**Cryptocurrency Prediction Using FinBERT and Time-Series Methods**

**Date:** 09-02-2024

**Prepared By:**

1. Sai Harsha Vardhan Reddy, Kolan – qg2321
2. Venkata Sai Preethi, Chilakala – hb5542
3. Sandeep, Jakkula – qg1264
4. Azmat Ahmed Shaik – ub9706
5. **Executive Summary**

The cryptocurrency market, characterized by its extreme volatility, offers both substantial risks and opportunities for investors. Accurately predicting price movements of leading cryptocurrencies is essential for developing robust investment strategies. This project focuses on exploring the predictive power of two distinct methodologies—FinBERT, a specialized natural language processing model for financial text, and advanced time-series analysis techniques—applied to five major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), Solana (SOL), and Ripple (XRP).

**Problem and Opportunity:** The volatile price movements of cryptocurrencies pose a significant challenge for investors striving to manage risk and maximize returns. Traditional forecasting models may capture some of the complex factors influencing these markets, but this project seeks to determine whether FinBERT's ability to analyse sentiment from financial news can yield accurate predictions for these critical cryptocurrencies. Additionally, we will develop time-series models that account for trend and seasonality in forecasting these assets, and we will compare the effectiveness of these approaches.

**Expected Outcomes and Impact:** This project aims to achieve a comprehensive understanding of the individual strengths and limitations of sentiment analysis using FinBERT and time-series forecasting for predicting cryptocurrency price movements. By evaluating each method independently, the exploration seeks to determine the accuracy and effectiveness of FinBERT's sentiment analysis and advanced time-series techniques in forecasting the price movements of the specified cryptocurrencies. The insights gained are expected to offer valuable perspectives on the capabilities of these methodologies in the context of cryptocurrency forecasting. This understanding could enhance future investment strategies by providing investors with more nuanced tools for decision-making. Additionally, the findings could contribute significantly to the broader discourse on financial forecasting methodologies, especially in the dynamic and rapidly evolving cryptocurrency sector.

1. **Project Background and Description**

**Context:** The cryptocurrency market is marked by rapid growth, extreme volatility, and complex dynamics that pose significant challenges for accurate price prediction. Traditional financial forecasting models, which often rely on historical data and linear patterns, may not fully capture the unique characteristics of cryptocurrencies. This project aims to bridge this gap by leveraging advanced methodologies for more precise predictions. Specifically, we will explore the potential of FinBERT, a specialized natural language processing model designed for financial text and compare it with advanced time-series forecasting techniques.

**Problem Statement**: The unpredictability of cryptocurrency prices is influenced by various factors, including market sentiment and historical trends. Traditional forecasting methods may not adequately address the rapid impact of news sentiment or the non-linear, volatile nature of cryptocurrency price movements. This project seeks to address these limitations by evaluating the effectiveness of FinBERT in analysing sentiment from financial news related to cryptocurrencies. We will develop a custom neural network using FinBERT embeddings and a linear layer to predict price movements for five major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), Solana (SOL), and Ripple (XRP).

**Objective Specificity:** The project aims to rigorously evaluate the predictive power of two distinct approaches—sentiment analysis using FinBERT and advanced time-series forecasting techniques. The specific objectives are:

1. Data Collection and Preprocessing:
   * Acquire and preprocess financial news articles and historical price data for BTC, ETH, BNB, SOL, and XRP.
   * Implement data cleaning and transformation processes to ensure high-quality inputs for model training.
2. Method Application:
   * Sentiment Analysis: Utilize FinBERT to analyse sentiment from the collected news articles and train a custom neural network incorporating FinBERT embeddings with a linear layer for prediction.
   * Time-Series Forecasting: Apply advanced time-series models, such as ARIMA, Holt-Winters, and others, to forecast price trends based on historical data.
3. Insight Generation:
   * Evaluate the accuracy and effectiveness of sentiment analysis and time-series forecasting in predicting cryptocurrency prices.
   * Compare the performance of both methodologies to determine their strengths and limitations in the context of cryptocurrency forecasting.

