Machine Learning Project Cryptocurrency Liquidity Prediction for Market Stability

Submitted by: Rohit Raikwar

Final Report

# 1. Introduction

The rise of cryptocurrencies has introduced a highly dynamic and volatile financial environment. Liquidity, defined as the ability to buy or sell an asset without causing drastic price changes, is a key determinant of market stability. This project focuses on developing a machine learning model to forecast liquidity changes using historical market data. Predicting liquidity levels aids in risk management and contributes to the resilience of crypto markets.

# 2. Background

Cryptocurrency markets are susceptible to rapid price fluctuations due to factors such as low trading volume, limited order book depth, and speculative behaviors. Low liquidity can exacerbate these fluctuations, leading to instability. Institutions and traders benefit greatly from predictive tools that highlight potential liquidity risks. This motivates the development of a robust ML pipeline tailored to the nuances of crypto markets.

# 3. Dataset Description

The dataset comprises historical cryptocurrency data from 2016 to 2017, collected from CoinGecko and other exchange platforms. It includes attributes such as:  
- Timestamp  
- Open, High, Low, Close Prices (OHLC)  
- Trading Volume  
- Market Capitalization  
- Number of Transactions  
Data from multiple CSV files were merged and cleaned for analysis.

# 4. Data Preprocessing

To ensure data consistency, we handled missing values by forward and backward filling. Numerical features were normalized using MinMaxScaler to bring them into the same scale. Data types were standardized, and outliers were checked and addressed based on statistical thresholds.

# 5. Exploratory Data Analysis (EDA)

EDA revealed seasonal trends in price and volume. Correlation matrices showed strong relationships between volume, market cap, and liquidity indicators. Histograms and line charts illustrated how liquidity varied across major cryptocurrencies like Bitcoin and Ethereum.

# 6. Feature Engineering

Derived features include:  
- 7-day and 30-day Moving Averages  
- Rolling Standard Deviation (Volatility)  
- Liquidity Ratios (Volume/Market Cap)  
- Price Momentum indicators  
These features enrich the model with temporal market dynamics.

# 7. Model Selection

Several regression models were evaluated. Random Forest Regressor and XGBoost were chosen for their performance and robustness. Time-based cross-validation was applied to account for temporal ordering in the dataset.

# 8. Model Evaluation

The models were evaluated using:  
- Root Mean Squared Error (RMSE)  
- Mean Absolute Error (MAE)  
- R² Score  
XGBoost achieved the best performance with an RMSE of 0.06 and R² of 0.91.

# 9. Deployment

A Streamlit-based web application was developed to serve the trained model. Users can upload input data and receive predictions on expected liquidity levels. This enables real-time application of the model’s predictions.

# 10. Challenges

Some of the challenges included:  
- Dealing with sparse or missing data from early years  
- Choosing relevant features from highly correlated variables  
- Ensuring the model generalized across different market phases

# 11. Future Work

In the future, we aim to integrate:   
- Real-time data APIs for continuous predictions  
- Deep learning models like LSTMs for temporal forecasting  
- Sentiment analysis from social media as external signals

# 12. Conclusion

This project demonstrates the feasibility of predicting cryptocurrency liquidity using machine learning. The system provides insights into market stability and helps mitigate trading risks. With further refinement, such tools can be integral to future crypto market infrastructure.