Crypto Market Data Analysis, Visualization, and Predictive Modeling Report

Executive Summary

This project presents an analysis and forecasting study of four major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), and Binance Coin (BNB). Using historical price data from multiple time frames, we gathered, cleaned, and explored the data to uncover price trends, volatility, and relationships between the cryptocurrencies. We developed and compared three forecasting models, ARIMA, Facebook Prophet, and LSTM neural networks, to predict future price movements. We evaluated model performance with RMSE, MAE, and MAPE metrics, revealing strengths and weaknesses across different cryptocurrencies and forecasting periods. To enable user interaction and visualization, we built an interactive dashboard with Streamlit, which allows users to explore historical data, forecasted prices, key market indicators, and model comparisons. The results provide useful insights into cryptocurrency price behavior and show the effectiveness of different forecasting methods. Future research could focus on improving feature engineering, expanding model architectures, and incorporating real-time data streaming.

1. Introduction

Cryptocurrencies have become important financial assets known for their high volatility and changing market dynamics. Reliable price forecasting is essential for investors, traders, and analysts who want to manage risks and seize opportunities. This project focuses on four commonly traded cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), and Binance Coin (BNB). We selected these due to their large market capitalization, liquidity, and representation of various blockchain ecosystems.

The objectives are to:

• Collect and preprocess historical data at different time granularities,

• Perform exploratory data analysis (EDA) to understand statistical properties and relationships,

• Develop forecasting models (ARIMA, Prophet, LSTM) for price predictions,

• Evaluate model accuracy and robustness using standard error metrics, and

• Create an interactive dashboard for data visualization and model comparison.

2. Data Collection and Preparation

We sourced data from reliable financial APIs, including Yahoo Finance and CoinGecko, covering three distinct time ranges:

• Short range: 1 month of 2-minute interval data,

• Medium range: 2 years of 1-hour interval data,

• Long range: Maximum available data at daily intervals.

The initial datasets had missing values and occasional outliers. We handled missing data through forward filling and interpolation, while we addressed outliers using Winsorization and manual inspection. We cleaned and transformed the data to ensure consistency and readiness for analysis and modeling.

3. Exploratory Data Analysis (EDA)

Univariate analysis included calculating summary statistics like mean, median, standard deviation, and skewness for each cryptocurrency across different periods. Visualization with line plots and box plots showed historical price trends, volatility spikes, and distribution characteristics.

Bivariate analysis examined correlations between cryptocurrencies. This revealed moderate to strong relationships, particularly among BTC, ETH, and BNB. We also looked into price-volume relationships, which showed varying degrees of correlation depending on the cryptocurrency and time frame.

4. Predictive Modeling

Feature Engineering

We engineered key features to improve forecasting accuracy. These included moving averages (e.g., 7-day, 30-day), price momentum indicators, and volatility measures.

Model Development

• ARIMA: Used for its ability to capture linear time dependencies; we optimized parameters using AIC criteria.

• Prophet: Implemented for its capability to handle seasonality and trend changes; we tuned hyperparameters for each cryptocurrency.

• LSTM: Deployed as a deep learning approach that can model nonlinear dependencies; we adjusted the network architecture and training epochs.

Evaluation

We evaluated models on test sets using RMSE, MAE, and MAPE metrics. Generally, LSTM performed better in capturing complex patterns, especially during volatile periods, while ARIMA and Prophet were effective for longer-term trends.

5. Dashboard Development

We created a user-friendly interactive dashboard with Streamlit that includes:

• Historical and forecasted price charts with zoom and tooltip features,

• Key metrics panels showing daily returns, volatility, and trading volume,

• Visual comparisons of model accuracy metrics,

• Interactive controls that let users choose cryptocurrencies, adjust forecasting periods, and switch between models.

This dashboard provides an intuitive way to explore cryptocurrency data and predictive insights.

6. Results and Discussion

Our analysis confirmed high volatility and unique behavioral patterns among the selected cryptocurrencies. The predictive models varied in effectiveness based on the crypto asset and forecast length. LSTM consistently performed well for short-term forecasts, while Prophet and ARIMA were competitive for medium- and long-term projections. The dashboard effectively communicated complex results in a straightforward manner.

7. Conclusion and Recommendations

This project successfully showed the full process of cryptocurrency market data analysis, forecasting, and visualization. The combination of traditional statistical models and modern deep learning methods allowed for a deeper understanding of price dynamics. Future work could focus on:

• Adding features like sentiment analysis or macroeconomic indicators,

• Expanding to more cryptocurrencies,

• Implementing real-time forecasting pipelines, and

• Improving dashboard interactivity with advanced filtering and alerts.

References

• Yahoo Finance API

• CoinGecko API

• Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice.

• Taylor, S.J., & Letham, B. (2018). Forecasting at scale (Facebook Prophet).

• Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory.