Predicting the price of BITCOIN Using Machine Learning

CONTENTS:

INTRODUCTION

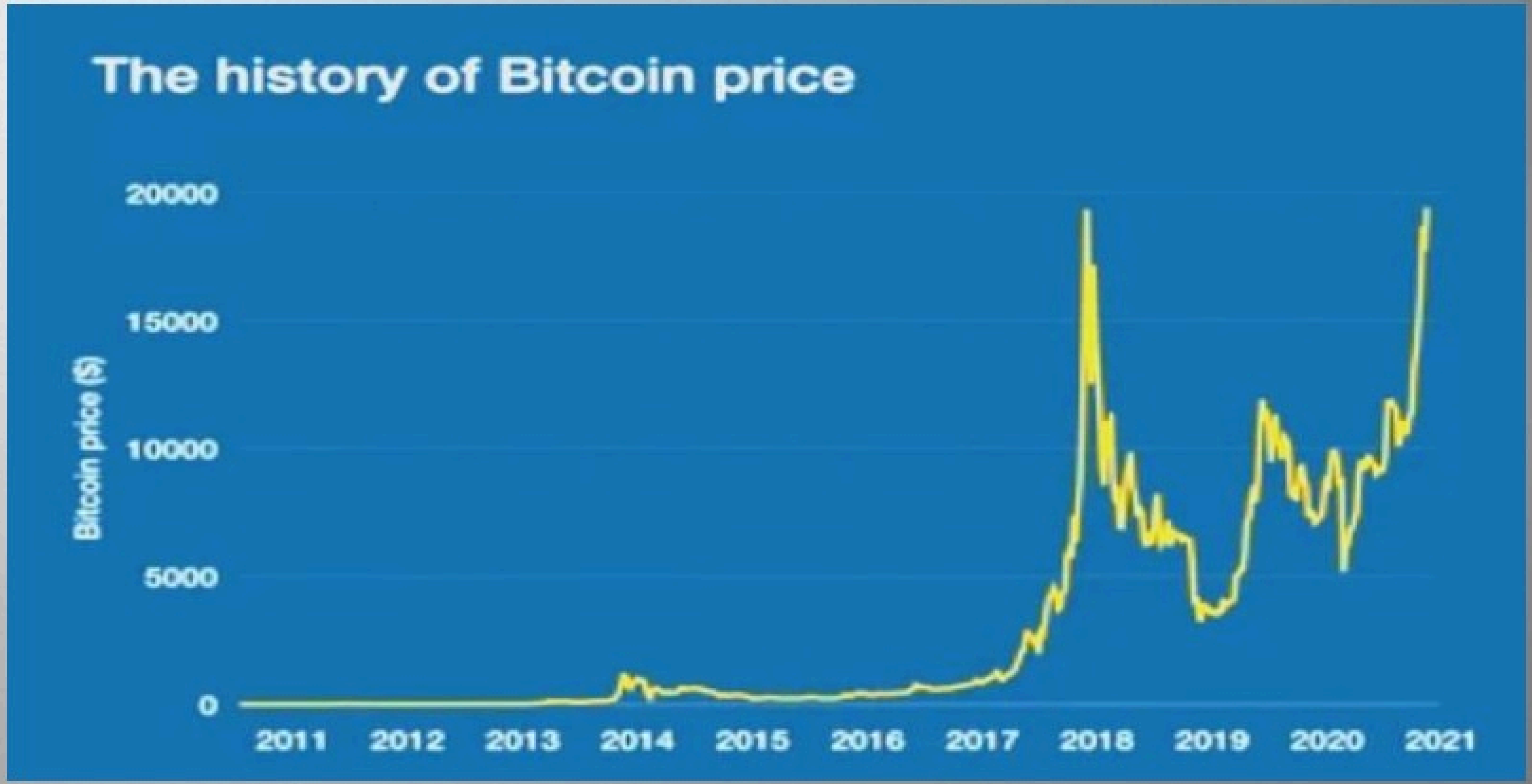
WHAT IS BITCOIN???

