High Level Design (HLD)

Project Title:

Cryptocurrency Liquidity Prediction for Market Stability

Objective:

The goal of this project is to predict the liquidity ratio of cryptocurrencies based on features like price, 24-hour trading volume, market capitalization, and volatility. This helps in understanding how easily cryptocurrencies can be bought or sold, thus improving market stability.

Overview:

The project involves:

- Data collection from two CoinGecko datasets (March 16 and March 17, 2022).

- Data cleaning by removing missing values and duplicates.

- Feature engineering: Creating moving averages, volatility, and liquidity ratio features.

- Exploratory Data Analysis (EDA): Understanding trends and relationships between variables.

- Model building:

  - Linear Regression (basic model)

  - Random Forest Regressor (final model after hyperparameter tuning)

- Model evaluation: Using metrics like RMSE, MAE, and R² Score.

- Saving the best model for future predictions.

Major Components:

1. **Data Ingestion**: Loading CSV files into Pandas DataFrames.

2. **Data Cleaning**: Handling missing values, duplicate removal, data type conversion.

3. **Feature Engineering**: Creation of new columns like `price_ma_2`, `mkt_cap_ma_2`, `volatility`, and `liquidity_ratio`.

4. **EDA**:

   ○ Plotting Bitcoin price over time.

   ○ Correlation heatmap between numeric features.

   ○ Summary statistics.

5. **Model Training and Tuning**:

   ○ Initial training with Linear Regression.

   ○ Hyperparameter tuning and training Random Forest Regressor.

6. **Model Evaluation**:

   ○ RMSE, MAE, R² Score calculated.

7. **Model Saving**: Final model saved as `liquidity_prediction_model.pkl` using Joblib.

8. **Deployment (optional)**: A simple Streamlit app for local testing.

Tools and Technologies:

- Python

- Pandas, NumPy

- Matplotlib, Seaborn

- Scikit-learn

- Joblib

- (Optional) Streamlit/Flask for local deployment

High Level Design (HLD) Diagram for Cryptocurrency Liquidity Prediction Project

Start
↓
Loading Dataset (CoinGecko CSV files)
↓
Cleaning Data (Remove Nulls, Duplicates)
↓
Feature Engineering (Moving Avg, Volatility, Liquidity Ratio)
↓
EDA (Bitcoin Price Graph, Correlation Heatmap)
↓
Model Building (Linear Regression → Random Forest)
↓
Model Evaluation (RMSE, MAE, R2)
↓
Saving Final Model
↓
(Optional) Deploy Using Streamlit
↓
End