Pipeline Architecture Document

Project Title: Cryptocurrency Liquidity Prediction for Market Stability

Overview:

This document explains the end-to-end data and model pipeline designed to predict market liquidity.

1. Pipeline Stages:

Stage 1: Data Collection

- Data source: CoinGecko historical dataset

- Format: CSV files (2016, 2017)

Stage 2: Data Cleaning

- Remove nulls, handle missing values

- Merge datasets

- Output: cleaned_data.csv

Stage 3: Feature Engineering

- Moving average of volume

- Liquidity index

- Volatility measures

- Output: feature_engineered_data.csv

Stage 4: EDA

- Correlation matrix

- Volume & price distribution

- Trend analysis

Stage 5: Model Development

- Algorithms: Random Forest, Linear Regression, XGBoost

- Feature selection

- Performance comparison

Stage 6: Hyperparameter Tuning

- Grid search for optimal model config

Stage 7: Model Evaluation

- Metrics: RMSE, MAE, R Score

Stage 8: Deployment

- Using Streamlit or Flask for user interaction

- Input: Market parameters

- Output: Predicted liquidity score

2. Automation:

Scripts executed in sequence using notebooks or automation script.

3. Version Control:

- Git/GitHub for source code management

- Final pipeline committed with .pkl model files