Next Hour Bitcoin Prediction (June 2022)

Lamiss Gargouri
Master student in Bahcesehir University
Department of Big Data Analytics and Management.

ABSTRACT Since the previous several years, Bitcoin and other well-known cryptocurrencies have attracted a lot of interest. This cryptocurrency, also known as a digital coin or virtual money, is acquired and traded through the blockchain technology. The use of blockchain technology in the use of cryptocurrencies has drawn criticism from the banking industry, the government, stakeholders, and private investors. Since the launch of Bitcoin in 2009, the emergence of cryptocurrencies has completely overtaken the industry. It is believed that cryptocurrency would eventually displace conventional paper money across the world.

INDEX TERMS Data Mining, Machine Learning, Data Visualization, Exploratory Data Analysis.

I. INTRODUCTION

When it comes to trading, there are several strategies that can be used to determine the best moment to invest. Some may examine the financials to determine if there is worth hidden under the surface. Others could pay attention to the daily news to see how it affects the present pricing. Another group could study historical price fluctuations to see if there are any patterns that can be used to predict future price movements.

What is Bitcoin?

Bitcoin is the first decentralized digital currency, meaning the system works without a single administrator or central bank, you can use them in every country, your account cannot be frozen, and there are no prerequisites or limits.

Bitcoins is known as peer to peer, that means that they are transferred directly from person to person. The cryptographic transactions are entered into a public distributed ledger known as the block chain and validated by a network of individuals. Keep in mind that after a payment has been made, it cannot be undone, and if your wallet is lost, your bitcoins are also lost.

II. BACKGROUND

A. BASIC CONCEPTS

- **Data Mining** consists of extracting patterns and other important information from huge data volumes.
- **Machine Learning** is a subset of artificial intelligence (AI) that enables computer programs to forecast outcomes more accurately without having been expressly taught to do so.
- **Exploratory Data Analysis** is the critical process of doing preliminary analyses on data utilizing statistical findings and visualizations in order to find trends, spot anomalies, test hypotheses, and confirm theories.
- **Data visualization** is referred to as the visual presentation of facts and information. By using visual elements like graphs and charts, data visualization tools make it simple to see and understand patterns and trends in data.

B. DISCUSSION

a. Cryptocurrency critics

The question of whether cryptocurrencies are a type of asset currency has drawn a lot of criticism. Bitcoin and cryptocurrencies are significantly closer to and satisfy the definition of currency in their current form since they can undertake financial transactions. The bulk of the three key features of currency—store value, unit of account, and means of transaction are present in cryptocurrencies, despite the fact that none of them meet the full requirements.

b. Possibilities and Benefits

Considering that cryptocurrencies are a relatively new good, the potential seem great. The benefits and potential for the future are still desired despite an increase in price and value. The following talks about the potential uses of cryptocurrencies for consumers, investors, and even the government.

III. DATASET

A. Context

My dataset is collected from “Binance” which is a platform of trading and investment of cryptocurrency that have a library which is also called “Binance”. This librairy provide me with API that give me the data.
I will send a request to binance server for bitcoin USD every one hour for 100 days ago. The data will be stored in the variable Klines.

B. Content

This API has 3 parameters:

- The 1st parameter is BTCUSDT named coin pair. This parameter will get the coin or the stock that I need.
- The 2nd parameter is the candle time. In our example: 1 hour.
- The 3rd parameter is the time frame length of data. That means X days ago from the second I will execute the program.

⇨ The following are the columns in the supplied dataset:

- OpenTime : The second the candle starts.
- Open: The starting price.
- High: The highest price for that hour.
- Low: The lowest price for that hour.
- Close: The final price.
- Volume: The volume of trades in that our.
- CloseTime: The second the candle ends.
- Quote_asset_volume
- Number of Trades
- Taker Buy base asset volume
- Take buy base asset volume
- Ignore