Bitcoin is a type of cryptocurrency. There are no physical bitcoins, only balances kept on a public ledger that everyone has transparent access to. All bitcoin transactions are verified by a massive amount of computing power. Bitcoins are not issued or backed by any banks or governments. Bitcoin is known as “BTC”.

- Cryptocurrencies, such as Bitcoin, are one of the most controversial and complex technological innovations in today's financial system.

- This study aims to forecast the movements of Bitcoin prices at a high degree of accuracy. To this aim, four different Machine Learning (ML) algorithms are applied, namely, the Support Vector Machines (*SVM*), the Artificial Neural Network (*ANN*), the Naïve Bayes (*NB)* and the Random Forest (*RF*) besides the logistic regression (LR) as a benchmark model.

- In order to test these algorithms, besides existing continuous dataset, discrete dataset was also created and used. For the evaluations of algorithm performances, the *F* statistic, accuracy statistic, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Root Absolute Error (RAE) metrics were used.



ABSTRACT

Crypto currencies, such as Bit coin, are one of the most controversial and complex technological innovations in today's financial system. This study aims to forecast the movements of Bit coin prices at a high degree of accuracy. In order to test these algorithms, besides existing continuous dataset, discrete dataset was also created and used. For the evaluations of algorithm performances, the *F* statistic, accuracy statistic, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Root Absolute Error (RAE) metrics were used. The test was used to compare the performances of the ANN, NB and RF with the performance . Empirical findings reveal that, while the *RF* has the highest forecasting performance in the continuous dataset, the *NB* has the lowest. On the other hand, while the *ANN* has the highest and the *NB* the lowest performance in the discrete dataset. Furthermore, the discrete dataset improves the overall forecasting performance in all algorithms (models) estimated.

LITERATURE SURVEY :

EXISTING SYSTEM:

- Linear regression model in linear regression is a linear approach to modeling the relationship between a dependent variable and independent variables. The case of linear variable is called simple linear regression. In this paper I am using the linear regression model for relationship between a dependent variable and one or more independent variables.

- K-Nearest Neighbor K-means creates k groups from a set of objects so that the members of a group are more similar and based on this data is clustered as normal, stressed or highly stressed.

DISADVANTAGES:

- Using regression to make predictions doesn’t necessarily involve predicting the future.
- Psychic predictions are things that just pop into mind and are not often verified against reality.
- Unsurprisingly, predictions in the regression context are more rigorous(difficult). We need to collect data for relevant variables, formulate a model, and evaluate how well the model fits the data.

PROPOSED SYSTEM:

- It is a kind of directed learning calculation that is for the most part utilized for order issues. Shockingly, it works for both clear cut and consistent ward factors. In this calculation we split the populace into at least two homogeneous sets. This is done dependent on most huge properties/autonomous factors to make as particular gatherings as could reasonably be BITCOIN PRICE PREDICTION USING MACHINE LEARNING.

- KNN (k- Nearest Neighbors) It very well may be utilized for both order and relapse issues. Be that as it may, it is all the more generally utilized in characterization issues in the business. K nearest neighbors is a straight forward calculation that stores every single accessible case and arranges new cases by a lion's share vote of its k neighbors.

- The ARIMA forecast was created by splitting the data into 5 periods and then predicting 30 days into the future. The data was differenced before being fifit with several ARIMA models. The best fifit was found by auto.arima from the R forecast package.
- Long short-term memory networks are an extension of recurrent neural networks, which basically extend the memory. Therefore it is well suited to learn from important experiences that have very long time lags in between.

Model Results

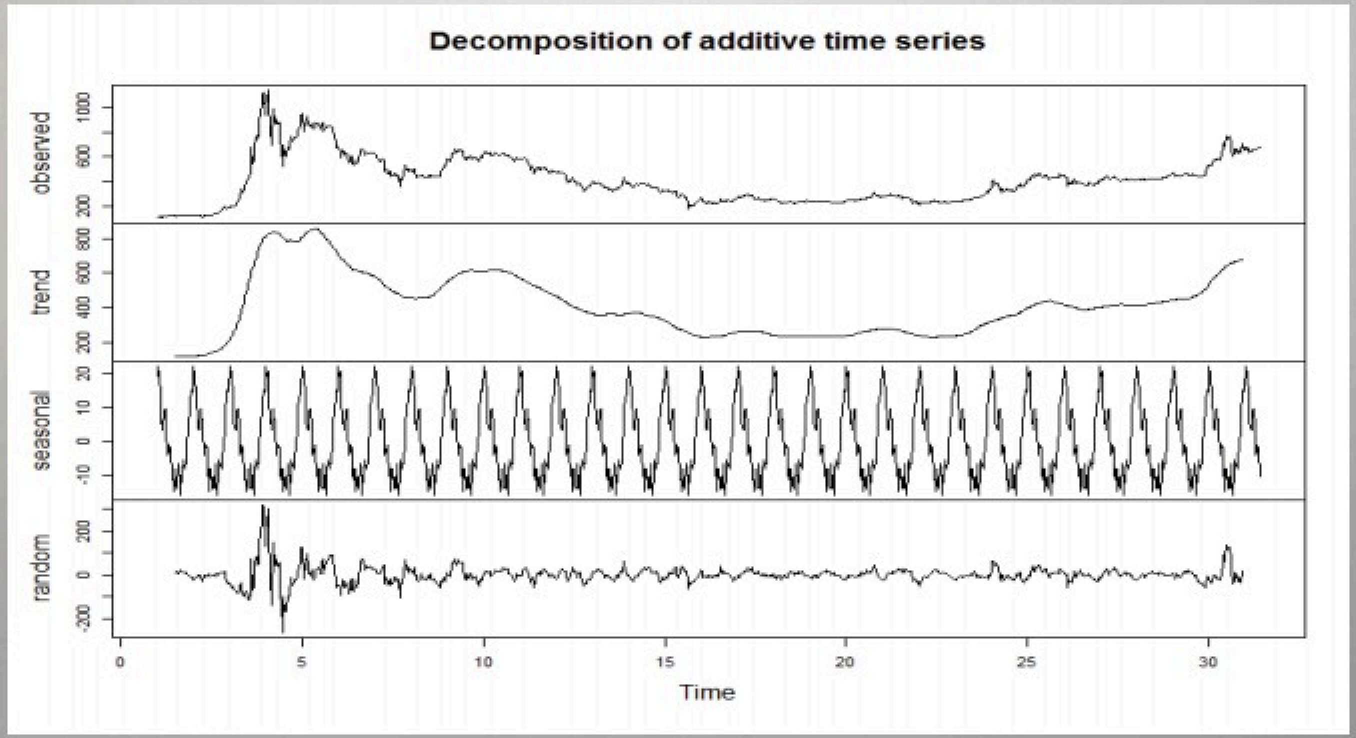
Temporal

Model	Length	Sensitivity	Specifificity	Precision	Accuracy	RMSE
LSTM	100	37%	61.30%	35.50%	52.78%	6.87%
RNN	20	40.40%	56.65%	39.08%	50.25%	5.45%
ARIMA	170	14.7%	100%	100%	50.05%	53.74%

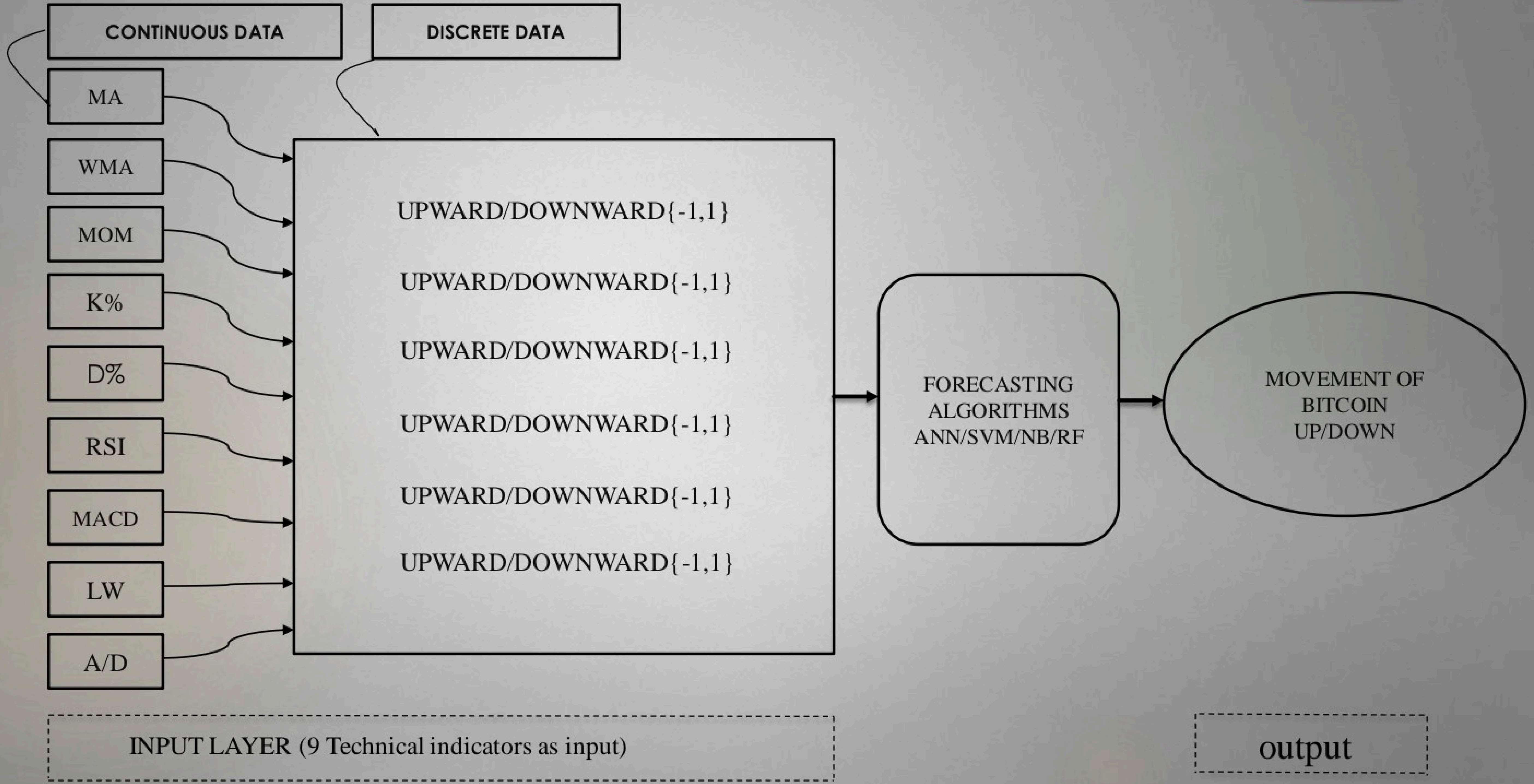
ADVANTAGES:

- Advantages of Proposed System LSTMs provide us with a large range of parameters such as learning rates, and input and output biases. Hence, no need for fine adjustments.

- The complexity to update each weight is reduced to O(1) with LSTMs, similar to that of Back Propagation Through Time (BPTT), which is an advantage.



BLOCK DIAGRAM:

