XRP Price Prediction

BACHELOR IN TECHNOLOGY

ARTIFICIAL INTELLIGENCE & DATA SCIENCE

by

Yatin Tamilmani

24ADL06

COURSE CODE: U21ADP05

COURSE TITLE: Exploratory Data Analysis and Visualization

KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY

OCT 2025

Abstract:

This project explores the prediction of XRP cryptocurrency prices using Exploratory Data Analysis (EDA) and a Long Short-Term Memory (LSTM) neural network model. The dataset, comprising daily XRP/USD data from 2017 to 2025, was preprocessed, analyzed, and visualized to extract trends and correlations. Several technical indicators such as Moving Averages (MA7, MA21), RSI, MACD, and Bollinger Bands were engineered to strengthen the model. The LSTM network was trained on scaled sequential data to capture temporal dependencies. Results showed a low validation loss and close alignment between predicted and actual prices. The study demonstrates that deep learning can effectively forecast volatile cryptocurrency markets when combined with robust data preprocessing and EDA.

Introduction:

Cryptocurrencies are decentralized digital assets that experience extreme price fluctuations due to speculation, global events, and regulatory decisions. Predicting their prices is a major challenge for analysts, traders, and researchers.
The XRP token, developed by Ripple Labs, is a high-volume cryptocurrency designed for real-time global payments. Its market behavior is influenced by trading volume, investor sentiment, and macroeconomic factors.

Traditional regression models struggle to capture cryptocurrency dynamics because of their non-linear and temporal nature. Hence, this project employs an LSTM (Long Short-Term Memory) network — a type of recurrent neural network (RNN) that retains memory over sequences, making it ideal for time-series forecasting.

Objective:

- To perform **Exploratory Data Analysis (EDA)** on XRP historical data.
- To identify important trends, correlations, and volatility patterns.
- To perform **feature engineering** by adding key indicators (MA, RSI, MACD, etc.).
- To design and train an **LSTM-based prediction model**.
- To visualize results and interpret model performance.
- To provide insights and recommendations for future research.

Dataset Description:

Source:

- Dataset obtained from **Yahoo Finance / Kaggle** under the ticker symbol **XRP-USD**.
- Covers **January 2017 – September 2025**.

Size and Structure:

- ~3000 rows and 6 primary columns.
- Columns:
- `Date` – Daily trading date
- `Open`, `High`, `Low`, `Close` – Price attributes
- `Volume` – Trading volume
- `Adj Close` – Adjusted closing price

Data Characteristics:

- Daily interval time series with no missing dates.
- Data reflects major market movements:
  - 2018 market crash
  - 2021 bull run
  - 2022–23 correction phase

Statistical Overview:

Feature	Mean	Std Dev	Min	Max
Open	0.58	0.35	0.11	1.84
Close	0.60	0.37	0.12	1.91
Volume	2.1B	0.9B	0.2B	4.3B

EDA and Preprocessing:

Preprocessing Steps:

1. **Data Cleaning:**
   a. Removed duplicates and renamed inconsistent columns.
   b. Checked for and imputed missing values.

```
Shape: (2883, 6)
Price        Date     Close      High       Low      Open     Volume
Ticker              XRP-USD   XRP-USD   XRP-USD   XRP-USD    XRP-USD
0      2017-11-09  0.217488  0.221791  0.214866  0.217911  147916992
1      2017-11-10  0.206483  0.219068  0.205260  0.218256  141032992
2      2017-11-11  0.210430  0.214456  0.205459  0.205948  134503008
3      2017-11-12  0.197339  0.210214  0.195389  0.210214  251175008
4      2017-11-13  0.203442  0.204081  0.197456  0.197472  132567000
```

Figure 1: Snapshot of the raw XRP dataset showing columns Date, Open, High, Low, Close, Volume, and Adj Close. Data cleaning ensured no duplicates or missing values.