	opentime	open	high	low	close	volume	close time	Quote_asset_volume	Number of trades	Taker buy base asset volume	Taker buy quote asset volume	Ignore
0	1645999200000	37417.75	37957.87	37346.23	37818.47	2013.80038	1646002799999	75819631.89975240	69133	874.78456000	32952039.47812830	0
1	1646002800000	37818.47	37854.53	37330.23	37699.07	2531.54953	1646006399999	95175555.67011850	70778	1270.43346000	47775593.36343690	0
2	1646006400000	37699.08	37802.04	37450.17	37574.27	1476.92822	1646009999999	55556810.35312730	46892	745.64854000	28048508.06623270	0
3	1646010000000	37574.27	38483.53	37574.27	37970.85	2758.22257	1646013599999	105072407.20162190	60016	1305.96191000	49756708.04003170	0
4	1646013600000	37970.86	38102.58	37559.38	37768.41	1443.75806	1646017199999	54602834.42840220	38041	657.25356000	24857866.08112270	0
5	1646017200000	37768.41	37884.60	37650.04	37793.80	1132.07150	1646020799999	42776271.95853050	34475	590.65314000	22319070.31025550	0
6	1646020800000	37793.80	38044.91	37731.41	37816.22	1015.50894	1646024399999	38454889.11902190	37097	560.04799000	21207299.35257320	0
7	1646024400000	37816.23	38011.81	37738.05	37829.44	1074.48439	1646027999999	40713615.05937880	31134	535.96592000	20306547.31869280	0
8	1646028000000	37829.44	38126.40	37451.56	37979.66	2023.92449	1646031599999	76520803.07507300	57482	976.54896000	36925255.83874780	0
9	1646031600000	37979.67	38478.28	37900.00	38347.58	2329.54711	1646035199999	89018884.32556170	59047	1207.47050000	46141893.50725280	0

IV. IMPORTING LIBRAIRIES & DATA EXPLORATION

To begin the project, import libraries like NumPy and Matplotlib for data visualization, and then load the data using Pandas. . The df.Head(10) function displays the first ten records of the data set, providing a quick overview of the data frame's rows and columns.

V. UNDERSTANDING THE STRUCTURE OF DATA

```
df.describe()
```

	opentime	open	high	low	close	volume	close time	Number of trades
count	2.400000e+03	2400.000000	2400.000000	2400.000000	2400.000000	2400.000000	2.400000e+03	2400.000000
mean	1.650317e+12	37880.580821	38056.368121	37702.684642	37878.065696	2358.800477	1.650321e+12	51088.520833
std	2.494673e+09	5712.716261	5710.348400	5712.023273	5714.250695	2330.408206	2.494673e+09	36296.423747
min	1.645999e+12	27114.840000	27867.970000	26700.000000	27114.840000	240.245560	1.646003e+12	11164.000000
25%	1.648158e+12	31043.340000	31368.992500	30680.450000	31043.332500	1057.280798	1.648162e+12	30294.750000
50%	1.650317e+12	39221.085000	39433.365000	39051.950000	39221.090000	1635.388920	1.650321e+12	40379.500000
75%	1.652476e+12	41867.907500	42045.182500	41717.370000	41867.907500	2758.510775	1.652480e+12	59962.750000
max	1.654636e+12	47970.980000	48189.840000	47811.400000	47970.990000	28613.983810	1.654639e+12	447737.000000

⇨ Pandas describe() is used to view some basic statistical details like percentile, mean , std , etc. of a dataframe or a series of numeric values. When I applied this method on my dataset, It returns a different output which is shown in the table below. The return type is a statistical summary of data frame.

```
df.shape
```

```
(2400, 12)
```

⇨ Pandas shape() returns the shape of my data which is 12 colums and for each columns there are 2400 data.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2400 entries, 0 to 2399
Data columns (total 12 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   opentime                        2400 non-null   int64
 1   open                            2400 non-null   float64
 2   high                            2400 non-null   float64
 3   low                             2400 non-null   float64
 4   close                           2400 non-null   float64
 5   volume                          2400 non-null   float64
 6   close time                      2400 non-null   int64
 7   Quote_asset_volume              2400 non-null   object
 8   Number of trades                2400 non-null   int64
 9   Taker buy base asset volume     2400 non-null   object
 10   Taker buy quote asset volume   2400 non-null   object
 11  Ignore                          2400 non-null   object
dtypes: float64(5), int64(3), object(4)
memory usage: 225.1+ KB
```

⇨ Pandas info() is used to help us get the concise summary of the dataframe (number of the data for each columns / if it's null or not / the type of the data/ and the memory usage which is 225.1+ KB)

```
df.isna().any()
```

```
opentime                          False
open                              False
high                              False
low                               False
close                             False
volume                            False
close time                        False
Quote_asset_volume                False
Number of trades                  False
Taker buy base asset volume       False
 Taker buy quote asset volume     False
Ignore                            False
dtype: bool
```

⇨ Pandas unsa() any() detect missing values for and array. Used to check whether any element is true, potentially over an axis. And in our case there is no missing values.

