Predicting Solana Cryptocurrency Prices Using LSTM-Based Linear Regression

A Financial Analysis and Predictive Modeling Approach

Objective and Financial Metrics

- Objective:
- - Predict Solana (SOL) closing prices at 1, 5, and 7 days ahead.
- - Use Bitcoin prices to compute the Beta coefficient for Solana.

- Key Financial Metrics:
- 1. **Beta Coefficient**: $\beta = Cov(SOL, BTC) / Var(BTC)$
- 2. **Bollinger Bands**: Upper Band = SMA + 2σ, Lower Band = SMA - 2σ
- 3. **Sharpe Ratio**: $(E[R] - Rf) / \sigma R$

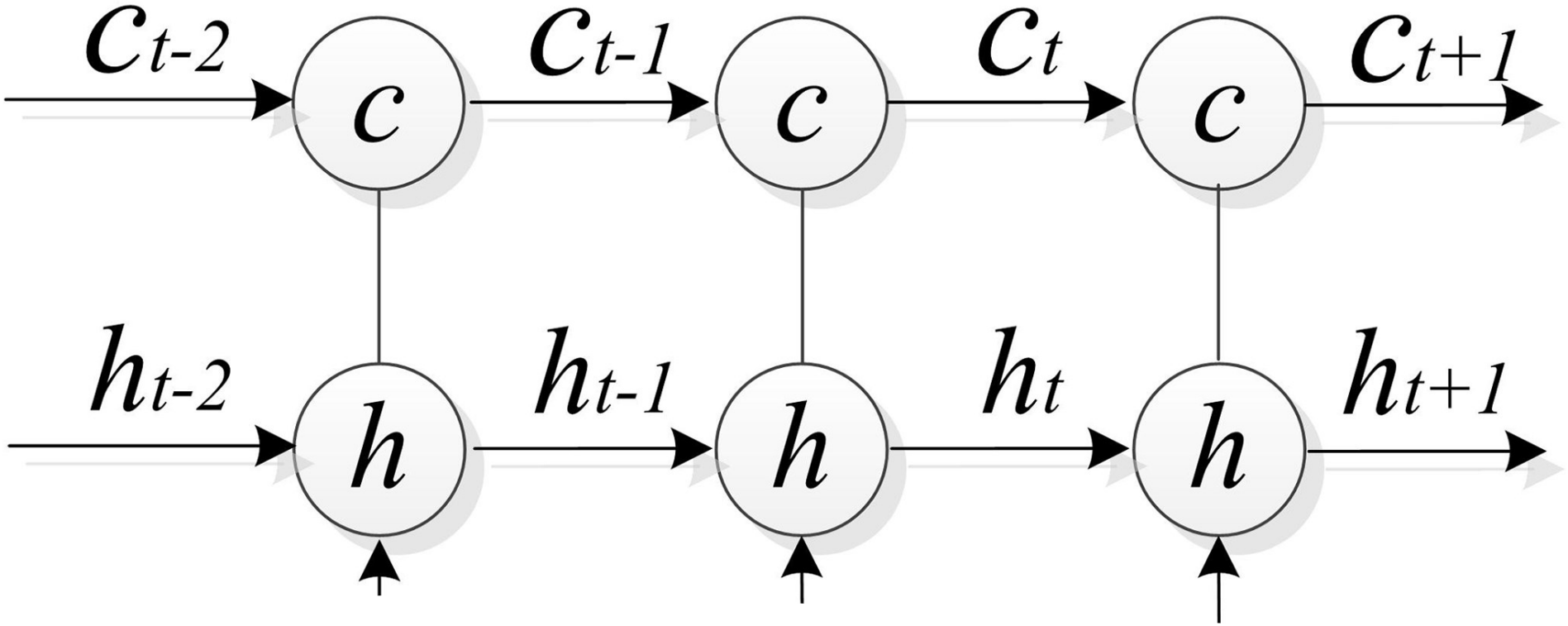
Exploratory Data Analysis (EDA)

- Step 1: Correlation Analysis
- Computed correlation matrix to check multicollinearity potential issues.
- Step 2: Distribution Analysis
- Assessed covariate distributions.
- Scatter Plot Analysis to check the relationship among the covariates.

Data Preparation and Benchmark Model

- Dataset:
- Source: Yahoo Finance (SOL-USD and BTC-USD).
- Split into training, validation, and test sets (historical order maintained).

