Project Kick-off Meeting Agenda: Bitcoin Price Prediction

Attendees: Nithin (team leader), Huan, Ognjen

Date: [11.2.2025]

Goal: To create a concrete action plan based on our project proposal and divide initial responsibilities.

1. Confirm Project Scope

- **Goal:** Make sure everyone is 100% on the same page.
- **Discussion:**
    - Briefly review our Bitcoin_project-proposal.pdf.
    - **Confirm Goal:** We are predicting daily/weekly closing prices.
    - **Confirm Features:** We need to get three types of data:
        1. **Market Data:** (Prices, Volume)
        2. **On-Chain Data:** (Hash Rate, Mining Difficulty, Tx Volume)
        3. **Macroeconomic Data:** (Interest Rates, S&P 500, Gold)
    - **Confirm Models:** We will compare ARIMA, KNN, LSTM, XGBoost, and LightGBM.
    - **Confirm Metrics:** We will judge success using RMSE, MAE, R-squared, and Directional Accuracy.

2. Data Collection Plan

- **Goal:** Figure out *exactly* how to get the data listed in our proposal.
- **Discussion:**
    - **Market Data (CoinGecko, Yahoo Finance):** How do we get this? Do we need an API key for CoinGecko? Can we download CSVs from Yahoo Finance?
    - **On-Chain Data:** Where is the best source for *historical* hash rate, difficulty, etc.? (e.g., Glassnode, Blockchain.com). Do they have free CSV downloads or an API?
    - **Macroeconomic Data:** This should be easy. We can probably find CSVs for historical S&P 500 (ticker: ^GSPC), Gold (ticker: GC=F), and US Treasury interest rates (from the Federal Reserve's FRED database).
- **Action Item:** Assign one person (or divide the task) to find and download all these data sources *before the next meeting*.

3. Division of Labor

- **Goal:** Split the entire project into three clear roles.
- **Discussion:** A 3-person team has a natural split. Which role does each person want?
    - **Role 1: The "Data Engineer"**
        - **Tasks:**

        - Collects all the data from the sources we just discussed.
        - Writes the Python (Pandas) script to clean *everything*.
        - Handles all missing values (as discussed in class slides, e.g., 10 - DataImputation.pdf).
        - Merges all three datasets (market, on-chain, macro) into **one final, clean CSV file** aligned by date.
        - Normalizes the data (e.g., using StandardScaler from class).
      - **Deliverable:** The final dataset_clean.csv that the team will use.
    - **Role 2: The "Modeling Engineer"**
      - **Tasks:**
        - Takes the dataset_clean.csv from Role 1.
        - Writes the Python scripts to implement all 5 models (ARIMA, KNN, LSTM, XGBoost, LightGBM).
        - Correctly implements the **Time-Series K-Fold Cross-Validation** (from slide 09 - Forecasting part3.pdf—this is critical!).
        - Trains all models and saves their predictions.
      - **Deliverable:** The prediction results (e.g., predictions.csv).
    - **Role 3: The "Analyst & PM"**
      - **Tasks:**
        - Takes the predictions.csv from Role 2.
        - Writes the script to calculate all the final evaluation metrics (RMSE, MAE, R2, DA).
        - Creates the final comparison tables (like in slide 09 - Forecasting part3.pdf, p. 30) and visualizations.
        - Acts as the project manager, keeping everyone on track.
        - Leads the writing of the final paper and creation of the presentation slides.
      - **Deliverable:** The final results, the paper, and the presentation.

4. Paper Structure & Timeline

- **Goal:** Agree on the paper's structure and set deadlines.
- **Discussion:**
  - **Writing Assignments:**
    - Role 1 (Data Engineer) writes the "Data and Methodology"
    - Role 2 (Modeling Engineer) writes the "Predictive Models" section
    - Role 3 (Analyst) writes the "Introduction," "Related Work," "Results," and "Conclusion"
  - **Set Deadlines:**
    - When does Role 1 need to have the clean data ready? (e.g., 1 week)
    - When does Role 2 need to have the model results ready? (e.g., 1 weeks after that)

        - When is the first draft of the full paper due? (e.g., 1 weeks after that)

5. Next Steps

- **Goal:** Make sure everyone knows exactly what to do *before the next meeting*.
- (Example)
    - **Ognjen:** Will take Role 1. Will find and download all data sources
    - **Huan:** Will take Role 2. Will start researching the Python libraries for ARIMA and LSTM.
    - **Nithin:** Will take Role 3. Will set up a shared Google Doc from the paper draft and a GitHub repository for our code.