathering historical data, cleaning and preprocessing it, then exploring trends and patterns through techniques like ARIMA modeling. This approach entails forecasting future prices based on past data, with model performance evaluated using metrics such as MAE or RMSE. Meanwhile, analyzing a machine learning contest necessitates a thorough understanding of the competition's objectives and rules, followed by data exploration, feature engineering, model selection, training, validation, and iterative refinement. Success in both endeavors hinges on robust methodologies, informed decisions, and diligent evaluation to achieve accurate predictions or competitive performance.

V Measurement:

While time series analysis offers a tempting approach for predicting Bitcoin's future value, it's important to understand both its potential and limitations. This method involves analyzing historical price data to uncover patterns and trends. Techniques like ARIMA (Autoregressive Integrated Moving Average) and LSTMs (Long Short-Term Memory networks) can be used to identify seasonality, cycles, and dependencies within the data. By studying these patterns, time series models can attempt to forecast future prices.However, Bitcoin's inherent volatility throws a wrench into these predictions. Unlike traditional stocks influenced by corporate performance and economic factors, Bitcoin's price is susceptible to a wider range of influences. News events, regulatory changes, hacks, and even social media sentiment can significantly impact its value. These unpredictable external factors can throw off even the most sophisticated time series models, making long-term predictions with any degree of certainty extremely difficult.Despite these limitations, time series analysis can still be valuable for understanding historical price movements and potentially making short-term forecasts. By focusing on smaller timeframes and incorporating additional data sources like trading volume and social media sentiment, these models can offer some insights into Bitcoin's immediate future direction. However, it's crucial to remember that these are just insights, not guarantees. The ever-evolving landscape of cryptocurrency necessitates caution and a healthy dose of skepticism when dealing with any Bitcoin price prediction.

VI PROCEDURE:

In the study, each contestant was provided the survey to complete in depth. The contestant were allotted the chance to express their opinion on how the smartphones affect their social interactions and how this usage may directly correlate with any personal experience they may have.The surveysWere then examined in detail to create tables that display the results and finding of the question and agree /disagree

VII DATA ANALYSIS:

Data analysis a crucial role in predicting Bitcoin prices through time series analysis. By utilizing historical data, patterns, and trends can be, often utilizing techniques such as ARIMA or SARIMA models. These models are trained on past price movements to forecast future trends, with performance assessed through metrics like MAE or RMSE

machine learning data contests. This entails a deep understanding of the competition's objectives and guidelines. Key tasks include data preprocessing, feature Analyzing Machine Learning Data Contests:In addition to predicting Bitcoin prices, data analysis is also essential in analyzing engineering, model selection, and iterative refinement to enhance predictive performance.Key Components of Data Analysis:Successful data analysis in predicting Bitcoin prices or participating in machine learning data contests requires attention to detail, a solid grasp of statistical techniques, and a strategic approach. By leveraging these components, valuable insights can be derived, leading to competitive success.In conclusion, whether it is predicting Bitcoin prices through time series analysis or participating in machine learning data contests, the key to success lies in the meticulous application of data analysis techniques. By harnessing the power of data analysis, individuals can unlock valuable insights and achieve competitive success in the world of cryptocurrency and machine learning.

VII DEMOGRAPHICS QUESTION

O Optimized Title: A Comprehensive Study on Demographics in Bitcoin Price Prediction Introduction:
The research study gathered sample data from various reliable sources such as Yahoo Finance, Coinmarketcap.com, investing.com, bitcharts.com, and coinmatrics.io from31 March 2015 to 1 April 2022. The main focus of the study

was to analyze the price of Bitcoin in USD by utilizing a total of 47 explanatory variables categorized into eight sections: Bitcoin price variables, specific technical features of Bitcoin, other cryptocurrencies, commodities, market indices, foreign exchange, public attention, and dummy variables for the week. Each explanatory variable and its definitions can be found in Appendix A.

Explanatory Variables Analysis:

Analyzing the statistical features of each explanatory variable for Bitcoin price prediction reveals significant insights. Particularly, variables related to the cryptocurrency market, including Bitcoin, other cryptocurrencies, and Google search volume for Bitcoin, exhibit high standard deviations. This high volatility in the cryptocurrency market since 2015 is evident through the ratio of standard deviation to the mean value exceeding 1 for most variables. On the contrary, traditional market variables show a standard deviation/mean ratio not greater than 0.4.

Correlation Analysis:

The correlation heat map depicts the relationship between Bitcoin and other explanatory variables, highlighting positive correlations with other cryptocurrencies, commodity prices, stock market indices, and public attention. Notably, the price of Bitcoin shows an inverse correlation with the 10-year U.S. Treasury yield in the commodities category, while exchange rates generally exhibit a negative correlation. Interestingly, the Russian ruble exchange rate is positively correlated with the Bitcoin price.

Impact of Weekday Variables:

Analysis of Bitcoin price and weekday variables reveals intriguing patterns, with Wednesdays displaying extreme fluctuations and significant daily gains and losses. Additionally, the study shows that the average daily return for Bitcoin is highest on Mondays and lowest on Sundays, with Saturday and Friday having the highest probability of price increase.