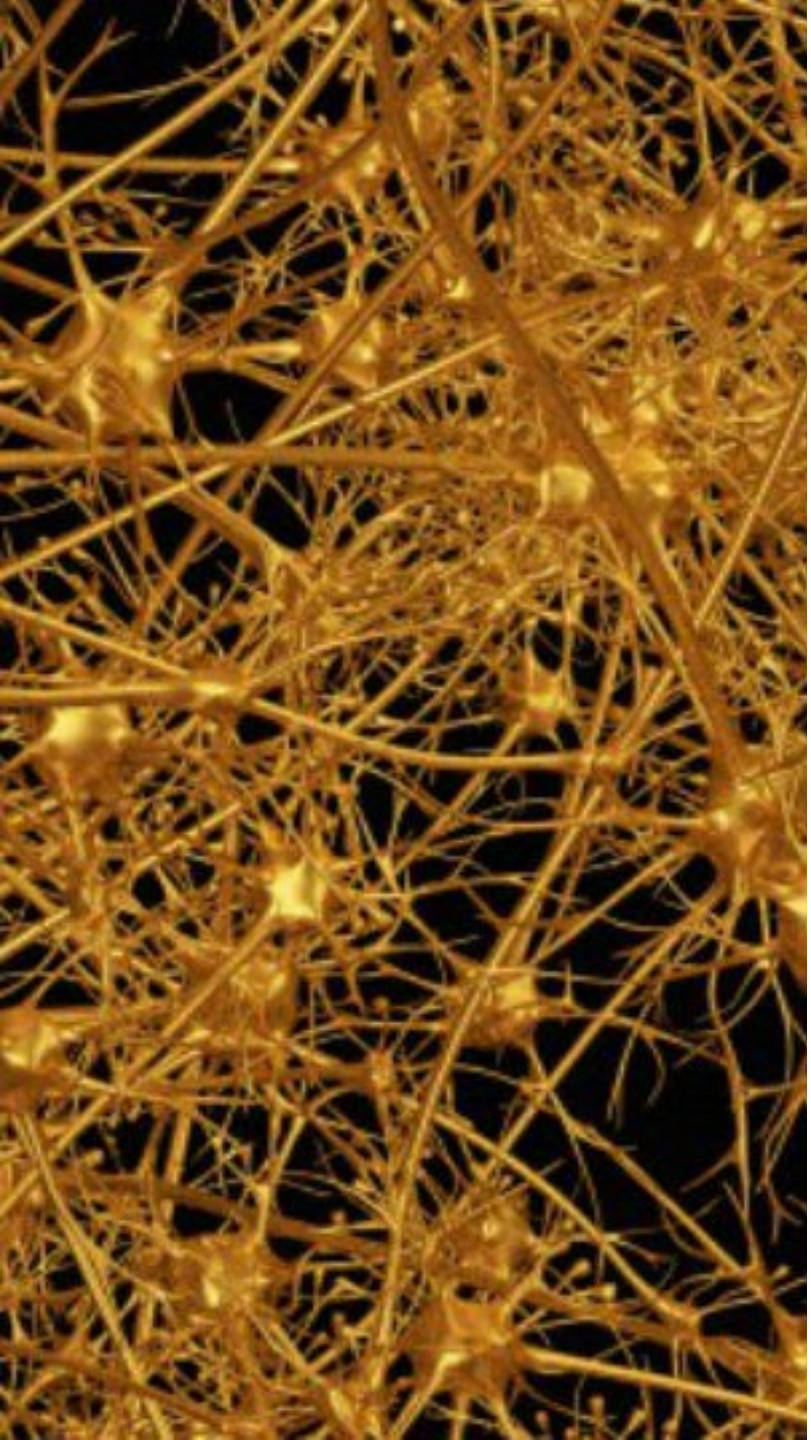
- Benchmark Model: Simple Moving Average (SMA)
- The SMA model averages prices over a fixed rolling window to predict future prices. For example, to forecast the price for the next day, it uses the mean of the most recent values within a window (e.g., 5 days).



Long Short-Term Memory (LSTM) Model

- Why LSTM for Time Series?
- Captures long-term dependencies in sequential data.
- Handles vanishing gradient problem.
- Ideal for financial forecasting tasks.

- Structure:
- Input: Time series of prices/indicators.
- Layers: LSTM layers with dropout for regularization.
- Output: Predictions for 1, 5, and 7 days ahead.



LSTM structure

- **First LSTM Layer**: Captures temporal patterns from input features (input_size).
- **Dropout Layer**: Adds regularization to prevent overfitting.
- **Second LSTM Layer**: Builds on the first layer, learning more complex dependencies.
- **Fully Connected Layer**: Maps hidden states to the final output (output_size).

Training the LSTM Model

- Trained on datasets for 1, 5, and 7 days ahead predictions.
- Mean Squared Error (MSE).
- Mean Absolute Error (MAE)
- Optimizer: Adam
- Trained for 50 epochs with early stopping based on validation.

Evaluation on test Set

Performance Metrics	
1-Day Ahead Predictions:	**Mean Absolute Error (MAE):** 12.3374 **Mean Squared Error (MSE):** 212.5564 **R^2 Score:** 0.7993
5-Day Ahead Predictions:	**Mean Absolute Error (MAE):** 17.9274 **Mean Squared Error (MSE):** 509.0615 **R^2 Score:** 0.5657
7-Day Ahead Predictions:	**Mean Absolute Error (MAE):** 20.3577 **Mean Squared Error (MSE):** 634.7626 **R^2 Score:** 0.4829

Conclusions

- **Effective Predictions**: LSTM models achieve low error and high R^2, excelling in short-term (1-day) forecasts.
- **Generalization**: Consistent performance on unseen test data.
- **Outperformance**: LSTM significantly surpasses the SMA benchmark.
- **Practical Use**: Residual analysis and trend alignment validate reliability.