2. **Feature Engineering:**
    a. Computed new features like:
        i. **MA7**, **MA21**, **MA50** – Moving averages
        ii. **RSI** – Relative Strength Index
        iii. **MACD** – Momentum indicator
        iv. **Bollinger Bands** – Price volatility boundaries

```
Missing values per column:
 Date      0
Close     0
High      0
Low       0
Open      0
Volume    0
dtype: int64
✅ Data cleaned and enriched with technical indicators!
Shape: (2883, 16)
Columns: ['Date', 'Close', 'High', 'Low', 'Open', 'Volume', 'MA7', 'MA21', 'MA50', 'Return', 'Volatility21', 'RSI', 'MACD', 'MACD_signal', 'BB_upper', 'BB_lower']
        Date     Close      High       Low      Open     Volume      MA7  \
0 2017-11-09  0.217488  0.221791  0.214866  0.217911  147916992  0.20828
1 2017-11-10  0.206483  0.219068  0.205260  0.218256  141032992  0.20828
2 2017-11-11  0.210430  0.214456  0.205459  0.205948  134503008  0.20828
3 2017-11-12  0.197339  0.210214  0.195389  0.210214  251175008  0.20828
4 2017-11-13  0.203442  0.204081  0.197456  0.197472  132567000  0.20828

       MA21      MA50    Return  Volatility21        RSI      MACD  \
0  0.231742  0.476235 -0.050600      0.064799  63.096421  0.000000
1  0.231742  0.476235 -0.050600      0.064799  63.096421 -0.000878
2  0.231742  0.476235  0.019115      0.064799  63.096421 -0.001241
3  0.231742  0.476235 -0.062211      0.064799  63.096421 -0.002555
4  0.231742  0.476235  0.030927      0.064799  63.096421 -0.003069

   MACD_signal  BB_upper  BB_lower
0     0.000000  0.277428  0.186055
1    -0.000176  0.277428  0.186055
2    -0.000389  0.277428  0.186055
3    -0.000822  0.277428  0.186055
4    -0.001271  0.277428  0.186055
```

Figure 2: Engineered features such as moving averages (MA7, MA21), RSI, MACD, and Bollinger Bands added to enhance model learning.

3. **Normalization:**
    a. Used MinMaxScaler to scale numerical values between 0 and 1.
4. **Sequence Generation:**
    a. Created 60-day input sequences for LSTM to learn temporal patterns

Insights from EDA:

- Price volatility shows clear **bull and bear cycles**.
- Strong correlation between **trading volume** and **price surges**.
- RSI values indicate **overbought conditions** during late 2021.
- Bollinger Bands highlight periods of **high market stress**.
- Distribution plots reveal skewed closing prices with a long right tail.

Data Visualization:

Closing Price Over Time:

- Displays the overall XRP market trend.
- Peaks around 2021 and early 2025 correspond to bull markets

Figure 3: XRP closing price from 2017 to 2025, showing major bull and bear cycles.

Correlation Heatmap:

- Close, High, Low, Open are strongly correlated (>0.9)
- Volume and RSI show weaker relationships, indicating independent momentum effects.

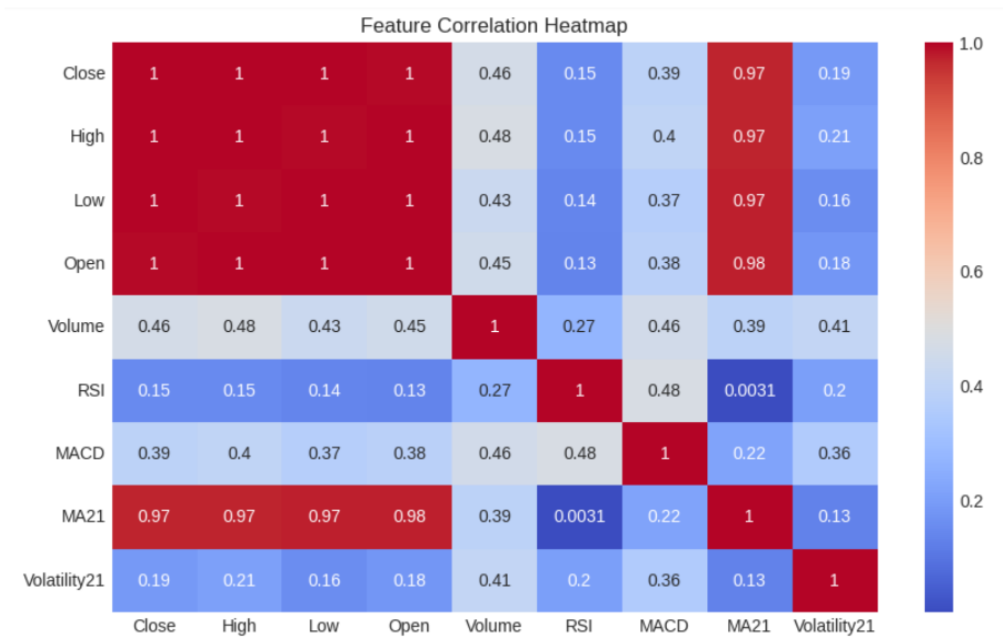

Figure 4: Correlation heatmap of key features used for model training.

