Price Forecasting using Time Series Analysis on Cryptocurrency Data

Rajvardhan Verma (2330504), Department of Electrical Engineering, IIT Kanpur

Mentor: Prof. Tushar Sandhan, Department of Electrical Engineering, IIT Kanpur

INTRODUCTION

The decentralized nature of cryptocurrencies, such as Bitcoin, Ethereum, and Litecoin, have attracted both investors and traders, leading to a highly **volatile** and **dynamic** market. As a result, accurately predicting cryptocurrency prices has become a challenging task for market participants.

Therefore, there is a growing need to explore advanced machine learning techniques to improve the accuracy and reliability of cryptocurrency price prediction.

OBJECTIVE

My research aims at developing an improved cryptocurrency price forecasting technique by studying and analysing the pre-existing state-of-the-art methods like ARIMA, LSTM, GRU, etc.

This is done using data preprocessing methods, sentiment analysis and various **machine learning** algorithms.

METHODOLOGY

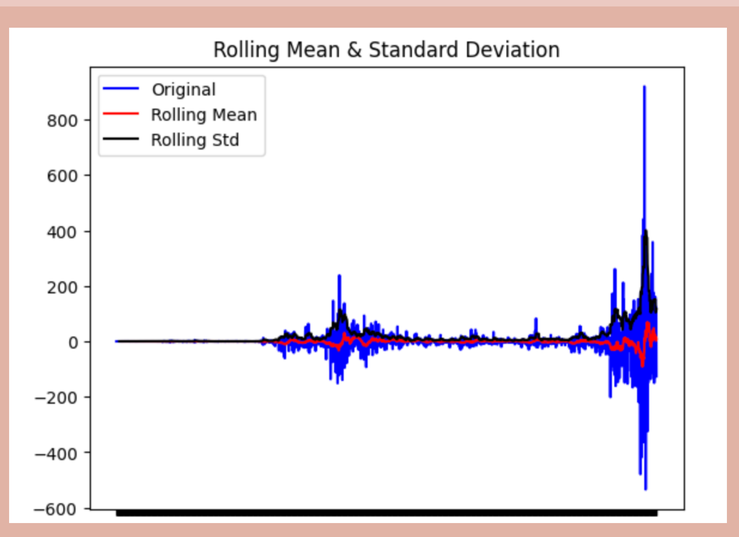
- The preprocessing and collection of data is done using Pandas and data was taken from kaggle and yfinance.
- Analysis of public sentiments is done using twitter dataset
- ARIMA, a statistical analysis model, is used as our baseline to better analyse the dataset.
- We improved our prediction accuracy by understanding deep learning where we implemented another state-of-the-art LSTM model using python's machine learning libraries.
- We have used 80% of the data for training and the remaining 20% for testing and validation. We have then made a sliding window of 60 days which will be used to predict the price of the next day.
- RMSE score is calculated for the predicted values and used as measure of validation.

DATA PREPROCESSING

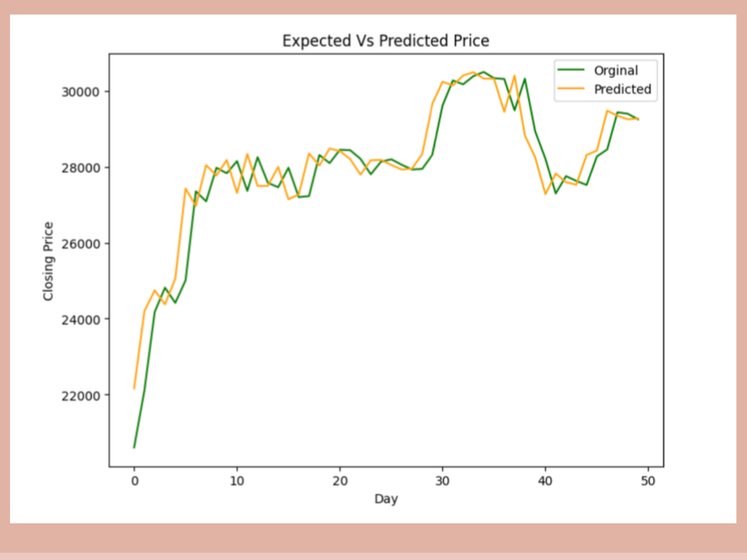
- We have conducted our analysis mainly on two cryptocurrencies, Bitcoin and Ethereum.
- To fill the missing points, I have taken first k instances closer to the missing value instance, and then get the mean of that attribute related to the k-nearest neighbors (KNN). This method is called **KNN Imputer**.
- We have used **Min-Max scaler**, which is one of the most common scalers and refers to scaling the data between a predefined range. This method is beneficial for Neural Networks since they don't assume any data distribution.

ARIMA MODEL

- ARIMA (Autoregressive Integrated Moving Average) is a popular time series forecasting model that combines autoregressive (AR), differencing (I), and moving average (MA) components.
- The **stationarity** of the time series data is assessed using **ADF test**, since this model works only on stationary data.
- If the data is not stationary, differencing is performed to transform the data into a stationary series. Differencing removes trends or seasonality from the data.