```
df.isnull().sum()
```

```
opentime                          0
open                              0
high                              0
low                               0
close                             0
volume                            0
close time                        0
Quote_asset_volume                0
Number of trades                  0
Taker buy base asset volume       0
 Taker buy quote asset volume     0
Ignore                            0
dtype: int64
```

⇨ Pandas isnull() sum() returns the number of missing values in the dataset. And as long as there is no missing values in our case the result will be 0 for each column.

VI. VISUALIZING THE DATA

a. Definition

- RSI: Relative Strength Index, sometimes known as RSI, is a scientific analysis indicator which essentially gauges the strength and speed of market oscillations. The RSI is a momentum indicator as a result, measuring the size and speed of market changes.
  The RSI equation: RSI = 100 - 100 / (1 + RS)

- CCI: A momentum indicator called the Commodity Channel Index (CCI) is used to identify pricing peaks, price changes of direction, and trend strength.
  The CCI equation: CCI = (Price – MA) / 0.015 x D

- ROC: A momentum-based technical indicator called the Price Rate of Change (ROC) calculates the change in its price between both the present price and the price from a predetermined number of time periods ago.
  The ROC equation: ROC=((Closing Price*p*-*n*Closing Price*p*-n) /Closing Price*p*-*n*)×100

- MOM: When a financial product (such as a stock, option, future, or currency) is going up or down in the marketplace, traders can use the momentum indicator to assess its momentum.
  The MOM equation: Momentum = (Current closing price)- (Closing price n periods ago)

b. Correlation Matrix

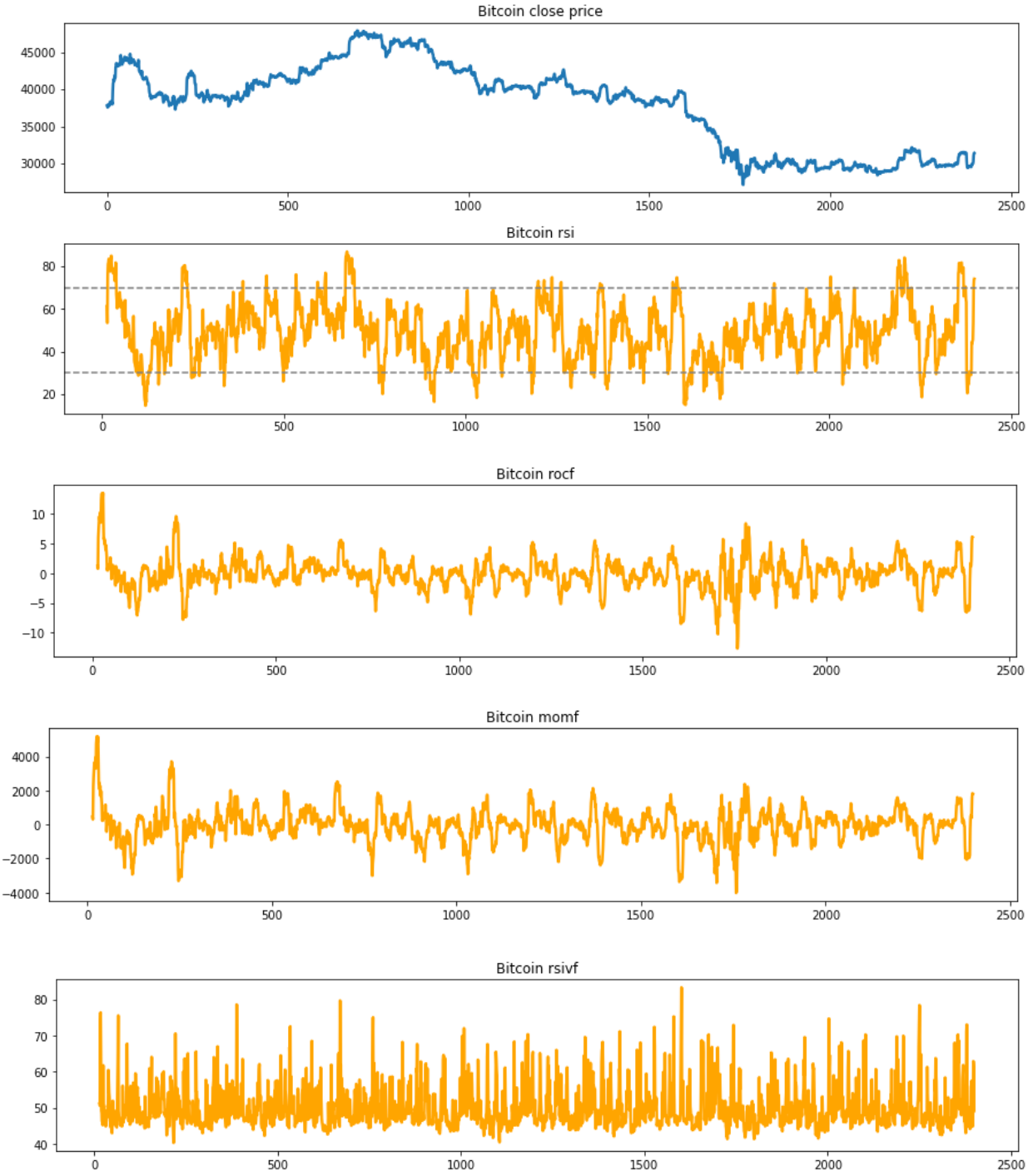
⇨ We will use visualizations to look at different combinations of characteristics while studying the data. This will assist us in better understanding our data and provide some insight into data patterns.

This correlation Matrix was created to illustrate relationship of two variables, one on each axis. I can see whether there are any trends in frequency with one or even both variables by looking at how cell colors vary across each axis.

⇨ To predict the bitcoin price, the best indicators are: RSI, ROC, MOM. Also, we will eliminate cci because it's the least correlated.

c. Plots

These are the plots of Bitcoin Close price, The Bitcoin RSIF that takes 14 candles backwards. If the RSI gets up in other words overbought it means that the bitcoin was bought in a short period of time and it will get down soon. But if if gets down in other words oversold, it means that the bitcoins was sold in a short period of time and it will get up soon.



- The plot of ROCF gives us the percentage of how much the bitcoin increased or decreased based on 14 candles backwards.
- The plot of MOM or let's say momentum indicator compares the current price with the previous price from 14 candles ago.

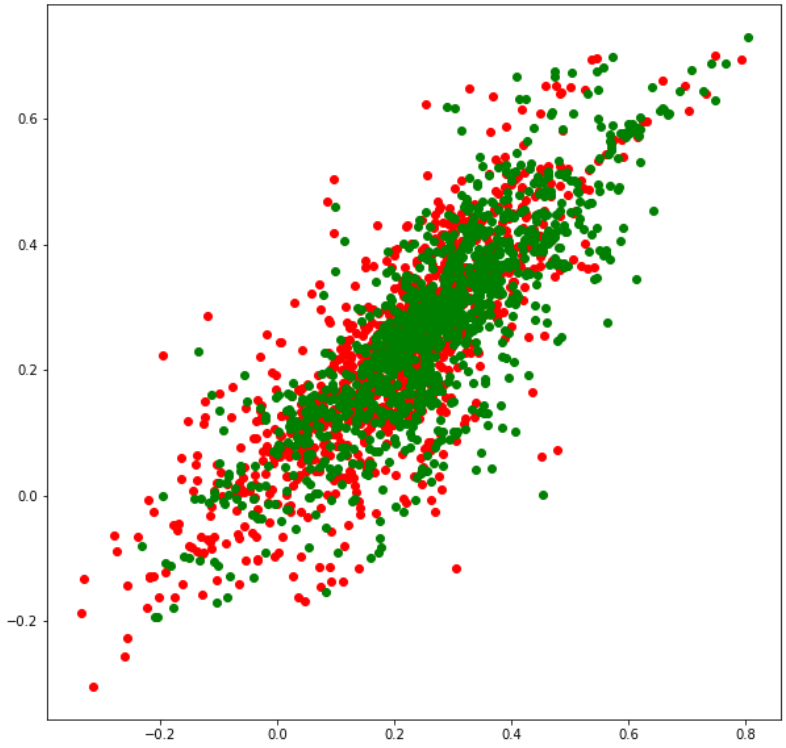
- The plot of RSIVF gives as a summary on how much bitcoin was sold and bought in the previous 14 hours.

