Technical Report

Cryptocurrency Price Prediction

30th June 2023

Table of Contents

List of Figures

Description Page No.

1. Introduction

Cryptocurrency, a digital or virtual form of currency, emerged with the invention of Bitcoin in 2009 by an individual or group known as Satoshi Nakamoto. Since then, the history of cryptocurrency has been marked by rapid growth (see Fig. 1), volatility, and evolving technologies.

The unpredictable cryptocurrency market may make predictions difficult, and a variety of methods have been used to study and predict price changes. These methods include fundamental analysis, which determines the inherent worth of cryptocurrencies based on aspects like technology, acceptance, and market demand, as well as technical analysis, which examines past price patterns, trading volumes, and market indicators to create forecasts. Popular cryptocurrency price prediction techniques may also be categorised and forecasted because of their nature, which is quite similar to stocks. Similar methods are used in our proposed research to leverage historical and present digital currencies market data. Our objective is to accurately anticipate the future prices of cryptocurrencies using effective feature engineering and machine learning.

Fig 1. Total Market Cap from Year 2013-Present for all cryptocurrencies

1.1 Proposition

The flow of the proposed work is as follows. First, historical and present cryptocurrency price data is collected from Yahoo Finance through the web scraping technique. This data includes features like opening price, closing price, high and low prices, and trading volume. Next, the collected dataset undergoes preprocessing to handle missing values, outliers, and any necessary transformations. The Prophet library, developed by Facebook, is then employed for machine learning. The trained Prophet model is evaluated using suitable metrics to assess its accuracy. Finally, the prediction system is interfaced with Streamlit, a user-friendly framework for building interactive web applications, to provide a user interface where users can input parameters, view predictions, and interact with the system.

1.2 Success Metrics

Success metrics are numerical measurements that are used to assess the performance and efficacy of a project. They offer a means of determining if the work has accomplished its goals and produced the expected results. Success indicators aid stakeholders and project teams in monitoring development, forming wise judgements, and assessing the project's overall impact. Potential success measures would likely be developed in the context of the "Cryptocurrency Price Prediction" project to evaluate the efficacy as well as the accuracy of the suggested model. The following are some possible project success indicators:

- Accuracy of Directional Predictions: This indicator assesses how well the model predicts the general trend of price changes, such as whether they will rise or fall. The percentage of accurate directional forecasts relative to actual price movements can be used to quantify it.
- R2 Score: The R2 score, sometimes referred to as the coefficient of determination, calculates how much of the variance in the independent variables (prediction features) can be accounted for by the dependent variable (cryptocurrency prices). A better match between the model and the data is shown by a higher R2 value.
- Visual Representation of Return on Investment (ROI): We compare the profits adhering to the anticipated buy/sell information to a benchmark or a baseline approach. This can be graphically visualized by observing the various graphs and plots in the system and making an informed decision.
- User Satisfaction: We can obtain feedback from users or other interested parties to gauge their satisfaction with the programme and the precision of the forecasts. User engagement data, polls, and user reviews can each be employed to do this.

2. Data Definition & Exploration

The dataset used by the cryptocurrency prediction system is taken from the Yahoo Finance website. Yahoo Finance is a well-known financial website that offers an extensive range of financial data, including stock and cryptocurrency statistics. Investors and data scientists may access historical and real-time market data, news, charts, and other financial indicators through its comprehensive platform.

The *yfinance* library was used to retrieve both current and historical data from Yahoo Finance. A popular and useful Python module called *yfinance* offers a user-friendly interface for gaining access to and downloading financial data from Yahoo Finance. In order to personalise data retrieval, *yfinance* provides a wide range of functions and attributes.

It allows users to specify the ticker symbols for the desired cryptocurrencies, define the time interval for data retrieval (e.g., daily, weekly, monthly), and choose the specific data attributes to retrieve, such as :

- Date and Time: The timestamps indicate the date and time when the data point was recorded.
- Opening Price: The price of the cryptocurrency at the beginning of the specified time interval.
- Closing Price: The price of the cryptocurrency at the end of the specified time interval.
- High and Low Prices: The highest and lowest prices reached by the cryptocurrency during the specified time interval.
- Volume: The trading volume, representing the number of units of the cryptocurrency traded during the specified time interval.

Therefore, an extensive dataset may be constructed by accessing Yahoo Finance for these characteristics, allowing the prediction engine to analyse previous trends and create projections based on the extracted digital currencies data.

2.1 Data Preprocessing

Data preprocessing refers to any sort of processing done on raw data to get it set up for another data processing technique.
However, our employed machine learning model already preprocesses much of the data for us. Hence no explicit preprocessing or feature extraction technique was used in our methodology.

Nevertheless, some small changes in data for preparing the data were still added to increase the readability of the dataset along with making it more efficient for our model training. This includes:

- Using the reset_index() method, the code adjusts the index of the downloaded data. By doing this, we can be trusting that the data frame will have a clear and orderly index for processing.
- Using the rename() method, the code renames the columns in the DataFrame. To comply with the specifications of the Prophet library, which is used for time series analysis and forecasting, it gives the columns more evocative names, such as changing "Date" to "ds" and "Open" to "y".

3. Machine Learning Model (Facebook's Prophet)

3.1 Overview

Facebook Prophet is an open-source time series forecasting library. It was initially launched in 2017, and as a result of its ease of use, adaptability, and capacity to handle a range of time series forecasting jobs, it has grown in popularity. Its features include:

1. Automatic Seasonality Detection: Without the need for explicit feature engineering, Prophet automatically recognises and models a variety of seasonality patterns in the data, including weekly and annual trends.

2. Trend Modelling: Forecasting is made more precise because of Prophet's ability to identify both short- and long-term patterns in time series data.

3. Holiday Effects: Prophet gives users the option to account for the effects of well-known holidays and other special occasions on the time series by including them in the forecasting model.

4. Estimating Uncertainty: Prophet creates uncertainty intervals around the predicted numbers, giving users a sense of how confident the forecasts can be.

5. Extensibility: Prophet is designed to be flexible and allows for the inclusion of additional regressor variables that may have an impact on the time series.

3.2 Implementation

In our implementation, the Prophet model is used by invoking the following library and also provides convenient functions for visualizing the forecasted results, such as the *plot_plotly* and *plot_components_plotly functions* used in the code to generate interactive Plotly graphs.

from prophet import Prophet
from prophet.plot import plot_plotly, plot_components_plotly

The model is then invoked while training as shown in the below code snapshot.

```
m = Prophet(seasonality_mode="multiplicative")
m.fit(df)

future = m.make_future_dataframe(periods=365)
forecast = m.predict(future)
```

Fig 2. Code Snapshot showing the call for Prophet model

The seasonality mode is taken as multiplicative which means that when this mode is selected, Prophet decomposes the time series into multiple components, including trend, seasonality, holidays, and error.
Finally, the ***fit*** method is used to train the model on the historical data, and the ***predict*** method is used to generate future price predictions.

3.3 Interfacing

The front-end development of the application was done through the Streamlit software. Streamlit is an open-source Python library that simplifies the process of creating interactive web applications for data science and machine learning projects. It enables developers to quickly build and deploy applications without extensive web development knowledge. One of the key advantages of Streamlit is its ability to automatically update the application interface as the code changes.
In our implementation, we used a wide range of built-in features and widgets for creating user interfaces. This included sliders for sliding across a month or year to visualize forecasted data,

checkboxes to navigate between pages, dropdowns when accessing different cryptocurrencies, and text inputs to write into the application, which can be easily integrated into the application to enable user interaction.
Additionally, we employed Streamlit to display the resulting predicting charts and history of cryptocurrency data, images, and interactive plots.

4. Evaluation

When evaluating the effectiveness and precision of machine learning models, evaluation measures are quite important. They offer numerical measurements that assist in assessing a model's performance and forecast accuracy. The R2 score, usually referred to as the coefficient of determination, is one assessment measure that is frequently employed. The R2 score shows how much of the variation in the dependent variable—the variable being predicted—can be anticipated from the independent variables—the characteristics used to make the forecast.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

The R2 score ranges from 0 to 1, where a value of 1 indicates a perfect fit and 0 indicates no correlation between the predicted and actual values.
The R2 score via running the model on Jupyter Notebook came out to be 0.98, when ran on a single coin for historical data. An R2 score of 0.98 indicates that the model explains approximately 98% of the variance in the dependent variable. This score was observed for the immediate next day. In the project, the confidence level, or confidence interval for each predicted price of the coin is seen to be between 88% to 95%. This indicates that the predicted price has a good probability to land between the upper and lower bound prices of the coin.

5. Project Walkthrough

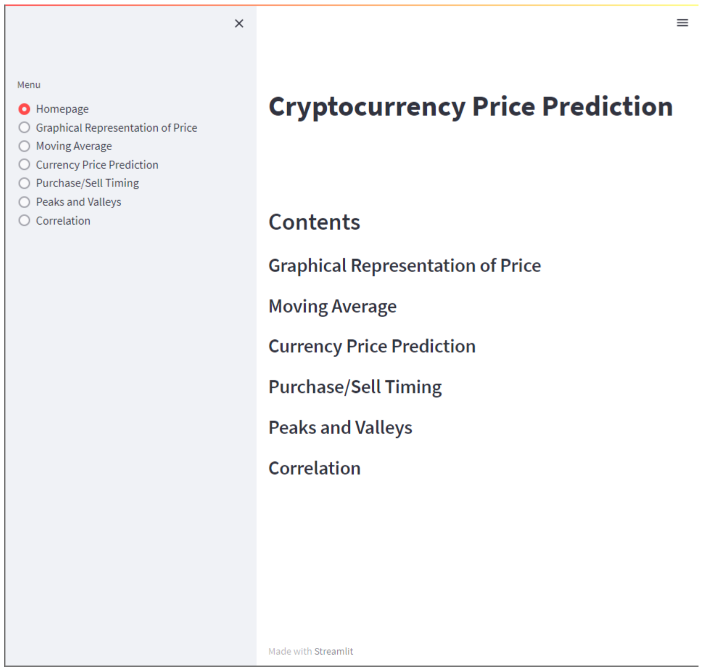

Fig 3. Homepage and Navigation Bar

To better understand the visual aspects of the project, we refer to the attached figures representing the features of the project.

The above Fig 3, displays the homepage which first appears when the application is loaded. Here the task details are shown along with the contents of the application. The navigation sidebar remains constant throughout the whole application and is useful for moving to a different page.

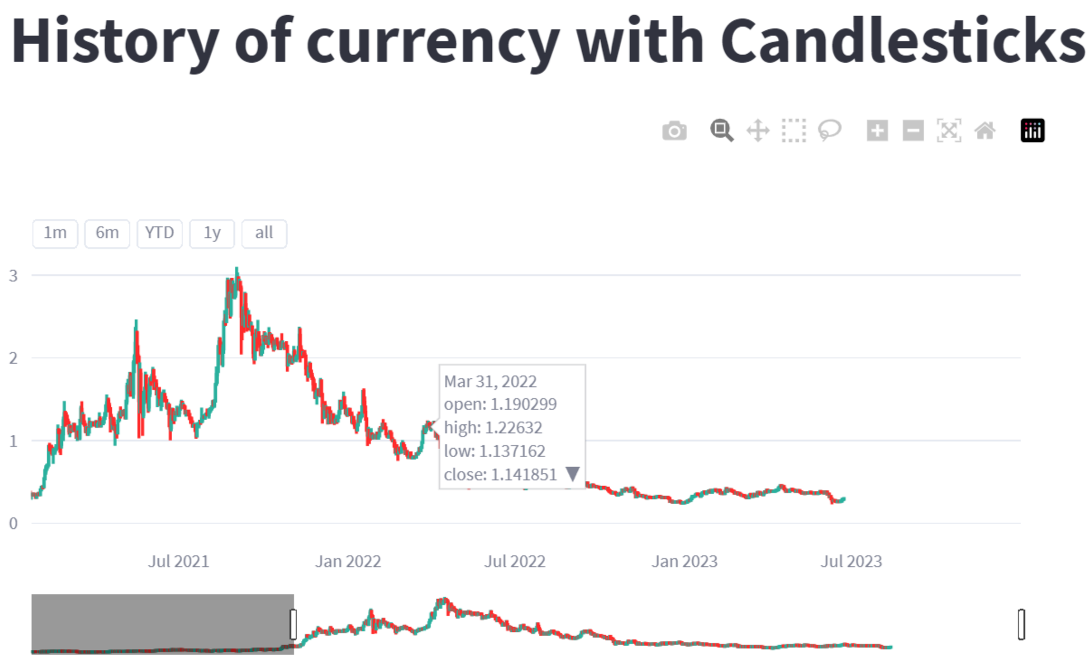

Fig 4. Historical prices for the chosen cryptocurrency

Figure 4 gives the historical cryptocurrency prices of opening from a defined start date up till the current day. Here, the start date is chosen to be 1/1/2020. A slider below the chart is attached for the user to navigate or define the period of time for which he/she wants to observe the prices.

