Price Forecasting using Time Series Analysis on Cryptocurrency Data

Rajvardhan Verma (2330504) under Prof. Tushar Sandhan

June 22, 2023

Abstract

Virtual currencies have been declared as one of the financial assets that are widely recognized as exchange currencies. The cryptocurrency trades caught the attention of investors as cryptocurrencies can be considered as highly profitable investments. Cryptocurrency, being a novel technique for transaction systems, has led to a lot of confusion among investors and any rumours or news on social media has been claimed to significantly affect the prices of cryptocurrencies. The huge percentage increase/decrease in its price over a short period of time is an intriguing phenomenon that cannot be foreseen, hence cryptocurrency price prediction has been a hot topic of study.

I studied some basic regression models which use feature extraction, and implemented the popular ARIMA model. Further, I moved to deep learning models and chose the LSTM model since it is much less susceptible to vanishing gradient problem. From the results, it is clear that ARIMA model is quite efficient in making prediction in short span of time, but as the time grows, the precision rate would decrease. However, after training, the LSTM could make prediction more efficiently, and the precision rate is also higher. In general case, taking less previous data to make prediction in LSTM could lead to better result.

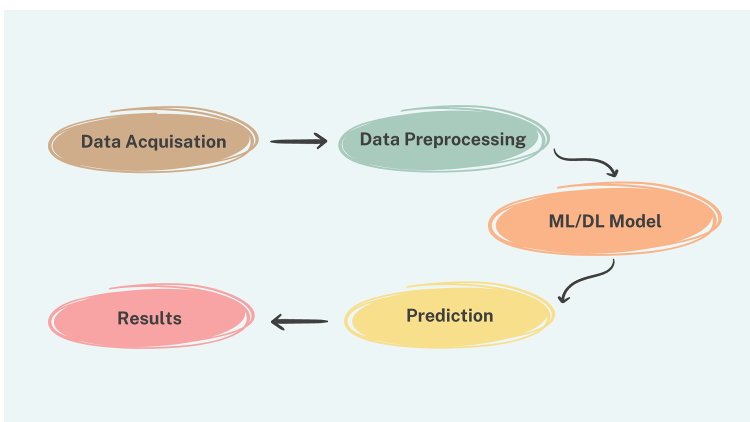

Figure 1: General Pipeline for Analysis of Prediction Models

1 ARIMA Model

1.1 Data Collection & Preprocessing

The cryptocurrency datasets used in my study includes *Bitcoin* and *Ethereum* which contains historical prices for the past five years.

The datasets that I have downloaded had some missing points, which should be handled since many ML models can't handle missing data. To fill the missing points, I have taken first k instances closer to the missing value instance, and then get the mean of that attribute related to the k-nearest neighbors (KNN). This method is called **KNN Imputer**.

The ARIMA models are valid only for stationary dataset. These are the two methods that I have used for checking stationarity.

- **Rolling Statistics -** Plotted the moving average or moving standard deviation to see if it varies with time. It is a visual technique to check for stationarity.
- **ADF Test -** Augmented Dickey–Fuller test is used to gives us various values that can help in identifying stationarity. The test gives a p-value, if this value is less than 0.05, our data is stationary.

To make the datasets stationary, the seasonality and trends from the series were reduced by using the differencing techniques.

1.2 Forecasting & Results

After preprocessing the data, model was trained and tested for both the crytocurrencies, Bitcoin and Ethereum. The RMSE was found to lower for Ethereum since there is less sudden variation in Ethereum as compared to Bitcoin.
The original and predicted data is plotted to observe how well the model predicts the test data.

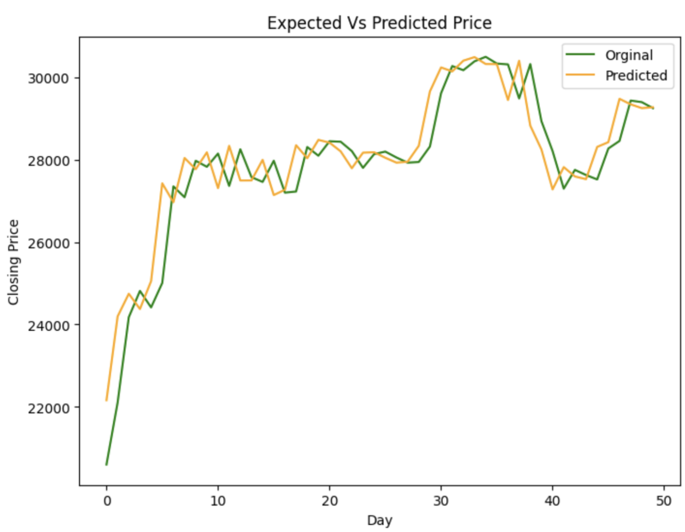

Figure 2: Plot of original and predicted values by ARIMA

2 LSTM Model

2.1 Data Collection & Preprocessing

The datasets used are similar to ones used in ARIMA model to draw a comparision between the two models. Also, the noise is handled by removing the *outliers* in dataset and the missing points are filled with suitable values using KNN Imputer method of gap filling.

To help the LSTM model to converge faster, it is important to scale the data. We have used **Min-Max scaler**, which is one of the most common scalers and refers to scaling the data between a predefined range (usually between 0 and 1). This method is beneficial for Neural Networks since they don't assume any data distribution.

$$\text{Scaled Value} = \frac{\text{Value} - \text{Min Value}}{\text{Max Value} - \text{Min Value}}$$

2.2 Forecasting & Results

After normalization, dataset was divided into train and test data. Then we train the LSTM model, which contains 2 LSTM layers and 2 dense layers.
The RMSE is calculated for the predicted values, and we found that the model is quite accurate for Ethereum and fairly accurate for Bitcoin as well.

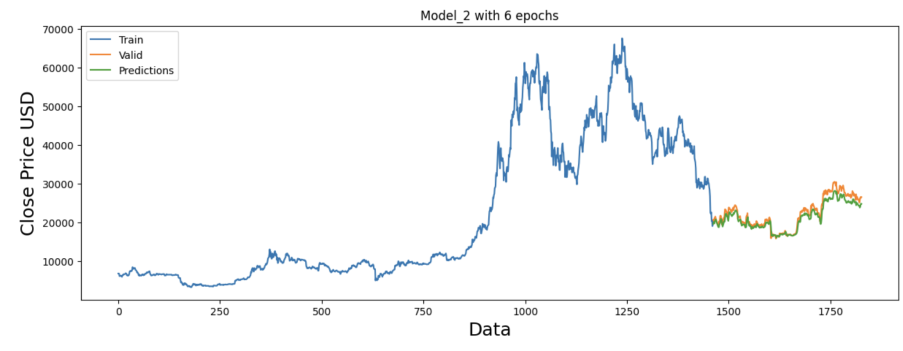

Figure 3: Plot of original and predicted values by LSTM

Results compiled (RMSE): The formula for Root Mean Square Error(RMSE) is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Model	**Bitcoin**	**Ethereum**
ARIMA	661.3740	193.3512
LSTM	127.1942	5.0199

Hence, the deep learning model LSTM outperforms the basic machine learning model ARIMA in terms of accuracy of prediction.

- Special Thanks to my mentor Mr. Pranjal Dubey who regularly guided me and helped me reach this far in this research project.

- Please note that the links to my work are accessible through clickable coloured headings.

Signature of Professor: