THOMPSON RIVERS UNIVERSITY

Time Series Analysis using Cryptocurrency Price Data

(Bitcoin Price)

Final Project Report

Masters of Science in Data Science

DASC 5420

(Winter 2024)

Group C:

Mengyang Zhang - T00696467

Waqar Ul Mulk - T00729986

Zhilin Song - T00731619

KAMLOOPS, BRITISH COLUMBIA

[November, 2024]

Professor: Dr. Erfan Hoque

Department of Statistics and Mathematics

Abstract

The rapid growth and volatility of cryptocurrencies have captured the attention of researchers and investors alike. Cryptocurrency prices cannot be determined with the same degree of certainty that the stock market can be. Therefore, this project focuses on using some advanced time series analysis techniques to forecast Bitcoin prices. The Forecasting is done using ARIMA, Dynamic Regression, and Neural Networks and the results demonstrated that Dynamic Regression and Neural Network outperforms most as compared to ARIMA in predicting Bitcoin prices on a daily time series basis in terms of root mean squared error RMSE and mean absolute percentage error MAPE. These models assist investors in making informed decisions regarding the buying or selling of cryptocurrency stocks.

Keywords: *Forecast, Accuracy, Volatility, Regression, Neurons.*

1. Introduction

Cryptocurrencies are a truly exceptional form of an asset class. Originally developed as a decentralized digital currency encrypted to the blockchain, cryptocurrencies are nowadays traded for the motive of speculation (Baur et al., 2018) [1]. Bitcoin has inspired extensive research into its price prediction, given its unique market dynamics compared to traditional stocks. There are a number of algorithms used on stock market data for price prediction. However, the parameters affecting Bitcoin are different. Therefore it is necessary to predict the value of Bitcoin so that correct investment decisions can be made. The price of Bitcoin does not depend on the business events or intervening government unlike the stock market. Thus, to predict the value we feel it is necessary to leverage advanced techniques to predict the price of Bitcoin [2].

In this study we analyzed Bitcoin stock data for the last ten years using noble Autoregressive Integrated Moving Average (ARIMA) model, to see how ARIMA performs in short and long-term time series dynamics, a Dynamic Regression model to extend the traditional regression models by including external predictors (exogenous variables) along with time series components, and a non-parametric machine learning approach (NNAR) which is specifically designed for univariate time series data. It basically extends the idea of traditional ARIMA models by using neural networks to capture complex, nonlinear relationships in the data.

2. Data

2.1 Dataset Description

The dataset contains historical daily price data for Bitcoin (BTC) from 2014-01-01 to 2024-11-07, sourced from Investing.com. It includes the following attributes for each day: Date, Price, Open, High, Low, Vol., and Change %. This data serves as a foundation for exploring various forecasting methods in time series analysis, with a focus on understanding market dynamics and predicting cryptocurrency prices.[3]

2.2 Data Preprocessing
Feature Selection: Relevant features, including 'Open','High','Low' prices and 'Vol.', were selected to predict stock prices.

Train-Test Split: The data from 2019-01-01 to 2023-12-31 was selected as the training set, and the data from 2024-01-01 to 2024-11-07 was used for testing.

2.3 Data Overview
While the data exhibits significant short-term volatility, it lacks a consistent long-term trend or clear seasonal patterns, as evident from the price fluctuations shown in [Figure 1]. This highlights the highly unpredictable and dynamic nature of the cryptocurrency market.

3. Methodology

3.1 ARIMA
The AutoRegressive Integrated Moving Average ARIMA(p, d, q) model is represented as:

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\epsilon_t$$

Where:
B: Backward shift operator ($By_t = y_{t-1}$)
$\phi(B)$:p-th order polynomial in B, representing the autoregressive (AR) component
$\theta(B)$:q-th order polynomial in B, representing the moving average (MA) component
d: Degree of differencing to achieve stationarity
c: Constant term
ϵ_t: Error term

3.2 Dynamic Regression
The dynamic regression model with ARIMA (p, d, q) errors is given by:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \ldots + \beta_k x_{k,t} + \eta_t$$
$$\phi(B)(1 - B)^d \eta_t = \theta(B)\epsilon_t$$
$$\epsilon_t \sim IID(0, \sigma^2)$$

Where:
y_t: The dependent (response) variable at time t
β_0: The intercept term, representing the baseline value of y_t when all predictors are zero
$x_{1,t}, \ldots, x_{k,t}$:Predictor (independent) variables at time t
$\beta_1, \ldots, \beta_k$: Coefficients representing the influence of each predictor variable on y_t
$\eta_t \sim ARIMA(p, d, q)$

3.3 Neural Network
The model utilized a fully connected neural network using the 'scikit-learn' library. The parameters of the network were optimized using the Adam algorithm, which dynamically adjusts the learning rate based on estimates of the first and second moments of the gradients. Specifically, the Adam algorithm computes the exponential moving averages of the gradients and their squares and uses these to update the parameters. This allows Adam to perform well in scenarios with sparse or noisy gradients.

