**CAPSTONE PROJECT**

**Predicting the price of Bitcoin**

**By: Rojesh Dhakal**

**15/04/2023**

Table Of Contents

Problem of Statement

Over the years, Bitcoin has gained significant recognition and at attention all around the world as a digital currency which can also be the alternative to traditional financial systems. Predicting the price of Bitcoin has been challenge due to its volatile nature and potential for rapid fluctuations in value. Therefore, it is a major problem for investors and traders. The purpose of this project is to develop a machine learning model through which prediction of price of Bitcoin can be done by looking at its historical data.

The project will focus on downloading data from y-finance and analyzing data of historical Bitcoin prices. The data obtained will be preprocessed and then be divided into training and testing sets and then fitted into two different machine learning models and evaluated.

Two machine leaning algorithms, ARIMA and LSTM model, will be trained and evaluated on the dataset to identify the best performing model. The model’s performance will be evaluated using a range of metrics, including mean squared error, mean absolute percentage error.

The goal of this project is to develop a machine learning model that can accurately predict the future price of Bitcoin based on historical data. The insights gained from this paper can be used by investors to make informed decisions about buying and selling Bitcoin and mitigate risk in the rapidly evolving world of cryptocurrency trading.

Business Question

People always invest in assets like Bitcoin by thinking that the price will go up and generate wealth from it. The business question would be how we can use machine learning techniques to predict the price of Bitcoin and how can we identify patterns to minimize losses?

Data Question

The data question would be what the historical price trend of Bitcoin would be. The main objective of this question is to collect and analyze the historical data on Bitcoin prices over time to identify various patterns and trends which might be predictive of future price movements.

Data

The data was downloaded using API from y-finance. Data ranges from 2014 to 2023. The data is clean, and no imputations are required. Initially, the datasets had 3110 rows and 10 columns.

Data Science Process

Data Analysis

For further analysis Year, Month & Days columns were combined as one column using pandas datetime format and saved as new column named as date and set that as index and then later dropped the other columns. Similarly, Dividends and Stock Splits columns were dropped as well, as it contains 0 as a value throughout.

Only the column named Close was used for further analysis because predicting the closing price of Bitcoin. Other columns were dropped because of multi collinearity.

Modelling

ARIMA model

ARIMA model is best used when we have time series. It has three components:

AR- Autoregressive dependent relationship between observation and some lagged observations(p)

I- Integrated used for differencing raw observation subtracting one observation from another from previous timestamps to make time series stationary(d)

MA- Moving Average dependency between an observation and residual errors from a moving average model applied to lagged observations It takes account in past values to predict future values(q)

Lagging a time series means to shift its values forward one or more-time steps, or equivalently, to shift the times in its index backward one or more steps.

The main assumption for ARIMA model is that time series should be stationary. We can check the stationarity by ADF test.

Stationary time series has statistical properties or moments (e.g., mean and variance) that do not vary in time.

Stationarity Test

We can see how time series can be stationary by ADF test.

ADF tries to prove null hypothesis wrong.

null hypothesis: not stationary (p-value>0.05)

alternate hypothesis: stationary

Results:

ADF Statistic: -1.549358057144826

p-value: 0.5089688390430713

Since p-value is more than 0.05, we can infer that it’s not stationarity therefore we need to find the order of differencing.

LSTM model

Outcomes

After EDA and correlation process helped to identify highly correlated features and then later removed them because of multicollinearity. LSTM model gave better results than ARIMA model. The process confirms the possibility to predict the price of bitcoin with high accuracy.

Implementation

The model predicted the price of Bitcoin with high accuracy, and it can be implemented by the investors if they want to predict the price. However, more features like news can be added to predict because the price can be hugely affected by different news around the world which can be analyzed by sentiment analysis.

Data answer

Business answer