⇨ These plots are made to compare the bitcoin price with its indicators to see if I can just predict visually, without programming or not.
⇨ We can also make plots for the slow indicator which means in our case 28 candles back.

d. KNN Model

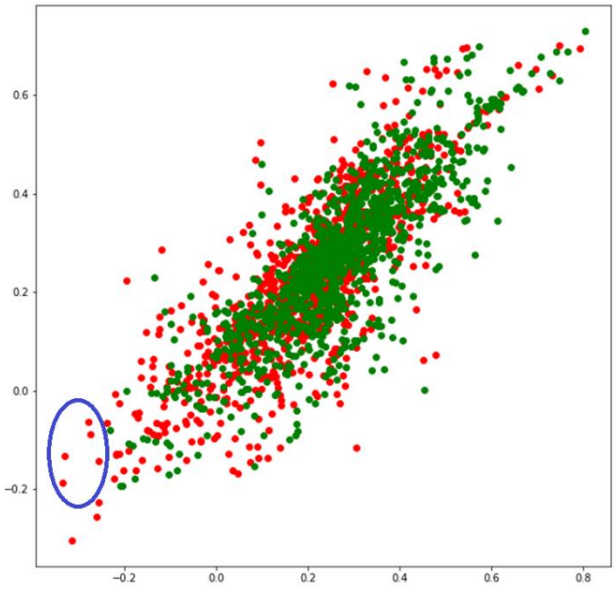
- A collection of data is essentially divided into classes using the training data notion of classification in deep learning.

- The k-nearest neighbors' algorithm, sometimes referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to produce classifications or forecasts about the clustering of a single data point.

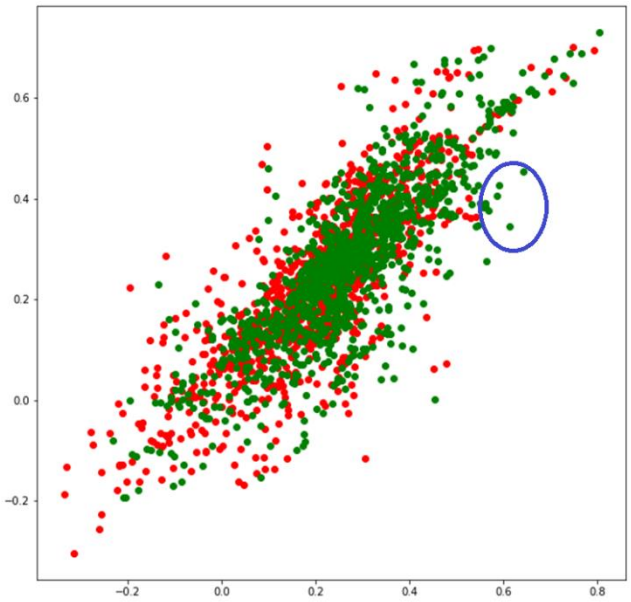


To make this knn plot, we have first normalised the 10 features to make them between -1 and 1. then I've devided them in 2 arrays the first one named features one contains all the fast indicators and the second array named features 2 contains all the slow indicators. Next we will visualize every hour, if the next hour the bitcoin increase it's a green or dot and if it decreases it's a red dot.

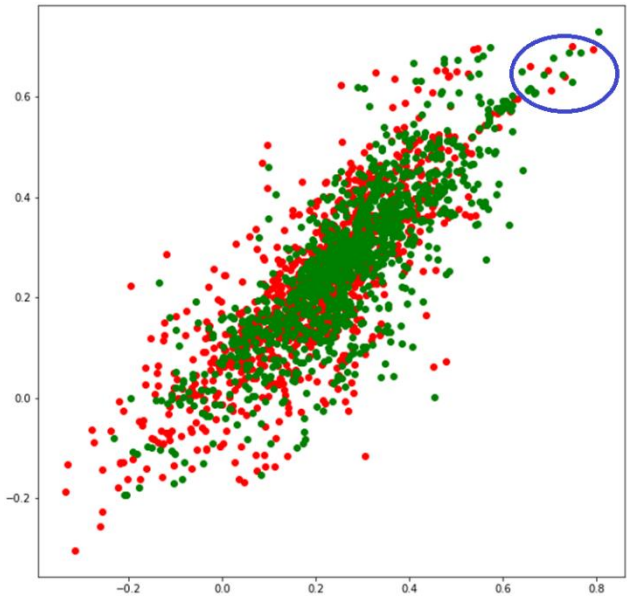
⇨ And we will experience this by changing the coefficient in order to seperate the red dots from the green dots.
⇨ The machine learning will make it easier until it finds the result, but in hour case, while programming we will change the coefficients until the result is more or less acceptable.



Let's make some examples. when the dot of our prediction for the next hour will be in this blue circle
➔ means that the bitcoin in the next hour will increase.

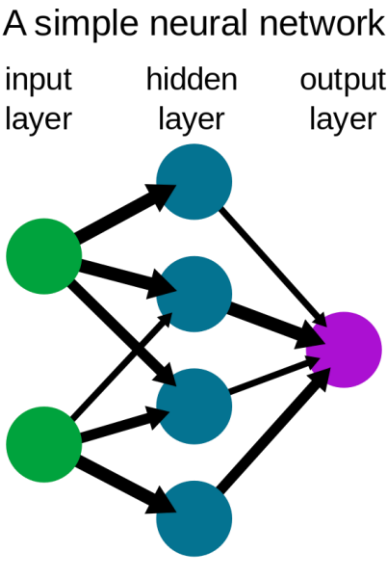