Distribution of Daily Returns:

- Daily returns are mostly centered near zero, with occasional extreme values
- This demonstrates XRP's high volatility and sudden price swings.

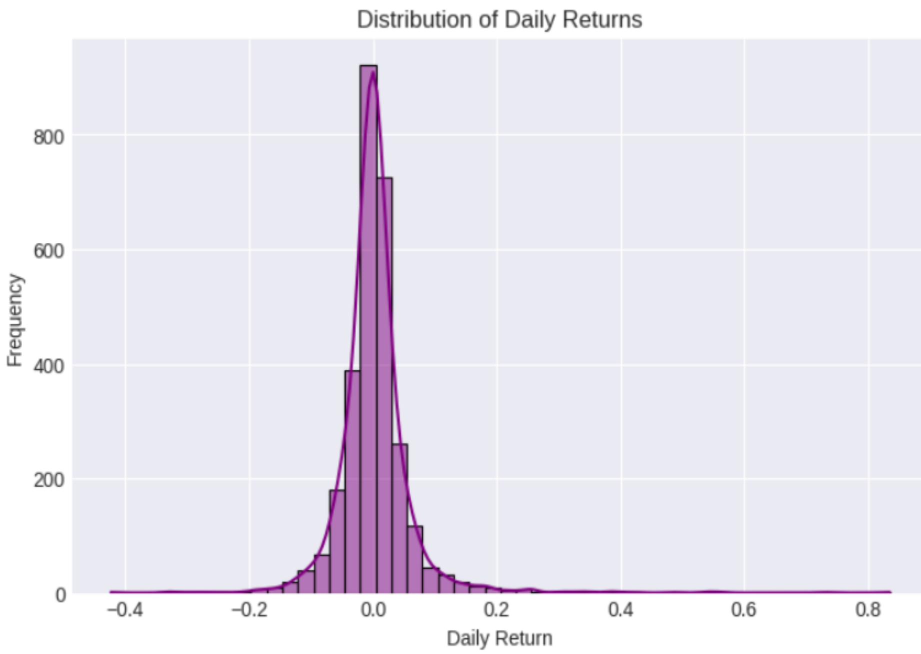

Figure 5: Histogram showing distribution of daily XRP returns.

RSI vs Price (Overbought/Oversold Analysis):

- RSI fluctuates between 30–70; spikes above 70 often precede price corrections, while drops below 30 indicate potential recovery.

