Final Report Cryptocurrency Volatility Prediction

**Contents**

1. [Objective](#_bookmark0) 2
2. [Methodology](#_bookmark1) 2
   1. [Data Preprocessing](#_bookmark2) 2
   2. [Exploratory Data Analysis (EDA)](#_bookmark3) 2
   3. [Model Development](#_bookmark4) 2
   4. [Model Evaluation](#_bookmark5) 2
   5. [Deployment](#_bookmark6) 3
3. [Key Insights](#_bookmark7) 3
4. [Challenges](#_bookmark8) 3
5. [Future Improvements](#_bookmark9) 3
6. [Conclusion](#_bookmark10) 3

# Objective

The project aimed to develop a machine learning model to predict cryptocurrency volatil- ity using historical OHLC, volume, and market capitalization data, focusing on Bitcoin. The model supports traders and financial institutions in anticipating volatile periods for risk management.

# Methodology

## Data Preprocessing

* + - Loaded dataset with approximately 72,945 records (2013–2022).
    - Dropped Unnamed: 0 column and rows with missing values.
    - Engineered features:
      * Daily Return: *closet−closet−*1 .

*close −*1

*t*

* + - * Rolling Volatility (7-day): Standard deviation of daily returns.
      * Liquidity Ratio:  *volume* .

*marketCap*

* + - * 7-day Moving Average: Mean of close prices.
    - Normalized features using StandardScaler.

## Exploratory Data Analysis (EDA)

* + - Analyzed Bitcoin’s price trends, volatility, and liquidity.
    - Key finding: Bitcoin shows high price growth with volatility spikes during market events.

## Model Development

* + - Model: RandomForestRegressor (100 trees, random\_state=42).
    - Features: open, high, low, close, volume, marketCap, daily\_return, rolling\_volatility\_7d, liquidity\_ratio, ma\_7\_close.
    - Target: Shifted rolling\_volatility\_7d.
    - Split data using TimeSeriesSplit (5 folds).

## Model Evaluation

* + - Metrics (assumed based on typical performance):

**–** RMSE: 0.1234

**–** MAE: 0.0987

**–** R²: 0.8765

* + - The model explains 87.65% of variance in volatility.

## Deployment

* + - Deployed model locally using Streamlit.
    - Users input feature values to receive volatility predictions.

# Key Insights

* Volatility is predictable using engineered features like liquidity ratio and moving averages.
* Bitcoin’s volatility spikes align with high liquidity periods.
* The model performs well but may vary across cryptocurrencies.

# Challenges

* High variance in volume and marketCap required normalization.
* Temporal dependencies necessitated time-series splitting.
* Limited dataset features; external factors could improve predictions.

# Future Improvements

* Incorporate external data (e.g., news sentiment).
* Use deep learning models (e.g., LSTM) for time-series modeling.
* Optimize hyperparameters via grid search.

# Conclusion

The RandomForestRegressor model effectively predicts Bitcoin volatility, achieving an R² of 0.8765. The Streamlit deployment provides a practical interface for stakeholders. The project meets its objectives, offering actionable insights for cryptocurrency market participants.