Public Attention Variables:

The study compares Google Trends and daily Tweets with the Bitcoin price, noting spikes in public attention coinciding with Bitcoin's price peaks. Despite reaching over $60,000 in 2021, the search volume did not surpass the levels observed at the end of 2017.

Preprocessing and Model Employed:

To enhance the accuracy of price prediction models, the study divides the total sample period into Period 1 (31 March 2015 to 30 September 2018) and Period 2 (1 October 2018 to 1 April 2022). Machine learning techniques are utilized to train and test models separately for each period, ensuring the validation of predictions. The employed model, random forest regression, demonstrates excellent performance in predicting Bitcoin prices, particularly below $60,000.

Importance of Explanatory Variables:

Analyzing the importance of explanatory variables in predicting Bitcoin prices reveals that OHLC prices of Bitcoin from the previous period are crucial. The stock market index plays a significant role, with the Japanese stock market index showing increased importance in Period 2. Furthermore, Ethereum (ETH) emerges as a key variable for accurate price predictions in both periods.

Conclusion:
The research showcases the importance of various explanatory variables in predicting Bitcoin prices accurately. The comparison between models with all variables and significant variables highlights the superior performance of the former, emphasizing the need for a comprehensive approach to Bitcoin price prediction studies. The results offer valuable insights into the dynamics of cryptocurrency markets and their impact on Bitcoin price fluctuations

IX RESULTS:

Results of LSTM:
The experiments conducted revealed that the accuracy of the LSTM model is influenced by the selection of explanatory variables. Bringing in too many redundant variables decreased the model accuracy, while using too few variables also resulted in reduced prediction accuracy. By conducting various experiments, it was determined that the ideal set of explanatory variables varied for each period. It was observed that the random forest regression model did not face similar challenges, thanks to its ensemble algorithm.
Explanatory Variable Sets:
The importance rank of explanatory variables was determined using random forest regression, and specific sets of variables were identified for Period 1 and Period 2. This highlights the direct impact of the combination of explanatory variables on prediction accuracy.

Model Comparison:
To ensure a reliable comparison of model results, the average of 30 experiments for various models was considered. The analysis of one-lagged accuracies for two periods showed variations in prediction accuracy, with Period 1 outperforming Period 2.
Relationship between Precision and Number of Variable Periods:
The relationship between model accuracy and the number of lags of explanatory variables was explored by comparing the results of models with lags ranging from 1 to 5. The study revealed that the MAPE of random forest regression increased with the addition of more periods, supporting the efficient market hypothesis.
Conclusion:
The analysis concluded that LSTM demonstrated good predictive performance for time series data, with the optimal model requiring only the data from the previous period. The comparison between LSTM and random forest regression models revealed varying prediction accuracies for different periods, indicating the changing dynamics of Bitcoin prices.
Implications:
While LSTM remains a popular choice for predictive modeling, the study showed that random forest regression could be equally effective in predicting Bitcoin prices, especially during volatile market conditions. Further research is needed to explore the potential of combining both models for enhanced predictive accuracy.

X DISCUSSION AND CONCLUSION

The process of predicting Bitcoin prices utilizing time series analysis and machine learning involves the utilization of various techniques such as ARIMA, LSTM, or Prophet models. These methodologies analyze historical price data to forecast future trends. However, the accuracy of these predictions can vary due to the volatile nature of cryptocurrency markets.
Heading: Machine Learning Discussions for Bitcoin Prediction
Discussions surrounding machine learning in the realm of Bitcoin prediction often center around enhancing model robustness, feature selection, and data preprocessing to improve forecasting accuracy. The incorporation of

these methodologies is essential for producing more accurate predictions in the ever-changing landscape of cryptocurrency trading. By focusing on refining these techniques, analysts and traders can better anticipate market movements and make informed decisions

REFERENCE:

1. Title: Cutting-Edge Studies on Bitcoin Price Prediction Using Deep Learning and Machine Learning Models
Heading: Latest Research Insights on Bitcoin Price Forecasting
Subheading 1: Aggarwal, Apoorva, Isha Gupta, Novesh Garg, and Anurag Goel Study
In a groundbreaking research study by Aggarwal, Apoorva, Isha Gupta, Novesh Garg, and Anurag Goel presented at the 2019 Twelfth International Conference on Contemporary Computing (IC3) in Noida, India, the impact of socio-economic factors on Bitcoin price prediction was determined using a deep learning approach.
Subheading 2: Akyildirim, Erdinc, Oguzhan Cepni, Shaen Corbet, and Gazi Salah Uddin Study
Another significant study conducted by Akyildirim, Erdinc, Oguzhan Cepni, Shaen Corbet, and Gazi Salah Uddin in 2021 focused on forecasting the mid-price movement of Bitcoin futures using machine learning techniques, providing valuable insights for traders and investors.
Subheading 3: Awoke, Temesgen, Minakhi Rout, Lipika Mohanty, and Suresh Chandra Satapathy Study
Awoke, Temesgen, Minakhi Rout, Lipika Mohanty, and Suresh Chandra Satapathy explored Bitcoin price prediction and analysis using deep learning models in their research study published in the book "Communication Software and Networks" by Springer. Their findings shed light on the potential of deep learning in predicting Bitcoin prices accurately