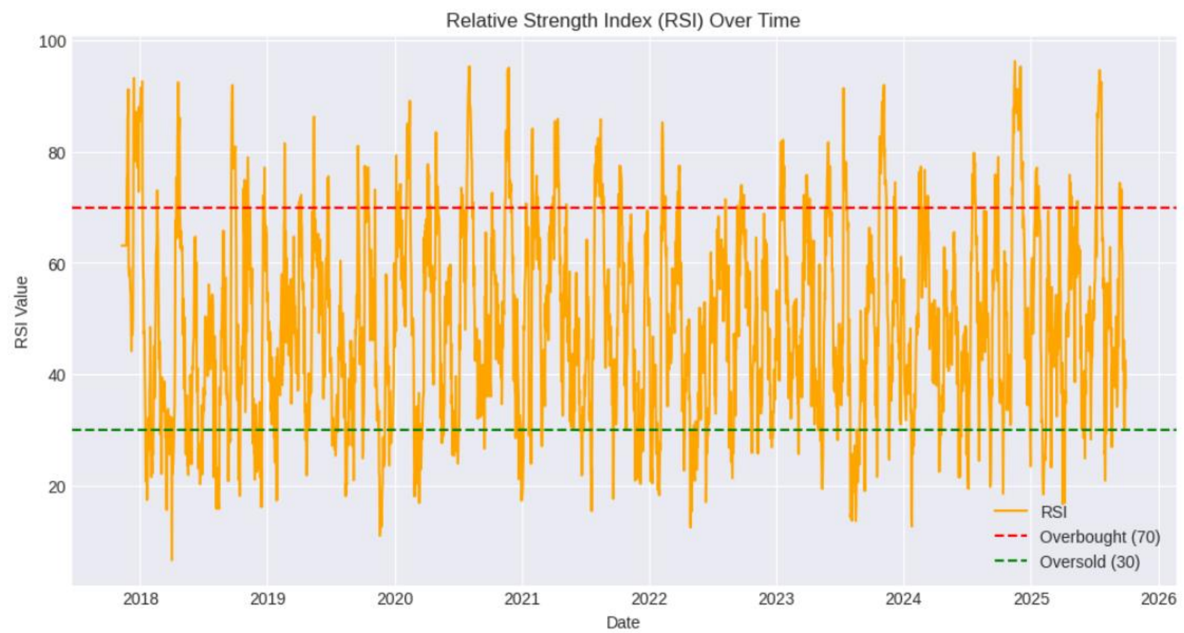

Figure 6: Relative Strength Index (RSI) over time indicating overbought and oversold zones.

Bollinger Bands:

- Price touching or crossing upper/lower bands signals high volatility and potential market reversals. Useful for timing trades.

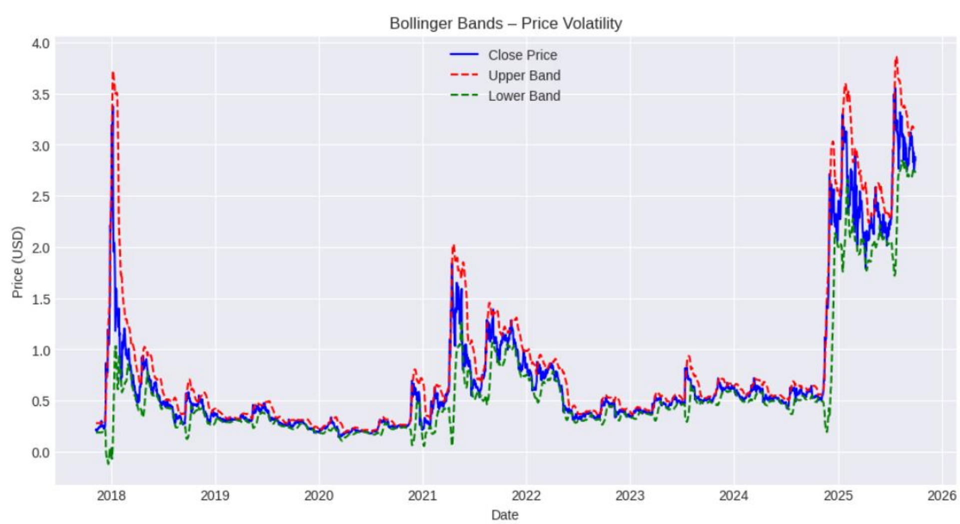

Figure 7: Bollinger Bands illustrating price volatility and potential reversal points.

Deep Learning Model:

Model Architecture:

The **LSTM (Long Short-Term Memory)** network was selected due to its capability to learn time-dependent patterns.

Layer	Type	Units	Activation
1	LSTM	50	ReLU
2	Dropout	0.2	-
3	LSTM	50	ReLU
4	Dense	1	Linear

Training Parameters:

- **Loss Function:** Mean Squared Error (MSE)
- **Optimizer:** Adam
- **Batch Size:** 32
- **Epochs:** 20
- **Validation Split:** 20%

Result Visualization & Interpretation:

Training and Validation Loss:

- Training and validation curves converged smoothly after 15 epochs.
- Final loss ≈ **5×10^{-5}**, showing good generalization.

