PREDICTIVE MODELING OF CRYPTOCURRENCY PRICES USING REGRESSION ANALYSIS

Project Members: Dodla Divya-UA03330, Kandregula Naga Satya Sai Naren-PW71937

Abstract

The value of Bitcoin, the first decentralized digital currency, has increased significantly; it started at $1 in 2010 and peaked at about $18,000 in 2017 (Tambe, 2023). Finance, economics, and computer science professionals have taken notice of this increase in value. Acquiring and preparing the Bitcoin dataset for analysis retrieved from the Yahoo historical data section of the Bitcoin USD (BTC-USD) page (Yahoo!, 2023). This Project uses a comprehensive methodology that combines more sophisticated approaches like XGBoost and Long Short-Term Memory (LSTM) with traditional forecasting methods like Autoregressive Integrative Moving Average (ARIMA) to acquire insights into the price trends of Bitcoin. The study intends to offer insightful forecasts for Bitcoin's pricing in 2023 by examining past data from 2014 to 2022, with a focus on the ARIMA model's accuracy in predicting future values.

Business Understanding

It is difficult to predict the value of Bitcoin due to the market's intrinsic volatility and decentralized structure. The financial sector will be greatly impacted by using blockchain technology, including hybrid, private, and public versions. Traditional forecasting models, such as Holt-Winters, find it difficult to describe the distinctive features of the Bitcoin market. To increase accuracy, there is a growing tendency towards the use of machine learning techniques like XGboost, LSTM, and ARIMA, which combine historical prices and market data. The inability to reliably predict Bitcoin prices, despite their growing popularity, emphasizes the limitations of machine learning in this domain.

Modeling Approach

Approach

For our project, where we take a structured approach to predicting bitcoin prices, we are adopting the CRISP-DM methodology. We will start by compiling and organizing historical Bitcoin records. We clean and standardize the data to ensure its accuracy.

Assumptions

We are aware that historical cryptocurrency trends can be useful in making forecasts. However, it can be challenging to estimate prices with accuracy because of the erratic nature of these markets and other influences. Although our model allows for the possibility of market conditions changing and adapting, it might not be able to instantly account for unanticipated occurrences and their immediate effects.

Feature Engineering

Our goal is to enhance market analysis, assess trends, and make better judgments by employing a variety of important indicators and trading tools, including SMA, EMA, MACD,

RSI, Bollinger Bands, True Range, and Williams %R. These indicators help see patterns, gauge momentum, and spot potentially overbought or oversold market circumstances (Levi, 2023).

Exploratory Data Analysis (EDA) & Correlation Matrix

Finding patterns and trends in the data through exploratory data analysis (EDA) provides the foundation for further modeling stages. We can learn more about the dataset's features by utilizing statistical metrics and visualizations. The correlation matrix facilitates the identification of correlations between variables, which helps us choose pertinent model features and makes it easier to understand the model's output.

Modelling

Regression modeling is the method we have chosen to forecast bitcoin prices. Our modeling method is diverse due to the inclusion of XGBoost, LSTM, and ARIMA models in our ensemble. We divided the dataset into training (80%), validation (10%), and testing (10%) sets to train, assess, and test the models. After training, the top-performing model is chosen using predetermined metrics. During the testing stage, the selected model forecasts future bitcoin prices based on the test data.

XGBoost Regression

We use the XGBoost regression in conjunction with Principal Component Analysis (PCA) to increase the precision of our Bitcoin price prediction. Ten PCA components, according to our study, are in the ideal balance. Randomized Search cross-validation is used for hyperparameter tuning to maximize the performance of the XGBoost regression model. 'TargetNextClose' price, our target variable, is forecasted by combining technical indicators with past stock data. Outstanding predictive performance is demonstrated by the XGBoost model, which scores highly on the training set (RMSE: 992.26, MAPE: 12.81%, R2: 0.996) and validation set (RMSE: 3917.45, MAPE: 12.81%, R2: 0.837). This demonstrates how well the model can forecast the future.