When the predicted dotwill be in this blue circle
➔ means that the bitcoin for the next hour will increase.



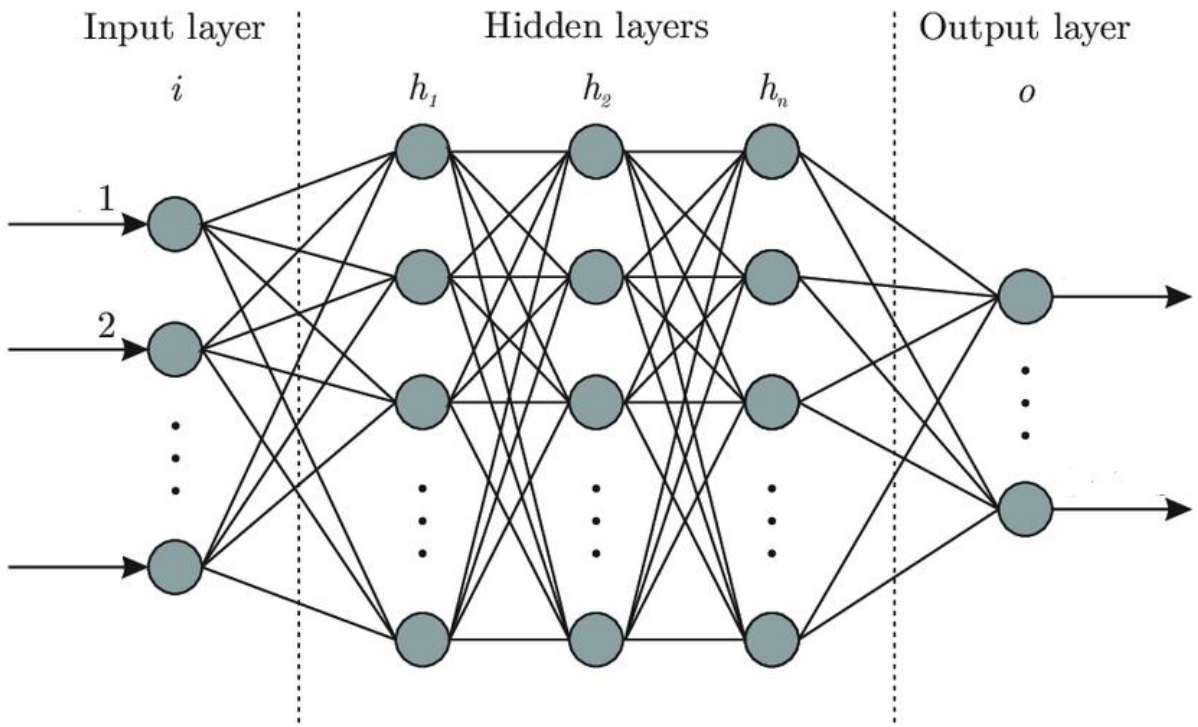
If it's in this blue circle
➔ means that we can't predict if next hour the bitcoin will increase or decrease.

VII. NEURAL NETWORK



Neural networks mimic the activity of the human brain in the fields of AI, machine learning, and deep learning, allowing computer programs to recognize patterns and solve simple issues.

A node layer, which includes an input layer, one or more hidden layers, and an output layer, makes up artificial neural networks. Each node, or artificial neuron, is connected to others and has a weight and threshold that go along with it. Any node whose output exceeds the defined threshold value is activated and begins providing data to the network's uppermost layer. Otherwise, no data is sent to the network's next tier.



In our case the input layer contains 10 features which are RSIF, RSIS, ROCF, ROCS, and so on. In the hidden layers H1 contains 512 neurons The H2 contains 512 as well and H3 contains 256 neurons. in the output layer we have 2 outputs which are Class 1 and Class 2.

I made a TensorFlow neural network code with all 10 features, load the dataset and divide my outputs into two classes: in Class 1 if the bitcoin increases the y will take (1,0) and in the second class if the bitcoin decreases the y will take (0,1). My data contains 2390 candles the train will take 1900 and the others will be for test. Then i defined the karas model, compiled it and fit it on the dataset.