Fig 5. 30-Day Moving Average

The next page shown as Moving Average displays a month's total moving average for the selected cryptocurrency up to the current date. A moving average (MA) is a cryptocurrency indicator used often in technical assessment for generating a coin's moving average to create a continuously updated average price in order to help even out the price data. The investment is in an uptrend if its moving average is increasing, whereas a downtrend is indicated by a decreasing moving average. A slider containing a range from 1-30 days is included to assess the cryptocurrency over a month's period.

The most important plot of the project is the price prediction plot which is generated as a result when the Prophet model is trained. The black dots shown in Fig 6. highlights the actual prices of the user-entered cryptocurrency. The intermixed blue line shown together is the predicted or forecasted price. By observation, it's evident that the predicted line does not falter in many of the periods and is actually close to the opening price of the actual cryptocurrency.

Fig 6. Cryptocurrency Price Prediction for historical and future dates.

We can also observe that the opening prices for future dates are also incorporated which can be seen just before the onset of July 2023 time period. It's important to note that the code of the trained model is set in such a way so that it inculcates future prices for up to 2 years.

Predictive Highest Price(Over Span of Two Years):

`0.6411590984423007`

Predictive Lowest Price(Over Span of Two Years):

`0.10950317403597024`

Fig 7. The highest & lowest value of a chosen coin

Fig. 7 associated with the page of Peaks & Valleys gives an output of a numerical value of the predictive price of a chosen cryptocurrency signifying the highest and the lowest achievable price by a coin over the span of two years.

Currency Price Prediction

Best time to purchase: 2023-07-26 01:03:23.537856

Best time to sell: 2024-09-02 01:03:23.537856

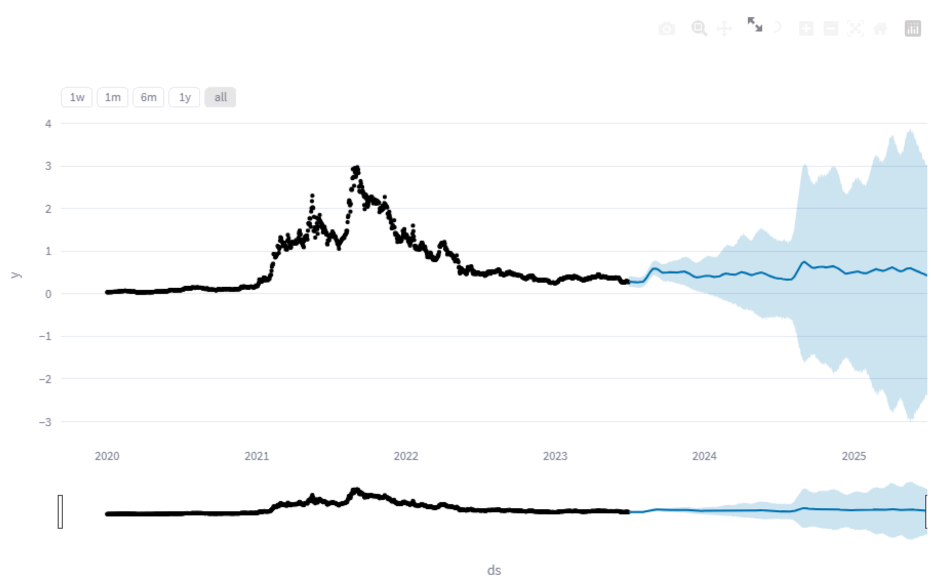

Fig 8. Best time to sell/purchase a cryptocurrency

On similar prediction criteria of the Prophet model, the Purchase/Sell page shows the best time for a chosen cryptocurrency to be purchased or sold. The time of purchase is signified by the drop in the price of the coin over the span of two years. Similarly, the best time to sell is signified by the highest attainable price of the market for the coin.

In the last pages of the project, we plot the positively and negatively correlated cryptocurrencies compared against each other defining the percentage of correlation. Here the names of the coin are hidden, however on inspection of the diagonal matrix from the heatmap given below we can identify that the most positive correlation occurs when the coin is compared against itself, hence the value of 1. Readers may note that only a handful of coins were selected so as to display a suitable heatmap of the correlation. The range of 1 to -0.5 is calculated only for handpicked cryptocurrencies.

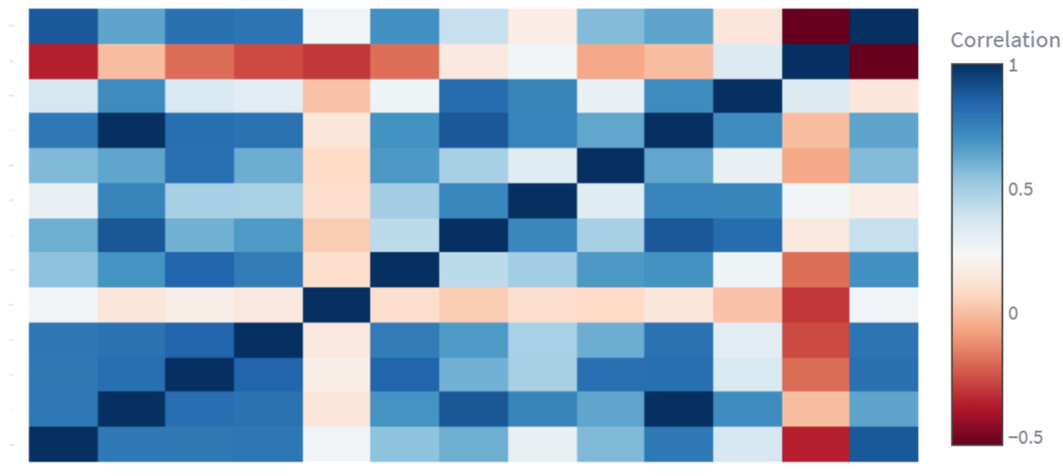

Fig 9. Cryptocurrency Correlation Matrix Heatmap

6. Conclusion

In this study, we suggested a cryptocurrency price prediction platform which employs the use of machine learning to predict cryptocurrency prices accurately. Our methodology involved data collection, preprocessing, and training the Prophet model on the collected dataset. We retrieved the cryptocurrency data from Yahoo Finance using the yfinance package. The R2 score was used to assess the model's performance, and it showed a high degree of validity with a score of 0.98. We used Streamlit for building the web application, to give a user-friendly interface. To aid users in analysing and comprehending the trends and patterns in coin values, historical price charts, candlestick charts, and price forecast plots were created. The prediction plots revealed a strong correlation between the projected and actual prices. Some limitations include no account for transaction costs, such as fees incurred during buying and selling cryptocurrencies. Future works in this area would involve applying deep learning models like Bi-LSTM and GRU to gain more accurate results.

7. References

1. Bitcoin: A Peer-to-Peer Electronic Cash System
2. Time & Seasonality Features in Time Series | by Alvin T. Tan, Ph.D. | Towards Data Science
3. https://finance.yahoo.com/crypto/
4. Chaudhari, A., 2020. *Forecasting cryptocurrency prices using machine learning* (Doctoral dissertation, Dublin, National College of Ireland).
5. Yenidoğan, I., Çayir, A., Kozan, O., Dağ, T. and Arslan, Ç., 2018, September. Bitcoin forecasting using ARIMA and PROPHET. In *2018 3rd international conference on computer science and engineering (UBMK)* (pp. 621-624). IEEE.
6. Time series prediction using Prophet in Python | by Renu Khandelwal | Towards Data Science
7. Goyal, C. (2021). The Game of Increasing R-squared in a Regression Model. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/05/the-game-of-increasing-r-squared-in-a-regression-model/
8. Prophet
9. Streamlit documentation
10. Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2
11. Prophet | Forecasting at scale.
12. Facebook Prophet Algorithm in Time Series Analysis | by Nwosu Rosemary | Analytics Vidhya | Medium
13. https://coinmarketcap.com/charts/
14. Patrick Loeber. (2021, February 7). Build A Stock Prediction Web App In Python [Video]. YouTube. https://www.youtube.com/watch?v=0E_31WqVzCY