CS470 Project Proposal

Thomas Ma
Emory University
Computer Science
thomas.ma@emory.edu
ID: 2583313

Alex Liu
Emory University
Computer Science/Math
alexander.liu@emory.edu
ID: 2588061

Jack Zhang
Emory University
Computer Science
jack.zhang@emory.edu
ID: 2548618

***Abstract*—Bitcoin, as a decentralized digital asset, exhibits highly volatile price movements influenced by various market, economic, and network-related factors. This study aims to develop a predictive model for Bitcoin prices by incorporating a diverse set of features, including blockchain network metrics, MACD (Moving average convergence/divergence), EMA (exponential moving average), SMA (simple moving average), RSI (relative strength index), gold prices, foreign exchange prices, and lastly bonds/federal funds rate. We plan to evaluate the effectiveness of different machine learning techniques in capturing Bitcoin's price dynamics and assess their predictive accuracy. By analyzing historical price data alongside these features, we seek to determine which variables contribute most significantly to price fluctuations. Our findings provide insights into Bitcoin's price behavior and the feasibility of using data-driven models for cryptocurrency forecasting.**

I. Introduction

This project seeks to take a deep dive into bitcoin and its correlation between different features which is upsampled from the data of bitcoin such as: price, blockchain data, MACD, and many more. We chose this topic because of our innate interest in bitcoin, especially in the recent modern landscape where it has slowly taken over the world as a digital form of currency.

Over the past decade, Bitcoin has emerged as a globally recognized digital asset, revolutionizing the financial landscape. As a decentralized currency, Bitcoin operates independently of traditional financial institutions, relying instead on blockchain technology and cryptographic security. Its limited supply, high volatility, and increasing adoption have attracted both retail and institutional investors. However, despite its growing prominence, Bitcoin's price movements remain highly unpredictable, influenced by a variety of market and network-related factors.

Below is a list of different ways to evaluate the success of our models:

1) Price Analysis:

- Historical price trends
- Moving Average Convergence Divergence (MACD)
- Relative Strength Index (RSI)

2) Blockchain Metrics:

- Hash rate and mining difficulty
- Transaction volume and fees
- Active wallet addresses

Market Sentiment & External Factors

- Trading volume and liquidity
- News sentiment analysis
- Correlation with traditional assets (e.g., stocks, gold)

II. Related Work

Bitcoin is highly stochastic and volatile as an asset. However, there is previous literature on Bitcoin price prediction. Application of Deep-Learning models such as CNN-LSTM, LSTNet, and TCN models have been benchmarked against ARIMA with CNN-LSTM achieving 82.44% accuracy (Omole & Enke, 2024) [1], although part of this success can be attributed to the time period on which the models were backtested on. Additionally, the same paper cites uses of features such as sentiment data for sentiment analysis, based off Google Trends, Reddit, and Bitcointalk as well as technical indicators, Bitcoin blockchain transacition records (on-chain data), macroeconomical variables, and price data, generating large datasets with high dimensionality. To deal with the Curse of Dimensionality, the same paper cites many feature selection methods, such as best-first search, PCA, particle swarm optimization, genetic algorithm, LightGBM, etc.

III. Intended Proposed Approaches

The similarity of Bitcoin's price behavior to that of stocks and similar financial investment assets leads to a regression model as the best choice for prediction. Given the well-documented financial history of Bitcoin, Amazon Web Service's Public Blockchain Data [2] offers plenty of background data on Bitcoin's price to train a model on. Since this is the metric of interest in our predictions, the primary focus will be on the price of Bitcoin.

From the prices, another measure to exploit is the moving average convergence/divergence (MACD) of the prices. This is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA. A nine-day EMA of the MACD line is a signal line. A MACD line above the signal line signals to buy Bitcoin, and an MACD line below the signal line signals to sell Bitcoin. A similar value is the relative strength index (RSI), which focuses on determining whether buyers are overbuying or overselling Bitcoin. Overbuying indicates that Bitcoin is at a higher value than its fair value, and overselling indicates that Bitcoin is at a lower value than its fair value. These provide analysis and data for recent and frequent price changes.

A simple moving average (SMA), on the other hand, offers an analysis of the price history over a longer period of time. SMA takes every closing price and provides the average, giving a more consistent and smooth look into the history of Bitcoin. Combined, all of these measures display a holistic narrative of Bitcoin's price.

Other cryptocurrencies struggle to compare to Bitcoin's sheer popularity and price, but several contenders of interest offer extra information on Bitcoin's performance. Longstanding currencies like Ethereum offer some reflection into the general confidence of buyers, and the history of new cryptocurrencies such as XRP (Ripple) being introduced and their respective success provides insight into activity in the market.

Federal influence through interest rates and bonds historically affects the prices of Bitcoin. When the Federal Reserve raises the federal funds rate, buyers are scared off riskier investments like cryptocurrency. Furthermore, higher United States bond yields mean less incentive to buy into Bitcoin. With federal interest and yield rates highly accessible, these data provide a good insight into the confidence of Bitcoin buyers.

IV. Intended System Design

We intend to use Python to conduct our data analysis. Our planned libraries include Pandas, Sklearn, YFinance, PyTorch, NumPy, Matplotlib, boto3 (for retrieval of Blockchain transaction logs from AWS), and Seaborn. although this is subject to change.

We plan to start by building our feature dataset from Bitcoin Prices (collected by Yahoo Finance) and AWS Public Blockchain data [2], which offers a free AWS CLI endpoint 's3://aws-public-blockchain/v1.0/btc/' for Bitcoin transaction logs.

To expand our dataset, we plan on extracting technical indicators directly from bitcoin price data including time-lagged variables and moving averages, and potentially adding other currencies such as Gold and USD, Japanese Yen, etc. Additionally for macroeconomic indicators, we can include key ETF prices as a measure of investor confidence in stock equities and metrics such as Treasury bond yields and federal funds rate/interest rates.

We intend on using Sklearn to build out "out-of-the-box" ML models such as random forest classifiers, XGBoost, and logistic regression. We intend on using Torch to build out more complex models such as the CNN-LSTM model as found in [1] and other models that are not already supported in XGBoost. Additionally, we can Grid Search on several hyperparameters such as feature selection algorithms. For evaluating our models, we should also consider other metrics besides accuracy, although more research needs to be done to better understand how to evaluate time-series models.

References

[1] Omole, O., Enke, D. Deep learning for Bitcoin price direction prediction: models and trading strategies empirically compared. Financ Innov 10, 117 (2024). https://doi.org/10.1186/s40854-024-00643-1

[2] AWS Public Blockchain Data. (n.d.). AWS Open Data Registry. Retrieved February 18, 2025, from https://registry.opendata.aws/aws-public-blockchain/

V. Appendix

A. Task Assignment

B. Time Schedule