```
19/19 [==============================] - 0s 8ms/step - loss: 0.3675 - accuracy: 0.8300
Epoch 143/150
19/19 [==============================] - 0s 8ms/step - loss: 0.3779 - accuracy: 0.8184
Epoch 144/150
19/19 [==============================] - 0s 8ms/step - loss: 0.3568 - accuracy: 0.8295
Epoch 145/150
19/19 [==============================] - 0s 8ms/step - loss: 0.3724 - accuracy: 0.8258
Epoch 146/150
19/19 [==============================] - 0s 10ms/step - loss: 0.3710 - accuracy: 0.8253
Epoch 147/150
19/19 [==============================] - 0s 9ms/step - loss: 0.3659 - accuracy: 0.8337
Epoch 148/150
19/19 [==============================] - 0s 8ms/step - loss: 0.3515 - accuracy: 0.8326
Epoch 149/150
19/19 [==============================] - 0s 9ms/step - loss: 0.3750 - accuracy: 0.8147
Epoch 150/150
19/19 [==============================] - 0s 8ms/step - loss: 0.3649 - accuracy: 0.8221
13/13 [==============================] - 0s 3ms/step - loss: 0.5099 - accuracy: 0.7450
Accuracy: 74.50
```

While evaluating the karas model, the accuracy was equal to 74.50 ➔ 0.8221 was the accuracy of train and 0.7450 is the accuracy of test.

- ⇨ I have also calculated the accuracy of the knn model and it was equal to 0.6725.
- ⇨ Then we can conclude that the neural network model is much more better than using the knn model.

VIII. CONCLUSION

When there isn't enough fundamental data to go on, we may rely on Technical Indicators to help us trade Bitcoin. We may incorporate new features (technical indicators) to the dataset instead of the typical price and volume history if we wanted to add more variables to a trading algorithm or a machine learning forecasting model. When using a machine learning model to estimate future prices, this strategy may be useful.

REFERENCES

By: IBM Cloud Education. (n.d.). *What are neural networks?* IBM. Retrieved June 29, 2022, from https://www.ibm.com/cloud/learn/neural-networks

FAUZI, M. A. P. A. I. M. A. N. (n.d.). *Bitcoin and cryptocurrency: Challenges, opportunities and future works*. The Journal of Asian Finance, Economics and Business. Retrieved June 29, 2022, from https://www.koreascience.or.kr/article/JAKO202026061031775.page

How to create a correlation matrix using pandas. Data to Fish. (2020, March 27). Retrieved June 29, 2022, from https://datatofish.com/correlation-matrix-pandas/?fbclid=IwAR0nXfBRxZUOn46cXGvFRYrUDyrh6aN2bEpBF2aQjvmoQBTMEy1d-ynxmOw

randerson112358. (2021, January 6). *Build a bitcoin price prediction program using machine learning and python*. Medium. Retrieved June 29, 2022, from https://randerson112358.medium.com/build-a-bitcoin-price-prediction-program-using-machine-learning-and-python-89f3dc6cb3b1

Santos, M. (2021, October 19). *Technical indicators on bitcoin using Python*. Medium. Retrieved June 29, 2022, from https://towardsdatascience.com/technical-indicators-on-bitcoin-using-python-c392b4a33810

Sean. (2021, March 1). *Momentum indicator (MOM) explained for Beginners*. Warrior Trading. Retrieved June 29, 2022, from https://www.warriortrading.com/momentum-indicator/

West, Z. (2022, June 16). *Calculating the RSI in python: 3 ways to predict market status & price movement*. αlphαrithms. Retrieved June 29, 2022, from https://www.alpharithms.com/relative-strength-index-rsi-in-python-470209/