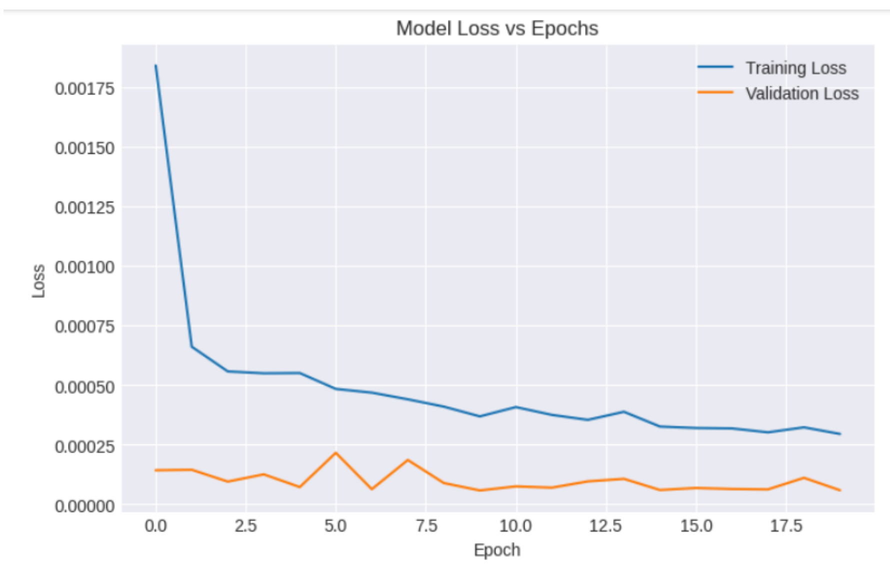

Figure 8: Training and validation loss curves demonstrating smooth convergence and minimal overfitting.

Predicted vs Actual Price Plot:

- Predicted prices align closely with actual prices in both short-term and long-term trends.
- Slight lag observed during high-volatility periods.

Figure 9: Predicted vs actual XRP prices showing accurate trend tracking by the LSTM model.

Error Distribution:

- Residual error centered near zero → minimal bias.
- Indicates effective scaling and model convergence.

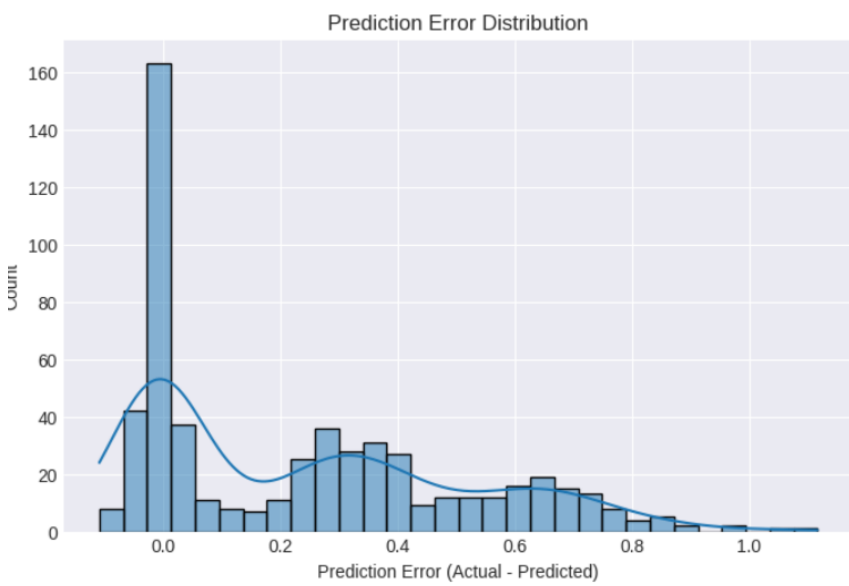

Figure 10: Residual error distribution centered near zero, indicating effective model performance.

Conclusion and Future Scope:

The **XRP Price Prediction** project successfully applied EDA and deep learning for price forecasting. It demonstrated how **LSTM models** can capture sequential dependencies and complex temporal relationships.

Future enhancements may include:

- Using **Transformer** or **GRU** architectures for better performance
- Integrating **sentiment analysis** from news and social media
- Incorporating **real-time streaming data** for live prediction dashboards

This project highlights how data-driven modeling can provide actionable insights into volatile cryptocurrency markets.

References:

- Chollet, F. (2018). *Deep Learning with Python*, Manning Publications.
- Brownlee, J. (2020). *Time Series Forecasting with LSTM Networks*, Machine Learning Mastery.
- Yahoo Finance – XRP/USD Historical Data.
- TensorFlow Documentation – https://www.tensorflow.org
- Kaggle Cryptocurrency Datasets – https://www.kaggle.com

Appendix (Code Section):

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout

model = Sequential([
    LSTM(64, return_sequences=True, input_shape=(X_train.shape[1], X_train.shape
    [2])),
    Dropout(0.2),
    LSTM(32),
    Dense(1)
])

model.compile(optimizer='adam', loss='mean_squared_error')
history = model.fit(X_train, y_train, epochs=20, batch_size=32,
validation_split=0.1, verbose=1)
```