Data Preprocessing

- **Missing values:** forward/backward fill per crypto.
- **Outliers:** winsorization at 1–99% per symbol.
- **Normalization:** robust scaling for numeric features.
- **Categorical encoding:** one-hot for `crypto_name`, `dow`, `month`, `is_weekend`.

Feature Engineering

- **Returns:** log returns, % change.
- **Volatility:** rolling std (7d, 14d, 30d), GK, Parkinson.
- **Trend:** moving averages, slope of 14-day MA.
- **Liquidity:** volume/market cap, rolling volume stats.
- **Calendar:** day-of-week, month, weekend flag.
- **Target:** next-day GK volatility (`shift(-1)`).

Model

- **Algorithm:** XGBoost Regressor.
- **Hyperparameters:** n_estimators=600, max_depth=6, learning_rate=0.05, subsample=0.8, colsample_bytree=0.8, reg_lambda=1.0.
- **Evaluation:** RMSE, MAE, R^2 across train/val/test splits.

Deployment

- **Streamlit app:**
  - Auto mode → latest features per crypto.
  - Manual mode → user inputs features.
  - Output → predicted volatility + risk regime classification (80th percentile threshold).