Long Short-Term Memory (LSTM)

We apply the Long Short-Term Memory (LSTM) model, which is specifically built to examine sequential data, to enhance the prediction accuracy of Bitcoin prices. The selection of 14 timesteps is strategically made to balance computational efficiency and pattern capture. Through strategic data transformation and callback integration, our LSTM model ensures optimal prediction by capturing complex patterns. We leverage key optimization parameters such as the ADAM optimizer and linear activation function to strike a balance between convergence speed and accuracy (GeeksforGeeks, 2023). The LSTM architecture, featuring three layers and employing dropout for regularization, effectively utilizes historical price and volume data, along with technical indicators, for predicting the 'TargetNextClose' price. Achieving strong performance on both training (RMSE: 0.0007, MAPE: inf%, R2: 0.987) and validation (RMSE: 0.0005, MAPE: 4.48%, R2: 0.973) sets, our LSTM model stands out for its capability to capture intricate patterns in Bitcoin's dynamic market. LSTM proves to be

a robust choice, surpassing traditional algorithms in handling sequential dependencies and demonstrating superior predictive accuracy in cryptocurrency price forecasting.

Autoregressive Integrated Moving Average (ARIMA)

We use the Autoregressive Integrated Moving Average (ARIMA) model in our quest for precise Bitcoin price prediction. We chose ARIMA because of its simplicity, statistical robustness, and capacity to manage seasonality. ARIMA is a straightforward univariate statistical model that works well for financial data analysis, in contrast to XGBoost and LSTM. It offers trustworthy short- to medium-term forecasts and works well with non-stationary data. ARIMA's flexibility and historical dependency modeling make it a useful tool for investors and traders. Our ARIMA model predicts future prices as the target variable using just past Bitcoin prices as input. We use walk-forward validation to simulate real-world settings, which enables the model to adjust to evolving patterns dynamically. The hyperparameters (p=5, d=1, q=0) were chosen after thorough experimentation to balance the need to preserve simplicity and capture significant patterns (Brownlee, 2023). With an RMSE of 969.51, R-squared score of 0.99, and MAPE of 2.26%, the model performs admirably on the validation data. These outcomes support ARIMA's efficacy as a price predictor for bitcoin.

Model Selection

Upon analyzing our models, thorough metrics for the validation set indicate that ARIMA is the best option. With an impressive R-squared score of 0.990 and a low Mean Absolute Percentage Error (MAPE) of 2.257%, ARIMA proves to be a successful model. ARIMA is a statistical time series model, and its strong performance on this dataset suggests that it captures the underlying patterns and trends effectively. Although they were competitive, XGBoost and LSTM models fell short of ARIMA's accuracy and fit. ARIMA's strong performance is additional evidence that it is the best model for predicting Bitcoin prices.

Conclusion

To sum up, the ARIMA model's excellent R-squared scores and low error metrics demonstrate how well it predicts Bitcoin prices. We find a strong degree of consistency between the projections produced by the ARIMA model and the actual Bitcoin closing prices based on historical data from Yahoo. This attests to the ARIMA model's predictive power for Bitcoin prices. Its ability to anticipate the price of Bitcoin successfully also suggests possible uses for other cryptocurrencies, such as Ethereum. The model will be further enhanced in the future by adding more features, applying ensemble modeling strategies, putting real-time updates into place, and producing user-friendly visuals. The modifications are intended to improve the model's flexibility and readability. An all-encompassing method for projecting Bitcoin prices can be produced by expanding the analysis by integrating external datasets. Overall, the ARIMA model is positioned as a useful tool for accurate and perceptive Bitcoin predictions due to its current performance and planned improvements.

References

Brownlee, J. (2023, November 18). *How to create an Arima model for time series forecasting in Python*. Retrieved from MachineLearningMastery.com: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/

GeeksforGeeks. (2023, June 5). *Understanding of LSTM networks*. Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/understanding-of-lstm-networks/

Levi, K. L. (2023, May 22). *Best 10 indicators for cryptocurrency trading: Primexbt *. Retrieved from PrimeXBT: https://primexbt.com/for-traders/best-indicators-for-cryptocurrency-trading/

Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). *Bitcoin and cryptocurrency technologies: a comprehensive introduction. Princeton University Press.*

Tambe, N. (2023, December 6). *Bitcoin price prediction: Can Bitcoin reach $1,000,000 by 2025?* Retrieved from Forbes: https://www.forbes.com/advisor/in/investing/cryptocurrency/bitcoin-prediction/

Yahoo! (2023, December 19). *Bitcoin USD (BTC-USD) price history & historical data*. Retrieved from Yahoo! Finance: https://finance.yahoo.com/quote/BTC-USD/history/?guccounter=1