

Final report

This final report summarizes the problem, methodology, results, deployment, and recommendations.

Executive summary

- **Objective:** Forecast next-day cryptocurrency volatility using engineered features and an XGBoost model.
- **Outcome:** A reproducible pipeline with a Streamlit UI that delivers predictions and risk regimes; strong alignment between EDA insights and model features.

Problem statement

- **Context:** Crypto markets exhibit high, regime-dependent volatility.
- **Goal:** Predict next-day GK volatility to support risk management and scenario analysis.

Methodology

Data and features

- **Inputs:** OHLCV + market cap; engineered features (returns, rolling vol, GK, spreads, turnover, calendar).
- **Target:** Next-day GK volatility.

Modeling

- **Algorithm:** XGBoost regressor within a pipeline (scaling/encoding as needed).
- **Validation:** Time-series split (train/val/test).
- **Metrics:** RMSE, MAE, R2 exported for evaluation.

Artifacts and deployment

- **Model artifact:** `models/xgb_volatility.joblib`.
- **App:** Streamlit with Auto (latest data) and Manual input modes; risk regime via 80th percentile of historical GK.

Results

- **Performance:** The metrics such as RMSE, MAE, R2 on validation/test splits are as below respectively:

Split	RMSE	MAE	R2
train	0.00314146005759421	0.0019989543139140417	0.9911122433606081
val	0.010185050293271306	0.006330527946508948	0.8789831981903933
test	0.008808947323285618	0.004846495915074571	0.8964669235606708

- **Qualitative behavior:**
 - Predictions track volatility regimes; higher GK periods flagged as “High” risk.
 - Manual mode enables scenario testing (e.g., higher spreads → higher predicted GK).
- **Visuals:**
 - Historical GK line chart with predicted point for selected asset.
 - Feature importance (optional): XGBoost gain/weight to highlight top drivers (e.g., GK, vol_14d, hl_spread).

Deployment and user experience

- **Modes:**
 - **Auto:** Uses latest features per crypto; guards against empty histories.



Cryptocurrency Volatility Forecast


Forecast next-day volatility using engineered features and XGBoost. Choose between using the latest data or entering your own scenario.

Choose input mode

- ☒ Auto (latest data)
☐ Manual input

Select cryptocurrency

Aave

 Predicted next-day volatility (GK)

0.0373

Risk regime: Normal (80th percentile threshold: 0.0683)



- **Manual:** User-defined feature inputs for scenario analysis.

Liquidity

0.01

-

+

Volume MA 7

1000000.00

-

+

Volume MA 14

1000000.00

-

+

Volume Volatility 14

0.01

-

+

Gap

0.01

-

+

High-Low Spread

0.01

-

+

Day of Week (0=Mon)

0

▼

Month


1

▼

Is Weekend

1

▼

 Predicted next-day volatility (GK)

0.0136

- **Outputs:**
 - **Prediction:** Next-day GK volatility.
 - **Risk regime:** “High” vs “Normal” using 80th percentile threshold; warning shown if history is empty.
- **Reliability:**

- Caching for features/model; validation checks for missing data; graceful warnings.

Limitations

- **Proxy choice:** GK is a daily proxy; intraday realized vol could improve fidelity.
- **Data gaps:** Some assets have sparse or noisy histories; thresholds may be unstable.
- **Stationarity:** Volatility regimes shift—periodic retraining and recalibration recommended.

Future work

- **Feature expansion:** ATR, liquidity depth, realized vol, regime indicators.
- **Modeling:** Gradient boosting with monotonic constraints; probabilistic forecasts (quantile regression).
- **Evaluation:** Backtesting with rolling windows; calibration of risk thresholds per asset.
- **Deployment:** Batch scoring and API; dashboard with multi-asset comparison.