**3. Scope of Work**

**Data Sources:**

1. The Crypto News dataset is available in the following repository:

<https://github.com/soheilrahsaz/cryptoNewsDataset.git>

1. The historical prices of the specified cryptocurrencies can be imported from the following website:

<https://coincodex.com/crypto/dogs/historical-data/>

**Data Specifications:**

* The repository mentioned above contains the crypto news which are fetched from the Cryptopanic.com. The table below shows the count and time range for the cryptocurrencies available in the dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| code | cnt | minDatetime | maxDatetime |
| BTC | 14540 | 2017-09-29 13:30:08 | 2024-07-12 08:58:20 |
| ETH | 10299 | 2017-09-23 10:00:42 | 2024-07-12 08:19:33 |
| SOL | 4684 | 2021-02-10 20:02:39 | 2024-07-12 08:19:33 |
| ADA | 4260 | 2018-01-11 21:35:52 | 2024-07-12 07:41:22 |
| XRP | 4202 | 2017-12-13 16:00:22 | 2024-07-12 08:00:00 |
| SHIB | 3640 | 2021-05-10 10:02:46 | 2024-07-12 07:18:35 |
| DOGE | 3628 | 2018-02-18 10:00:05 | 2024-07-12 07:18:35 |
| USDT | 2653 | 2018-01-27 23:43:00 | 2024-07-12 07:00:24 |
| MATIC | 2279 | 2019-05-22 19:06:25 | 2024-07-12 07:39:09 |
| BNB | 1959 | 2018-11-27 17:17:02 | 2024-07-11 21:00:51 |
| LINK | 1849 | 2020-02-28 13:00:30 | 2024-07-11 21:38:51 |
| AVAX | 1784 | 2021-01-24 22:15:44 | 2024-07-11 21:00:00 |
| DOT | 1730 | 2020-08-29 08:39:09 | 2024-07-12 06:24:59 |
| TRX | 1652 | 2018-01-08 06:00:21 | 2024-07-11 23:10:13 |
| UNI | 1580 | 2020-09-21 08:58:52 | 2024-07-12 00:00:27 |

* We can get past 5 years prices of these currencies from the coincodex.com website.
* By performing data manipulation and preprocessing, we will integrate information from these sources to support our analysis.
* We are going to use monthly level granularity for our analysis.

**Analysis Techniques:**

1. **Data Preprocessing and Transformations:**

* **Tools:** Python, SQL
* **Methods:**
* Use SQL queries to merge tables, perform joins, aggregations etc.
* Use pandas from python for data cleaning and preparation.

1. **Time-Series Forecasting:**
   * **Tools:** R/Python
   * **Methods:** 
     + Linear Regression with Trend and Seasonality
     + Linear Regression with Trend and Seasonality + Moving Average for residuals
     + Linear Regression with Trend and Seasonality + AutoRegressive for residuals
     + Holt-Winters (Exponential Smoothing)
     + ARIMA (AutoRegressive Integrated Moving Average)
2. **Prediction with FinBERT:**
   * **Tools:** Python
   * **Methods:**
     + Hugging Face Transformers Library: Obtain attention-enhanced embeddings from FinBERT, a large language model designed for financial text.
     + Pytorch: Implement a custom neural network using FinBERT embeddings for predicting cryptocurrency price movements.
     + Faiss (Facebook AI Similarity Search Library): Perform semantic searches to identify the top k news headlines most relevant to cryptocurrency price predictions.

**Deliverables:**

1. **Comprehensive Report:** A detailed report documenting the methodology, analysis, results, and insights from both FinBERT sentiment analysis and time-series forecasting techniques. This report will include performance metrics, comparative analysis, and recommendations for future work.
2. **Predictive Models:** Model weights for the custom neural network, as well as time-series models.
3. **Visualizations:** Interactive visualizations showcasing the results of the two methods. These will include graphs of historical prices, predicted trends, and sentiment impacts.
4. **Presentation:** A summary presentation highlighting key findings, insights, and recommendations from the project, suitable for stakeholders or presentations.

**4. Timeline and Action Plan**

|  |
| --- |
| S.NO Phase Start End |
| 1. Data Preprocessing and Transformations 09/04/2024 09/25/2024 |
| 1. Time-Series Forecasting 09/26/2024 10/17/2024 |
| 1. Prediction using FinBERT 10/18/2024 11/07/2024 |
| 1. Creating Report 11/08/2024 11/17/2024 |
| 1. Creating Presentations 11/18/2024 11/25/2024 |
|  |

**5. Evaluation and Performance Metrics**

Measuring and comparing the predictive accuracy of both methods using 3 key metrics

1. MSE (Mean Squared error)
2. MAE (Mean Absolute error)
3. RMSE (Root Mean Squared error)
4. **Appendices and References**
5. Oikonomopoulos, Sotirios, et al. "Cryptocurrency price prediction using social media sentiment analysis." *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2022.
6. Frohmann, Markus, et al. "Predicting the price of Bitcoin using sentiment-enriched time series forecasting." *Big Data and Cognitive Computing* 7.3 (2023): 137.
7. Pant, Dibakar Raj, et al. "Recurrent neural network based bitcoin price prediction by twitter sentiment analysis." *2018 IEEE 3rd international conference on computing, communication and security (ICCCS)*. IEEE, 2018.
8. Lee, Hokyung. "FINANCIALBERT: PREDICTING STOCK PRICE WITH BERT AND SEMANTIC ANALYSIS." (2023).