Predicting Cryptocurrency Prices Using Historical Market Indicators and News

Detailed Project Plan

Group: FYP24043

Members:

Guo Yue 3035835014 BASc (Fintech)

Wong Man Lok 3035931222 BEng (CompSc)

Zhu Suying 3035845710 BEng (CompSc)

Project Background

Cryptocurrencies have gained significant attention in recent years, with their prices demonstrating high volatility [1]. Accurately predicting cryptocurrency prices is of great interest to investors and traders. Previous studies have explored the use of historical market indicators, such as past prices, to forecast future cryptocurrency prices using machine learning techniques [2]. However, the impact of news events on cryptocurrency prices cannot be overlooked. A notable example is when Elon Musk replaced Twitter's bluebird icon with the Shiba Inu digital currency symbol, causing the price of Dogecoin to surge by over 30% [5].

While some studies have investigated the effect of news sentiment on cryptocurrency prices independently, the results have been mixed, with some findings suggesting it is effective [6] and others indicating otherwise [7]. Moreover, the combined impact of news itself and technical indicators on cryptocurrency prices has not been extensively explored.

To address this gap, we propose a project that aims to develop an effective model for predicting cryptocurrency prices by considering both historical technical indicators and news data. By leveraging advanced machine learning techniques such as Long Short-Term Memory (LSTM) and Transformers, we seek to capture the complex relationships between news topics, market sentiment, and price movements. The objective is to create a predictive model that outperforms existing approaches by incorporating a comprehensive set of features and employing advanced machine learning algorithms.

Through this project, we aim to provide valuable insights into the factors influencing cryptocurrency prices and contribute to the development of more accurate and reliable price prediction models. The findings of this study could have significant implications for investors, traders, and researchers in the field of cryptocurrency market analysis.

Project Objective

This project aims to achieve two key objectives:

1. **Cryptocurrency Price Prediction:**
    - Employ machine learning techniques, such as Long Short-Term Memory (LSTM) networks and Transformer architectures to predict future prices with historical cryptocurrency market indicators and relevant news articles.
    - Train and fine-tune the models to accurately forecast cryptocurrency prices over specified time horizons.
    - Evaluate the performance of the predictive models using appropriate metrics and validation techniques to ensure their reliability and effectiveness.

2. **User Interface Development:**
    - Design and implement an intuitive and user-friendly interface that allows users to interact with the cryptocurrency price prediction model seamlessly.
    - Incorporate a feature that allows users to select a desired range of days for price prediction, such as 7 days, 14 days, or 21 days, providing flexibility and customization options.
    - Display the predicted prices on the user interface in a visually appealing and easily interpretable format, such as charts or graphs, to facilitate understanding and analysis.

The combination of a robust predictive model and an intuitive user interface will provide a proof of concept for the practical application of the model, enabling users to make informed decisions and gain valuable insights into the cryptocurrency market.

Project Methodology

To achieve the project goals, the methodology encompasses the following stages:

A. Data Collection and Preprocessing

The initial phase of our project involves a systematic approach to collecting cryptocurrency market data and relevant news. We will begin by determining the evaluation metrics for our data sources, such as data scope, volume, granularity, source credibility, and relevance to specific cryptocurrencies. These metrics will guide us in selecting the most suitable data sources for our project.

After determining the metrics, we will finalize the data source to obtain the required dataset. We would obtain the datasets from data sources primarily by various APIs, such as CoinMarketCap API or CoinAPI.io for market data and Crypto News API for crypto-related news.

To align with our project objectives and resources, we plan to focus on a single reputable news source that offers a wide range of topics and extensive time coverage. Additionally, we will concentrate our analysis on several leading cryptocurrencies e.g. Bitcoin, Ethereum, Dogecoin, etc.

Throughout the data collection process, we will ensure that the market data and news are aligned in terms of time range and cryptocurrency type. We will also conduct a preliminary study to gain insights from previous research that has successfully utilized similar models. This literature review will further refine our criteria for selecting the most suitable dataset.

By following this structured approach, we aim to collect high-quality, relevant data that will lay a solid foundation for our machine learning model to predict cryptocurrency prices accurately.

B. Sentiment Analysis and Feature Engineering

Feature engineering is a crucial step in preparing the input data for our machine learning models. We will extract and create relevant features from various data sources to capture the factors that potentially influence cryptocurrency prices.

Our raw data will include historical price data and other market indices, as well as news articles covering a wide range of topics such as entertainment, medicine, technology, and politics. By incorporating diverse news topics, we aim to uncover interesting patterns and relationships that may not be immediately apparent.

