Cryptocurrency Price Predictor - Project Documentation

Course: Introduction to Data Science

Author: Hasan Ali

Date: 14-06-2025

# 1. Objective

The objective of this project is to:  
- Select a unique cryptocurrency dataset.  
- Perform in-depth Exploratory Data Analysis (EDA).  
- Engineer and preprocess features for machine learning.  
- Train an XGBoost regression model to predict the next day's closing price.  
- Develop a Streamlit web application that allows real-time user interaction and predictions.

# 2. Dataset

Name: preprocessed\_crypto\_data1.csv  
Source: Custom-preprocessed historical crypto data.  
Features include:  
- Prices: Open, High, Low, Close, Volume, Marketcap  
- Technical Indicators: MA (7/30), RSI, Bollinger Bands  
- Lag Features: Close\_lag1, Volume\_lag3, etc.  
- Temporal Features: Month/Weekday sin/cos encodings  
Coins Covered: Bitcoin, Ethereum, Aave, etc.

# 3. Exploratory Data Analysis (EDA)

Performed using Python libraries (pandas, matplotlib, seaborn) within the Streamlit interface.  
Key Analyses:  
- Summary statistics (describe())  
- Missing value analysis  
- Volume distribution (histograms)  
- Close price trends (line charts)  
- Correlation heatmap  
- Feature importance from the trained model

# 4. Data Preprocessing

Each coin’s data is individually cleaned and transformed:  
- Missing values handled by rolling stats and shifting  
- Lag features added (1, 3, 7 days)  
- Technical indicators added (RSI, Bollinger Bands)  
- Time encodings using sin/cos functions  
- Features scaled using StandardScaler  
- Sorted by Date for time-series consistency

# 5. Machine Learning Model

Model Used: XGBoost Regressor (XGBRegressor)  
Input: Engineered features  
Output: Next-day Close Price  
Evaluation Metrics:  
- MAE (Mean Absolute Error)  
- RMSE (Root Mean Squared Error)  
- MAPE (Mean Absolute Percentage Error)  
- Directional Accuracy (% of times model predicts correct direction)

# 6. Streamlit Web Application

The project is deployed via an interactive Streamlit interface with:  
  
Sections:  
- Introduction – Project summary and dataset overview  
- EDA – Live charts and statistics based on coin selection  
- Model & Prediction – Train and predict closing prices based on user input  
- Conclusion – Key takeaways and future improvements  
  
Interactive Features:  
- Coin selection via dropdown  
- Real-time model training and next-day prediction  
- Charts: Line, Histogram, Heatmap, Feature Importance  
- User-friendly layout with Streamlit widgets and markdowns

# 7. File Structure

project/  
├── final\_code.py # Model and feature engineering functions  
├── streamlit\_app.py # Streamlit interface  
├── preprocessed\_crypto\_data1.csv # Dataset  
├── coin\_prediction\_metrics.csv # Evaluation metrics output  
├── coin\_predictions.csv # Predicted close prices

# 8. Runtime Prediction Requirement

✔️ Model makes predictions at runtime based on user-selected coin.  
✔️ Users receive the next day's closing price prediction with feature insight.

# 9. Conclusion

This project demonstrates a practical use of data science and machine learning in the financial domain. By combining feature engineering, modeling, and visualization, we built a complete pipeline to:  
- Analyze historical cryptocurrency data  
- Predict future prices  
- Deploy a user-friendly app for interaction

# 10. Future Enhancements

- Add live API-based crypto price updates  
- Implement LSTM/GRU models for sequential prediction  
- Allow multi-day forecasting  
- Enable CSV download of predictions and metrics