The model consists of two fully connected layers (dense layers). The input layer has 52 features, corresponding to our predictor variables such as "Open" and "High" prices. The hidden layer contains 10 neurons and applies the ReLU activation function. The output layer consists of a single neuron that outputs the prediction result.

3.4 GitHub Link
https://github.com/WaqarMohmand/Time-Series-Group-Project

4. Results

The ACF and PACF of the training data is checked and is shown in [Figure 2].

It can be seen that ACF decreases slowly and the values of PACF decreases rapidly after lag 2. Therefore, ARIMA(2,0,0) is chosen to be the ARIMA model used for forecasting.
The comparison of the predicted value and the actual value using the ARIMA(2,0,0) model is shown in [Figure 3]. It can be seen that the gap between the predicted value and the true value is large. Therefore, dynamic regression model and neural network is then used for forecast.

The dynamic regression model we use takes the previous day's opening price, highest price, lowest price, and trading volume as predictor variables, and with the error ARIMA(1,0,0). The comparison of the predicted value and the actual value of the dynamic model is shown in [Figure 4]. As can be seen in [Figure 4], the predicted value of the dynamic regression model is close to the true value, and the prediction performance is better than that of the ARIMA (2,0,0) model.

The comparison between the forecasts of the Neural Network Autoregressive model (NNAR) and the true values is shown in [Figure 5].

It can be seen that the predicted value of NNAR is also similar to the true value, and the predicted value is close to that of the dynamic regression model.We also used RMSE and MAPE to compare the performance of the three models. The results are shown in Table 1.

Table 1: The RMSE and MAPE of the models

Model	RMSE	MAPE
ARIMA(2,0,0)	24755.32	35.62
Dynamic Regression	1813.26	2.20
NNAR	1929j.56	2.37

As can be seen from Table 1, the RMSE and MAPE of ARIMA (2, 0, 0) are not ideal, while that of the dynamic regression model and NNAR are similar, with the dynamic regression model slightly better.

5. Conclusion

This project compared the performance of three models—ARIMA, Dynamic Regression, and Neural Network (NNAR)—in predicting Bitcoin prices. Among them, the Dynamic Regression Model demonstrated the best performance, achieving the lowest RMSE and MAPE, effectively capturing patterns in the data and delivering the most accurate predictions.

On the other hand, the ARIMA Model performed the worst, with the highest RMSE and MAPE, struggling to handle the complexity and volatility of the dataset, particularly in the absence of external predictors. The NNAR Model achieved moderate success, surpassing ARIMA but falling short of the accuracy achieved by the Dynamic Regression Model.

In conclusion, the results emphasize the importance of integrating external factors and capturing the correlation and volatility of time series data, as seen in the Dynamic Regression Model, for improving predictive accuracy. Future work could explore hybrid approaches that combine the strengths of these models to achieve even better performance.

References

1. Oliver Borgards,Dynamic time series momentum of cryptocurrencies,The North American Journal of Economics and Finance,Volume 57,2021,101428,ISSN 1062-9408, https://doi.org/10.1016/j.najef.2021.101428.

2. S. Velankar, S. Valecha and S. Maji, "Bitcoin price prediction using machine learning," 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon, Korea (South), 2018, pp. 144-147, doi:10.23919/ICACT.2018.8323676

3. Investing.com. Historical Data for Bitcoin USD [Internet]. [cited 2024 Nov 29]. Available from: https://ca.investing.com/crypto/bitcoin/historical-data