From the historical price data, we will derive technical indicators such as moving averages, RSI (Relative Strength Index), MACD (Moving Average Convergence Divergence), and Bollinger Bands. These indicators can provide insights into price trends, momentum, and volatility. The indicators will be determined after we do a thorough literature review.

To leverage the news data, we will employ sentiment analysis techniques to quantify the sentiment expressed in each article. Sentiment scores will be used as features to capture the overall market sentiment. Additionally, we will explore other potential features derived from the news data using techniques such as topic modeling, keyword extraction, and named entity recognition. These features may help identify specific events, themes, or entities that have an impact on cryptocurrency prices.

To handle the high-dimensional feature space resulting from the diverse data sources, we will apply dimensionality reduction techniques such as Principal Component Analysis (PCA). PCA will help identify the most important features and reduce the computational complexity of our models.

By generating an extensive set of features from various data sources, we lay the foundation for feature selection in subsequent stages of the project. The selected features will be used to train and test our machine learning models, enabling us to identify the most informative and predictive variables for cryptocurrency price prediction.

C. Model Development and Validation

The model development phase will explore various machine learning techniques suitable for time-series forecasting, focusing on three representative cryptocurrencies: Bitcoin, Ethereum, and Dogecoin. These cryptocurrencies were chosen for the following reasons:

- Bitcoin and Ethereum: As the two cryptocurrencies with the largest market capitalization, they represent the top coins from different blockchain ecosystems.
- Dogecoin: A popular coin forked from the Bitcoin blockchain, known for its volatility and sensitivity to news events. Its inclusion allows for comparison with Bitcoin and represents coins with high media influence.

This focused approach allows us to:

- Simplify data gathering and processing
- Concentrate on developing robust prediction models
- Draw conclusions that may be applicable to a wider range of cryptocurrencies

We will investigate models such as LSTM (Long Short-Term Memory networks), CNN (Convolutional Neural Networks), transformer and other relevant algorithms to determine the most effective approach for cryptocurrency price prediction. The models to be applied will be determined after a thorough literature review.

Our primary objective is to train a model for each of the three selected cryptocurrencies that can predict future price trends based on historical data and news. We will experiment with multiple input-output configurations, represented as (X, Y), where X is the number of input days and Y is the number of days to predict. Configurations such as (30, 1), (30, 7), (30, 14), (45, 1), (45, 7), and (45, 14) will be evaluated to identify the most accurate combination for each cryptocurrency.

To maximize the utilization of available data, we will employ a rolling window approach for data preparation. For instance, if we have data from January 1 to December 31, we might create training samples as follows:

- January 1 - January 30 (input) to predict January 31 – January 6 (output)

- January 2 - January 31 (input) to predict February 1 – January 7 (output)

This method will generate a substantial number of training samples from our raw data for each cryptocurrency.

The dataset for each cryptocurrency will be divided into training, validation, and test sets to ensure model generalization. The training process will involve iterative feature selection, hyperparameter tuning and validation to optimize performance. Model evaluation will utilize metrics such as RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) to assess prediction accuracy.

By focusing on these three diverse yet representative cryptocurrencies, we aim to develop a robust methodology that can potentially be extended to other cryptocurrencies in future research. This approach allows us to thoroughly investigate the nuances of each selected coin while managing the complexity of data collection and model training.

D. Implementation of Real-Time Predictive System

The trained model will be integrated into a backend system capable of serving real-time predictions. This system will be developed using Python with frameworks such as Flask or Django for efficient implementation.

The real-time aspect of the system will involve fetching the most recent X days of data (where X is the input window size determined during model development) through APIs. This data will include both price information and relevant news articles. The backend will process this information and feed it into the trained model to generate predictions for the specified future period.

E. User Interface Development

We will develop a user-friendly interface using React to make the predictive system accessible. This user interface will display real-time predictions in an interactive format.

Users will have the ability to:

- View predictions for different cryptocurrencies
- Potentially select different (X, Y) combinations for input and prediction periods if multiple models are implemented

The UI will connect the backend to dynamically fetch real-time data and display up-to-date predictions based on the most recent market conditions.