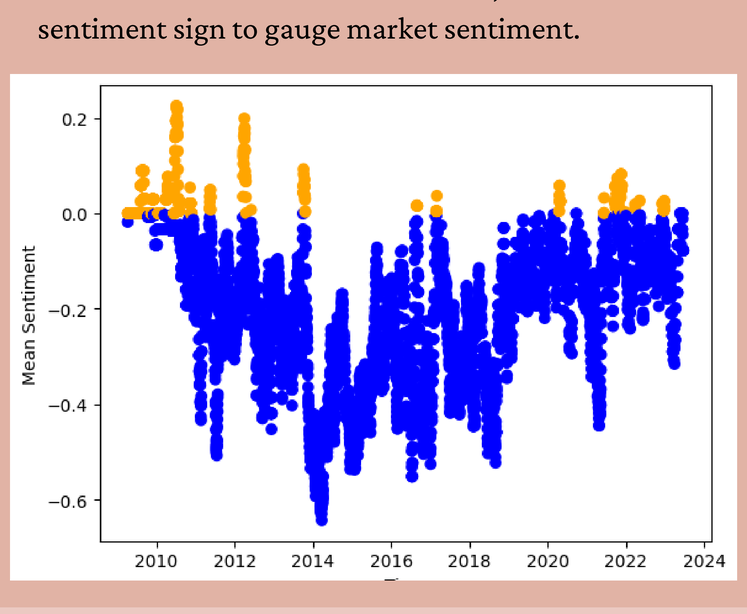


- The appropriate values of p, d, and q are determined by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots.
- Once the model is fitted, it can be used to generate forecasts for future time periods.If the model performance is not satisfactory, the process can be repeated by adjusting the parameters or considering other variations of the ARIMA model (e.g., seasonal ARIMA) until an acceptable level of accuracy is achieved.



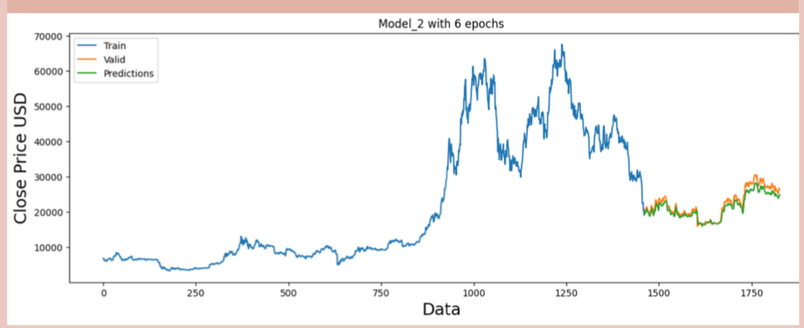
SENTIMENT ANALYSIS

- Sentiment analysis, often known as opinion mining, is a natural language processing (NLP) technique that includes identifying the sentiment or subjective information represented in a piece of text.
- It seeks to extract and categorise the feeling or emotion expressed by a text, whether good, negative, or neutral.
- We conducted a sentiment analysis on tweets data related to Bitcoin and other cryptocurrencies.
- For the analysis, we used the **transformers** library's **Sentiment pipeline**. It was used to identify whether the sentiment was positive or negative and to provide a score to it. We used this score in conjunction with the sentiment sign to gauge market sentiment.



LSTM MODEL

- The LSTM (Long Short-Term Memory) model is a type of recurrent neural network (RNN) that is commonly used for sequence modeling and prediction tasks, particularly with time series data.
- It is designed to address the issue of **vanishing gradients** in traditional RNNs, which can make it challenging for the network to learn long-term dependencies.
- 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.



RESULTS

- The RMSE for Bitcoin and Ethereum is 661.3740 and 193.3512 using ARIMA model. Moreover, it is clear that this model is quite efficient in making prediction in short span of time, but as the time grows, the precision rate would decrease.
- The RMSE for Bitcoin and Ethereum is 127.1942 and 5.0199 using ARIMA model. However, after training, the LSTM could make prediction more efficiently, and the precision rate is also higher.

FUTURE WORKS

- We sought to employ Q-Learning as a reinforcement learning training model. We started by defining our Environment.
- The environment is made up of a __init__() constructor, a reset() function to reset the environment, and a step function whose actions are buy and sell. When we buy, we save the closing prices from that day. We reward positively when profit is positive and adversely when profit is negative.
- Using the Chainer library, the code defines an approach for training a Deep Q- Network (DQN) agent. The chainer library was used to construct the Q network architecture, which consists of three linear layers.
- During DQN network training, the model begins by performing random actions to investigate the environment we have established. It gradually swings towards utilising its gained knowledge by picking actions based on the DQN's maximum Q-value prediction. This method is used to teach the DQN agent how to make the best judgements in the given environment.

REFERENCES

- M. Fernandes, S. Khanna, L. Monteiro, A. Thomas and G. Tripathi, "Bitcoin Price Prediction," IEEE journal.
- Chen, J. Analysis of Bitcoin Price Prediction Using Machine Learning. J. Risk Financial Manag. 2023, 16, 51.