Blockchain Cryptocurrency - Price Prediction (COMP3125 Individual Project)

Nathaly D. Phrasavath
Wetnworth Institute of Technology

***Abstract*—This report explores the application of data science and blockchain principles to analyze and predict short-term cryptocurrency price movements, focusing on Bitcoin and Ethereum. Leveraging historical price data, trading volume, and sentiment data from external sources, the study investigates key market indicators that influence price changes, evaluates the performance of machine learning models in forecasting prices, and compares the volatility of Bitcoin with other cryptocurrencies. The analysis is structured around four research questions, incorporating blockchain-based data structures and concepts where applicable to enhance data integrity and transparency. The project combines exploratory data analysis, statistical testing, and predictive modeling to provide insights into the challenges of forecasting in decentralized, volatile financial markets.**

Keywords—blockchain, cryptocurrency, Bitcoin, Ethereum, price predication, machine learning, volatility analysis

I. Introduction (*Heading 1*)

The rise of cryptocurrency has revolutionized the financial landscape, introducing a decentralized and highly volatile digital asset class that continues to challenge traditional forecasting methods. Among the most prominent cryptocurrencies, Bitcoin (BTC) and Ethereum (ETH) dominate both market capitalization and public attention. As speculative interest and institutional adoption grow, accurately predicting short-term cryptocurrency price movements has become a critical objective for traders, investors, and researchers alike.

This project applies data science techniques to investigate and predict cryptocurrency price fluctuations, with particular attention to Bitcoin and Ethereum. The unpredictable nature of these assets has sparked widespread interest in understanding the underlying factors that drive their behavior. Specifically, this report is guided by four core research questions: (1) What factors—such as market trends, transaction volume, or trading patterns—have the strongest correlation with cryptocurrency price changes? (2) How accurately can a machine learning model predict the price of Bitcoin or Ethereum for the next 24 hours? (3) How does the volatility of Bitcoin compare with other cryptocurrencies over time? and (4) How do significant spikes in trading volume or sudden price changes impact short-term price movements of Bitcoin and Ethereum?

The structure and analysis of this report are organized around answering these four questions. Through exploratory data analysis, statistical correlation, time series forecasting, and machine learning modeling, the project aims to uncover actionable insights into the behavior of crypto markets. In doing so, it not only evaluates the predictive power of various features and algorithms but also investigates patterns in market volatility and short-term reactivity to trading spikes. Ultimately, this project contributes to the growing body of research on cryptocurrency forecasting and offers a data-driven perspective on the complexities of modeling digital asset markets.Datasets

A. *Source of dataset (Heading 2)*

The primary dataset is sourced from Kaggle and contains minute-by-minute historical price data for Bitcoin and Ethereum dating back to 2014. This high-resolution dataset enables granular time series analysis and is well-suited for training machine learning models to predict short-term price changes. The dataset includes key attributes such as timestamps, opening and closing prices, high and low values, and trade volumes. To complement the Kaggle data, additional datasets are obtained from CryptoDataDownload, which provides historical OHLC and volume data at daily, hourly, and minute-level intervals across various exchanges. These datasets offer flexibility for analyzing longer-term trends and comparing market behaviors across different time frames.

Example: XXXX

B. *Character of the datasets*

This project incorporates four main datasets to analyze cryptocurrency price dynamics and the effect of external events. The primary Kaggle dataset includes approximately 13 million rows of minute-level price data for Bitcoin and Ethereum and is around 480 MB in size. Key columns include Timestamp, Open, High, Low, Close, and Volume, all in USD and units traded. The CryptoDataDownload datasets complement this with lower-frequency OHLC data, allowing for longer-term trend analysis. Yahoo Finance adds roughly 2,500 rows of daily ETH-BTC pricing data with standard financial fields such as Open, Close, Adj Close, and Volume. To incorporate external sentiment, the Crypto News+ dataset includes over 280,000 news articles, with columns such as Date, Title, Text, Subject, Sentiment, and Source. Sentiment is categorized as Positive, Neutral, or Negative.

All datasets were cleaned and standardized before analysis. Timestamps were converted to a unified UTC format using pandas.to_datetime(). Missing values in critical columns were either forward-filled or dropped. ETH-BTC prices were converted to USD using concurrent BTC/USD rates from the primary dataset. Datasets were merged on Timestamp or Date using inner and outer joins depending on the analysis granularity. New features were created to enhance analysis, including Price Change %, rolling volatility (7-period standard deviation of log returns), Daily Sentiment Avg, 10-day and 30-day moving averages, and a binary External Event Flag marking dates associated with major headlines or news spikes. These transformations produced a unified dataset suitable for time series forecasting, volatility comparison, and feature-based correlation analysis.

Example: XXXX

979-8-3315-0993-4/25/$31.00 ©2025 IEEE

II. Methodology

In this part, you should give an introduction of the methods/model. First, what's the method/model. What's the assumption of this method/model. What's the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extra work did you adjust to make the results better. If you have multiple methods, feel free to use subsection A., B. to separate them.

Example: Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

A. *Method A*

Example: The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

$$a + b = \gamma \quad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

B. *Method B*

- Bulletin 1
- Bulletin 2.
- Bulletin 3

C. *Method C*

Example: The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

$$a + b = \gamma \quad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

An excellent style manual for science writers is [7].

III. Results

In this section, present your findings using an appropriate method, such as equations, numerical summaries, or visualizations like charts and graphs. Clearly explain all results and provide guidance on how to interpret them. If any unexpected results arise, discuss possible reasons or contributing factors. To improve clarity and organization, consider using subsections (e.g., A, B) to separate different aspects of your results.

Example: After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. *Result A*

Example: XXX

1) For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

2) For papers with less than six authors: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) Deletion: Delete the author and affiliation lines for the extra authors.

B. *Results B*

Example: Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

C. *Results C*

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. Table Type Styles

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a] Sample of a Table footnote. (*Table footnote*)

Fig. 1. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

IV. Discussion

Every method/project has its shortage or weakness. Please discuss the unsatisfied results in your project. And discuss the

Identify applicable funding agency here. If none, delete this text box.

feasible suggestions of future work to revise/improve your result.

Example: xxx

V. Conclusion

In this part, you should summarize your project. What important results did you find for your topic and what's the effect of this result on the real-world?

Example: xxx

Acknowledgment *(Heading 5)*

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

Use the IEEE format for the citation. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..." Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.