Project Documentation: HLD & LLD

# 1. High-Level Design (HLD)

The goal of the project is to predict cryptocurrency liquidity to ensure market stability using historical data and machine learning models. The system takes in features like trading volume, market cap, price, and social media sentiment to predict a liquidity score.

Major Components:

- Data Collection from CSV (historical data from CoinGecko)

- Data Preprocessing (handling missing values, normalization, feature creation)

- Exploratory Data Analysis (EDA)

- Feature Engineering (liquidity ratio, volume-to-market-cap, etc.)

- Model Building (Linear Models, Tree-based Models, XGBoost)

- Model Evaluation using RMSE, MAE, and R²

- Model Deployment (optional using Streamlit or Flask)

# 2. Low-Level Design (LLD)

🔽 Detailed Breakdown of Each Component:

## 2.1 Data Preprocessing

- Removed nulls using mean/mode  
- Encoded categorical columns (e.g., 'symbol')  
- Normalized numerical columns using MinMaxScaler

## 2.2 Feature Engineering

- Created new features:  
 \* liquidity\_score = volume / market\_cap  
 \* price\_change\_ratio = 24h / price  
 \* is\_stable\_coin = 1 if coin is USDT/USDC else 0

## 2.3 Model Training

- Models used:  
 \* Linear Regression (Baseline)  
 \* Random Forest Regressor  
 \* Gradient Boosting Regressor  
 \* XGBoost Regressor  
 \* SVR  
- Best model selected using GridSearchCV

## 2.4 Evaluation

- Metrics:  
 \* RMSE (Root Mean Squared Error)  
 \* MAE (Mean Absolute Error)  
 \* R² Score

## 2.5 Model Saving

- Saved final XGBoost model using joblib for future predictions