Project Schedule and Milestones

(Subject to Change)

Stage	Objective	Deadline	Remark
Sem 1			
Project Setup (4 Weeks)	Feasibility Assessment	9/22	• Identify the possible data source. • Identify APIs to collect market data and news. • Identify languages and tech stacks applied.
	Environment Setup	10/1	• Select programming languages and tools • Setup development environment and necessary libraries
	Project Website Creation	**10/1**	• Updated with Each Milestone
	Detailed Project Plan	**10/1**	
Preliminary Study (5 Weeks)	Milestone1: Preliminary Research Summary	10/31	• Milestone Deliverable: Preliminary Study Summary (Will be updated on Webpage)
* Simultaneous work on Data Manipulation & ML implementation			
Dataset Manipulation (8 Weeks)	Data Collection/Prepare Dataset	12/31	• Collect historical cryptocurrency market data. • Collect news articles and social media posts related to cryptocurrencies. • Find dataset with cryptocurrency market data and news
	Data preprocessing /cleaning	12/31	• Brainstorm as many features as possible • Generate some features by LLM • Cleaning, normalization, and structuring for future analysis.

ML Implementation (8 Weeks)	Train models using both historical market indicators and news	12/31	• Choose appropriate machine learning models based on preliminary study. • Split data into training, validation, and test sets.
	Milestone 2: Pave the path for research	12/31	• Ensure models can be trained with limited data and limited features
Sem 2			
Midterm Wrap-up (3 weeks)	**First Presentation**	**1/13**	• Midterm Paperwork • No workload allocated in winter holiday
	Interim Report	**1/26**	
* Simultaneous work on ML Model Reinforcement and Testing & UI Implementation			
ML Model Reinforcement and Testing (8 Weeks)	Feature Selection & Models Comparison		• Try different feature groups and find the most relation features • Try different combinations of tested and predicted time and find the model with the best combination
	Reinforce Models		• Implement advanced techniques like cross-validation to avoid overfitting. (To be confirmed by the assessment result) • Iterative work on feature selection, hyperparameter tuning and validation to optimize performance.
	Milestone 3: Finalize ML Model	3/17	• Milestone Deliverable: Predictive models which are well developed and evaluated
UI Implementation (8 Weeks)	Frontend Development		• Design UI and Implement the frontend with React
	Integration with ML		• Connect the frontend with the machine learning model through a backend in

	models		Django / Flask
	Milestone 4: Finish all coding deliverables	3/17	• Milestone Deliverable: A functional UI developed and integrated with the models.
Final Wrap-Up (5 Weeks)	**Final Individual Report**	**4/21**	• Final Paperwork • Another extra report focusing on the project details should be written by the group
	Final Presentation	4/21	
	Final Project Website	4/30	
	Final Report	5/30	

References:

[1] A. Mishra, V. Gupta, S. Srivastava, A. K. Pandey, L. Kumar, and T. Choudhury, “Social Media Role in Pricing Value of Cryptocurrency,” *Machine Intelligence and Data Science Applications*, pp. 819–829, 2022, doi: https://doi.org/10.1007/978-981-19-2347-0_63.

article=1039&context=datasciencereview

[2] F. Sabry, W. Labda, A. Erbad, and Q. Malluhi, “Cryptocurrencies and Artificial Intelligence: Challenges and Opportunities,” *IEEE Access*, vol. 8, pp. 175840–175858, 2020, doi: https://doi.org/10.1109/access.2020.3025211.

[3] Cryptocurrency price prediction based on multiple market sentiment, semantic scholar, (accessed Oct. 1, 2024).

[4] J. Abraham, D. Higdon, J. Nelson, J. Ibarra, and J. Nelson, “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis,” *SMU Data Science Review*, vol. 1, no. 3, 2018, Available: https://scholar.smu.edu/cgi/viewcontent.cgi?

[5] R. G. Kolodny, “Dogecoin Jumps More than 30% after Musk Changes Twitter Logo to Image of Shiba Inu,” *CNBC*, Apr. 03, 2023. https://www.cnbc.com/2023/04/03/dogecoin-jumps-over-30percent-after-twitter-changes-logo-to-doges-symbol.html

[6] S. Penmetsa and M. Vemula, "Cryptocurrency Price Prediction with LSTM and Transformer Models Leveraging Momentum and Volatility Technical Indicators," 2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA), Dalian, China, 2023, pp. 411-416, doi: 10.1109/ICDSCA59871.2023.10393319.

[7] Zhao, H.; Crane, M. and Bezbradica, M. (2022). Attention! Transformer with Sentiment on Cryptocurrencies Price Prediction. In Proceedings of the 7th International Conference on Complexity, Future Information Systems and Risk - COMPLEXIS; ISBN 978-989-758-565-4; ISSN 2184-5034, SciTePress, pages 98-104. DOI: 10.5220/0011103400003197