A Appendix

Figure 1: Bitcoin Price VS Time Overview

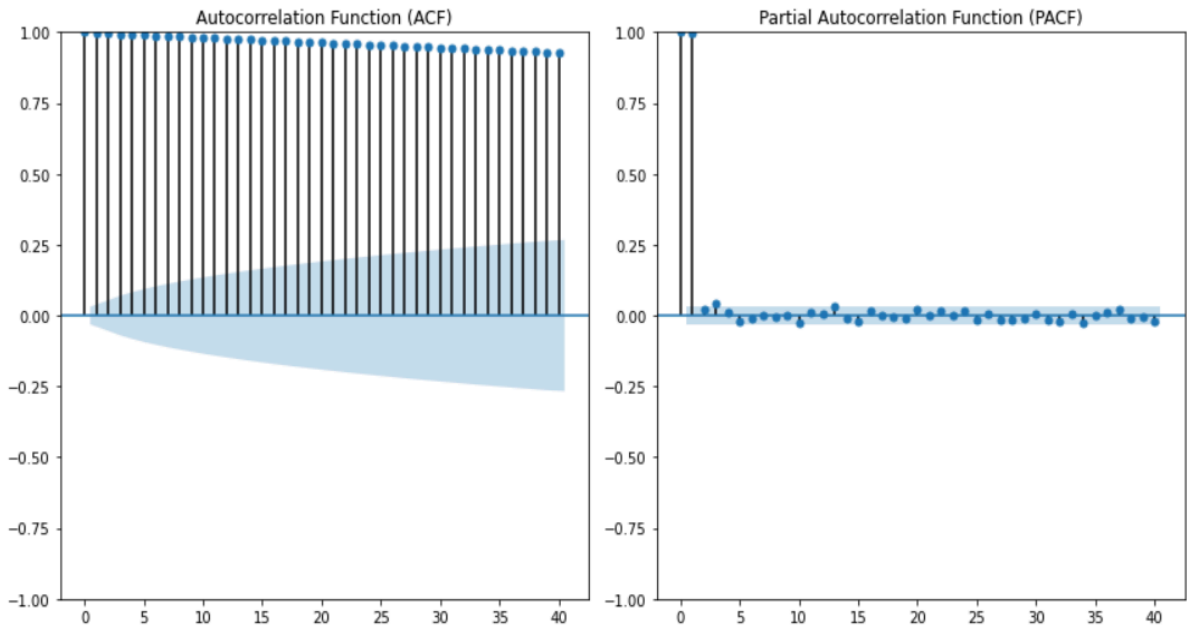

Figure 2: The figures of ACF and PACF

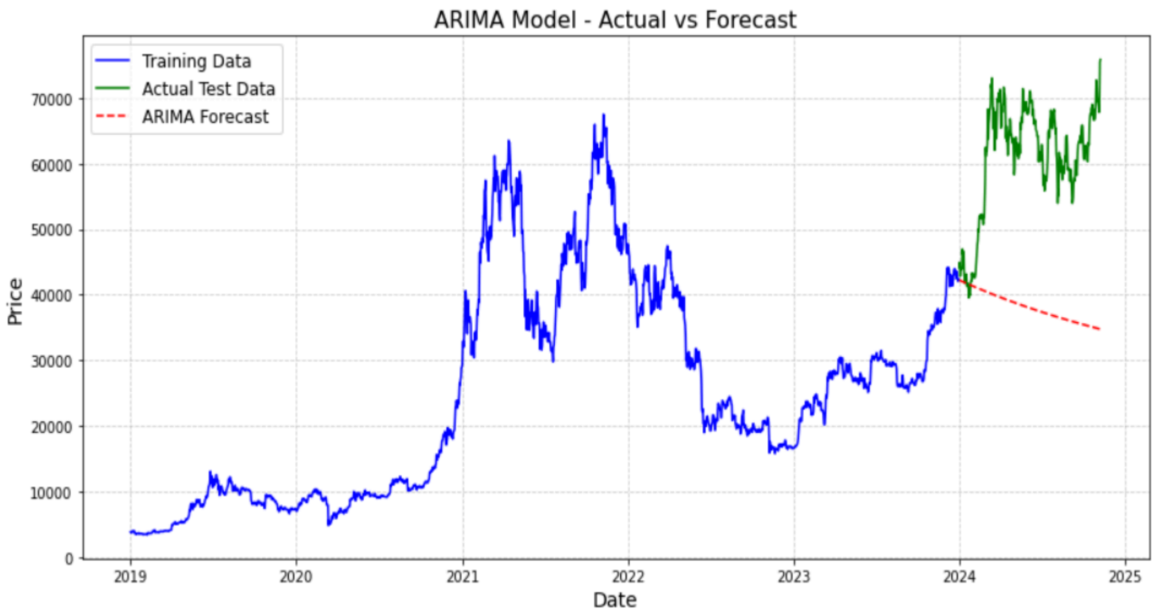

Figure 3: The comparison of the true values and predicted values of ARIMA(2,0,0)

