

CSCI 6444 Introduction to Big Data and Analytics

Computer Project Milestone 1 (10 pts)

Bitcoin Price Prediction

Team Members:

Srinadh Doppalapudi - G30742950

Nandini Gogineni - G48744249

Problem Definition:

Bitcoin is the most popular cryptocurrency, noted for its volatile price fluctuations. Its value is determined by a variety of factors, including trade activity, global economic conditions, regulations, and market sentiment. These rapid price changes make Bitcoin forecasting challenging yet necessary for investors and researchers.

The question we want to answer is if we can use machine learning and deep learning models to estimate Bitcoin's daily closing price based on past market data.

Background Review:

Bitcoin was first introduced in 2009 and has since become an important financial asset. Its price has fluctuated dramatically throughout the years, going from a few hundred dollars to tens of thousands and then decreasing again in short times. This kind of volatility makes it difficult to employ typical forecasting tools.

Statistical models such as ARIMA have been employed in financial forecasting, however they are unable to capture nonlinear and abrupt changes. Machine learning approaches like regression and random forests perform better, but they cannot handle long-term dependencies in time series data.

Deep learning methods, particularly Long Short-Term Memory (LSTM) networks, have shown useful for time-series forecasting. LSTMs can learn patterns over time and handle sequential dependencies more effectively than classical models. This makes them an excellent choice for Bitcoin price prediction.

Proposed Solution:

Our research will develop and compare models for forecasting Bitcoin's daily closing price.

Data collection: Use historical daily Bitcoin data, such as open, high, low, close, and volume.

Data Preparation: Clean and normalize the dataset, then add characteristics like as moving averages, volatility measurements, and lag values.

Models: Begin with linear regression and ARIMA to get baseline results.

Deep Learning Model: Create an LSTM network to forecast the next day's closing price based on previous data sequences.

Evaluation: To evaluate performance, use RMSE, MAE, and R^2 scores, as well as assess directional accuracy (trends upward or downward).

Visualization: Plot actual vs. expected values to improve understanding. We expect the LSTM model to outperform the baselines since it can identify time-based patterns in the data.

Data Resources:

Primary Dataset:

- Historical Bitcoin daily data from [Investing.com](https://www.investing.com).
- Coverage: January 1, 2014 – Present.
- Features: Open, High, Low, Close, Volume, % Change.