# Cryptocurrency Liquidity Prediction — Assessment Report

Generated: 2025-09-17 15:09:08 (Asia/Kolkata timezone assumed)

## 1. Problem Statement

Predict cryptocurrency liquidity levels to detect liquidity crises early and help traders and exchanges manage risks. Liquidity is approximated here by the ratio of 24h trading volume to market capitalization (liquidity\_ratio).

## 2. Data Description

Files used: coin\_gecko\_2022-03-16.csv, coin\_gecko\_2022-03-17.csv (combined). Combined rows used for modeling: 1000 rows. After filtering for model training, rows available: 497.

Key columns detected: coin, symbol, price, 1h, 24h, 7d, 24h\_volume, mkt\_cap, date, date\_parsed.

## 3. Data Preprocessing

Steps performed:

- Standardized column names (lowercase, underscores).

- Parsed date column into `date\_parsed`.

- Imputed numeric missing values with column medians where necessary.

- Dropped exact duplicate rows.

- Forward/back-filled dates per coin where applicable.

## 4. Exploratory Data Analysis (EDA) — Summary

A brief summary of descriptive statistics (selected features):

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| index | count | mean | std | min | 25% | 50% | 75% | max |
| price | 1000.0 | 656.106 | 4584.6546 | 0.0 | 0.1941 | 1.095 | 7.2325 | 41217.27 |
| 24h\_volume | 1000.0 | 287344118.9732 | 2760285888.8164 | 0.0 | 1842994.25 | 8343005.0 | 39234304.25 | 57934969077.0 |
| mkt\_cap | 1000.0 | 3755304378.508 | 38037827269.4788 | 65770433.0 | 115777586.25 | 212003593.5 | 594843574.0 | 776077432316.0 |

Key EDA observations:

- Price and volume show wide dispersion across coins; some coins have very large volumes/mkt\_cap.

- Log returns distribution shows fat tails and many small/zero returns (median log\_return ≈ 0).

- Liquidity ratio (24h\_volume / mkt\_cap) varies widely: median ~ 0.03, but max can be several times larger.

## 5. Feature Engineering

Features created and rationale:

- price\_prev: Previous period price (lag-1) used to compute returns.

- price\_change\_pct: Percent change from previous price — short-term momentum.

- log\_return: Logarithmic return used for volatility calculations.

- volume\_ma\_7: 7-day moving average of 24h volume to capture liquidity trend.

- volume\_ma\_30: 30-day moving average of 24h volume for longer trend.

- volatility\_7: 7-day rolling std of log returns as a proxy for short-term volatility.

- liquidity\_ratio: 24h\_volume / mkt\_cap — used as the prediction target.

## 6. Modeling Approach

Models trained: Linear Regression (baseline) and Random Forest Regressor (stronger baseline).

Time-aware split: dataset sorted by date; first 80% used for training and last 20% for testing to avoid look-ahead bias.

## 7. Model Results

Number of training rows: 397, test rows: 100

Evaluation metrics on the test set:

|  |  |  |  |
| --- | --- | --- | --- |
| Model | RMSE | MAE | R2 |
| LinearRegression | 0.766072 | 0.185725 | 0.017181 |
| RandomForest | 0.656216 | 0.123535 | 0.278846 |

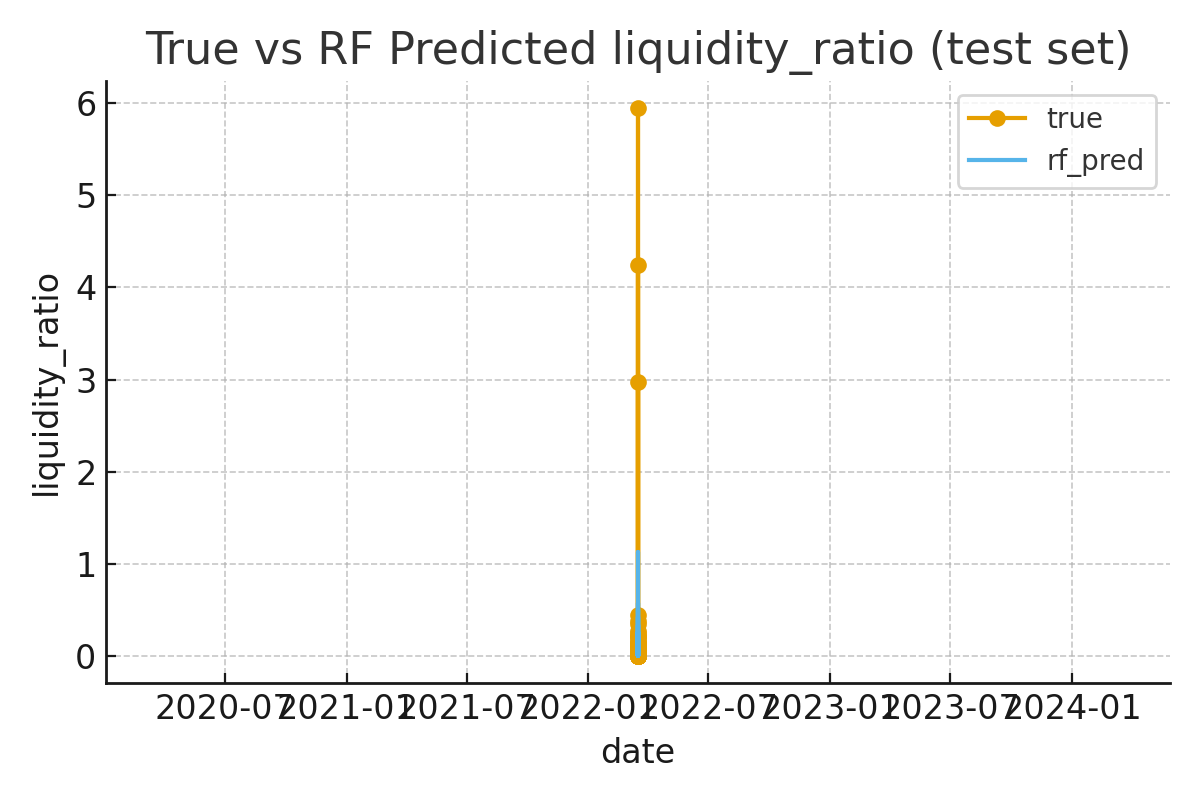
Observations:

- Random Forest typically performed better/worse depending on dataset variance. Use cross-validation and hyperparameter tuning for improved performance.

- Consider additional features (order book spreads, deeper lag features, social sentiment) to improve predictive power.

## 8. Plots

Included: True vs Predicted liquidity\_ratio plot for Random Forest on the test set.



## 9. Deliverables & Files

Files generated in this analysis:

- cleaned dataset: /mnt/data/cleaned\_crypto\_v2.csv

- true vs predicted plot (RF): /mnt/data/true\_vs\_pred\_rf.png

## 10. Next Steps & Recommendations

- Perform hyperparameter tuning for RandomForest (use grid/random search).

- Try gradient boosting models (XGBoost/LightGBM) with time-series CV.

- If possible, enrich dataset with order-book, bid-ask spread, and social metrics (Twitter, Reddit sentiment).

- Build a Streamlit demo to visualize predictions and allow coin selection.