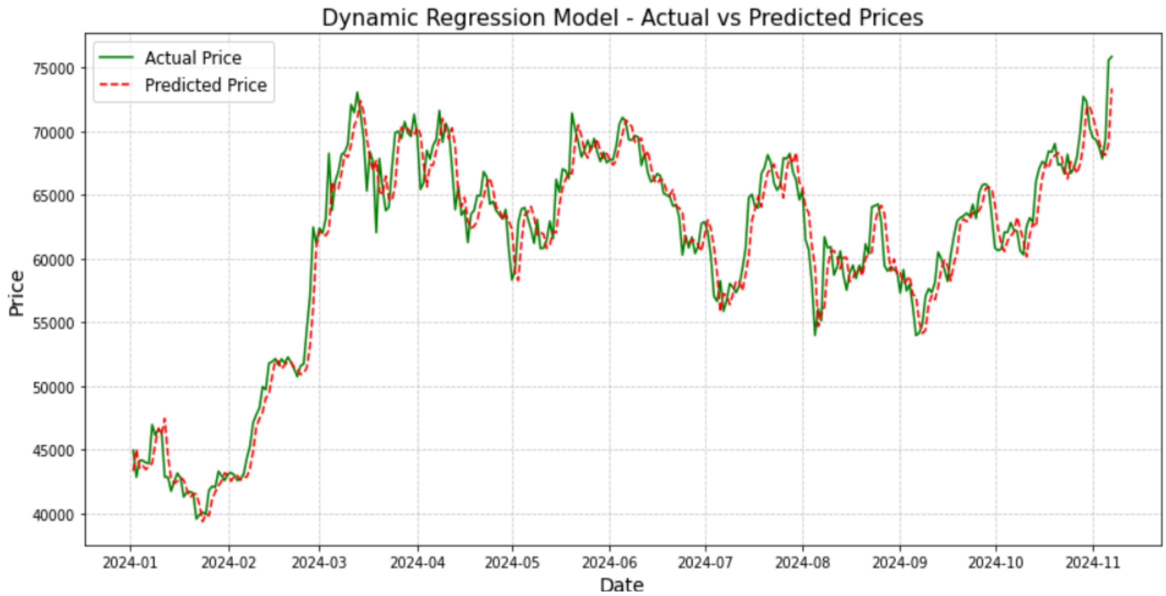

Figure 4: The comparison of the true values and forecasts of dynamic model

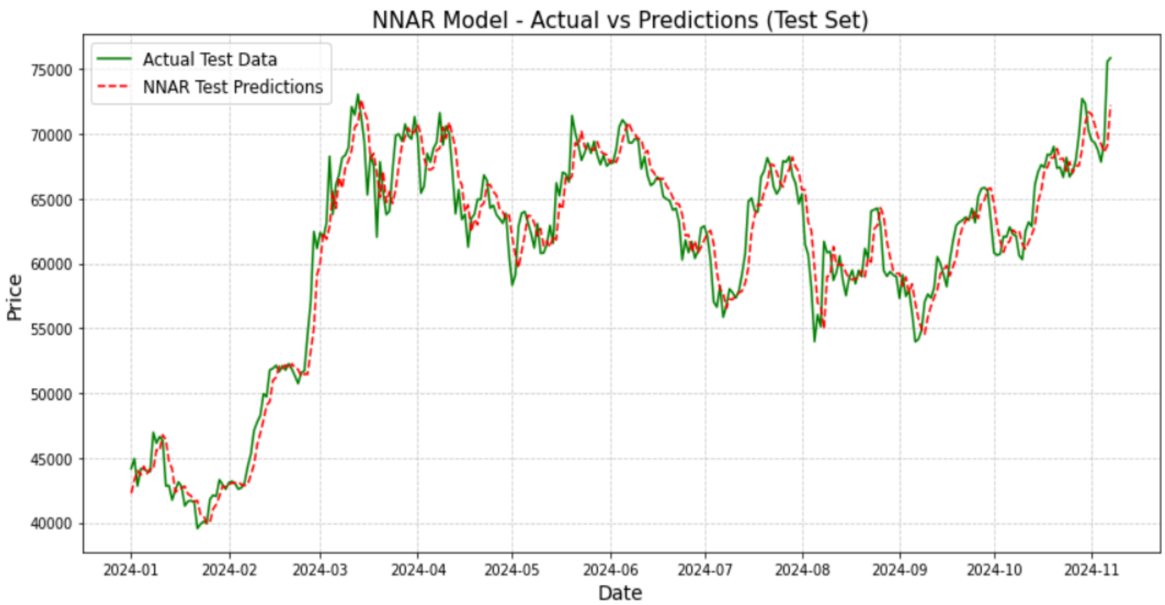

Figure 5: The comparison of the true values and forecasts of NNAR