**Project Follow-Up-Report**

**Title of Project**: Cryptocurrency Prediction Using FinBERT and Time-Series Methods

**Date of Report**: 09-26-2024  
**Team Members**:

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

**1. Project Overview:**  
The project embarks on an innovative journey to assess the predictive potential of both FinBERT sentiment analysis and cutting-edge time-series forecasting for cryptocurrency markets. Key objectives include:

1. **Data Acquisition & Preparation:** Gather and preprocess a rich dataset of crypto news and historical price data for around 400 crypto currencies, ensuring high-quality inputs for modelling.
2. **Methodology:**
   * Sentiment-Driven Predictions: Build a custom neural network for crypto price predictions leveraging FinBert a transformer-based state-of-art-model fine-tuned on financial data.
   * Time-Series Mastery: Apply sophisticated time-series models- LSTM, Prophet, and others—to predict trends from historical data.
3. **Performance Insights:**
   * Evaluate and compare the accuracy of both approaches across Bitcoin, Ethereum, Binance Coin, Solana, and Ripple, highlighting each method's unique advantages in cryptocurrency forecasting.

This comprehensive analysis promises to uncover powerful insights into how sentiment and time-series methods can drive more informed predictions in the volatile crypto landscape.

**2. Progress Made:**

* **Milestones Achieved**: Successfully pre-processed approximately 153,000 cryptocurrency-related news headlines, transforming them into a format ready for modelling. Additionally, historical price data for around 400 cryptocurrencies was sourced from the Coin Codex website and refined for model training. The news and price datasets were logically merged into JSON files structured as follows:

1. **News Format**

{"currency": "-------",

"month\_year": "-------",

"data": [{"title": "……"}]

}

1. **Prices Format**

{"currency": "-------",

"month\_year": "-------",

"data": {"this\_month\_price": "-------",

"next\_month\_price": "-------",

"percentage\_change": "--------"

}}

* **Tasks Completed:**

1. **get\_news.SQL**: Established the news database on the MySQL server and developed the SQL script to efficiently retrieve cryptocurrency news along with the relevant currency codes.
2. **crypto-Datapreprocessing-1**: Created a Jupyter notebook to preprocess and transform the news data obtained from the previous step, ensuring it was in an optimal format for modeling.
3. **crypto-Datapreprocessing-2**: Developed a second Jupyter notebook to preprocess and transform the price data sourced from the Coin Codex website, preparing it for integration with the news dataset.

The completion of the above tasks resulted in the creation of two logically connected JSON files: **training\_data.json** and **price\_history\_data.json**. The formats for these files have already been highlighted above.

**3. Challenges Encountered:**

* **List of Challenges**:

1. The news dataset was quite disorganized, with each news title linked to multiple currencies. Grouping the news according to currency codes proved to be a challenging task when a single news item was associated with several currencies.
2. Another significant challenge arose from the fact that the currency prices were fetched from different source. Ensuring that the time frame of the news data and price data matched was a difficult task.
3. The prices obtained from the website resulted in 400 unique files each corresponding to different currency code. Manually reviewing all the files and consolidating them according to currency codes into a single file is time consuming.

* **Steps Taken to Overcome Challenges:**

1. Employed the pandas library in Python to efficiently handle rows associated with multiple currencies by splitting them apart, allowing for effective grouping based on currency codes and time frames.
2. To tackle the issue of mismatched time frames resulting from data sourced from coin codex website, we organized group meetings to collaboratively determine an appropriate data structure. This consensus led to a uniform preprocessing approach to ensure consistency across the datasets.
3. Created a custom Python script that systematically processes all 400 price data files, organizing them by currency code into the desired data structure format, thereby streamlining the consolidation process.

**4. Member Contributions**

* **Venkata Sai Preethi, Chilakala:**

1. Played a key role in establishing the MySQL database environment and developed the SQL script to retrieve news data from the database.
2. Retrieved price data for 100 currencies from the Coin Codex website.

* **Sandeep, Jakkula**

1. Assisted in Data Preprocessing-1 by utilizing pandas for data transformations and effectively grouping the rows.
2. Collected price data for 100 currencies from the Coin Codex website.

* **Azmat Ahmed Shaik**

1. Contributed to Data Preprocessing-2 by developing an automated Python script that consolidates data from 400 CSV files into a single data frame.
2. Gathered price data for 100 currencies from the Coin Codex website.

* **Sai Harsha Vardhan Reddy, Kolan**

1. Led the project to ensure it progressed effectively.
2. Contributed to both Data Preprocessing-1 and Data Preprocessing-2, ensuring the data was structured correctly and facilitating the creation of the JSON files.
3. Sourced price data for 100 currencies from the Coin Codex website.

**5. Next Steps:**

* **Upcoming Tasks:** Data preprocessing is a critical step in predictive modelling. With the data now prepared, our next task will be to train the machine learning models.
* **Goals for Next Report:** We aim to explore two variants of sentiment-based predictions: 1. FinBERT and 2. FinBERT using FAISS.

**6. Additional Comments:**

The team has shown remarkable collaboration and commitment throughout the project, effectively navigating challenges and leveraging each member's strengths. Regular check-ins and open communication have facilitated the exchange of ideas, leading to innovative solutions and a positive team dynamic. As we move forward, maintaining this collaborative spirit will be essential, especially as we delve into model training and exploration of advanced techniques. The enthusiasm for learning and experimenting with different methodologies is encouraging, and we look forward to sharing significant progress in our next report.