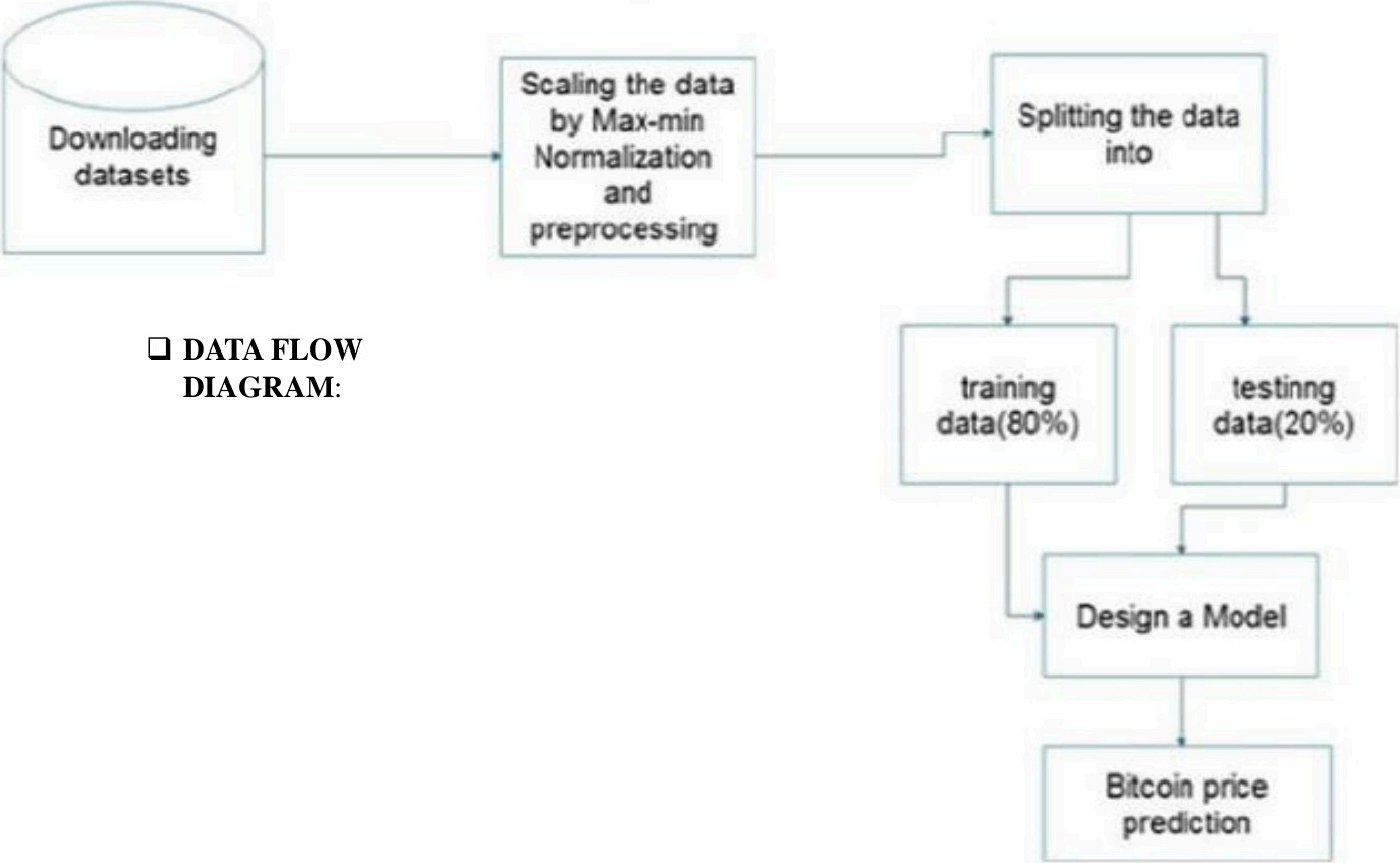


ABBREVIATIONS:

MA - Moving Average
WMA - Weighted Moving Average
MOM - Momentum
RSI - Relative Strength Index
MACD - Moving Average Convergence Divergence

USD — United States Dollar
RNN — Recurrent Neural Network
ARIMA — Autoregressive Integrated Moving Average
LSTM — Long Short-Term Memory

- ❑ **DATA FLOW DIAGRAM**:



METHODOLOGY:

1. CRISP data mining methodology:

✓ Feature engineering is the art of extracting useful patterns from data to make it easier for machine learning models to Perform their predictions. It can be considered one of the most important parts of the data mining process in order to Achieve good results in prediction tasks.

2. Deep Learning Models:

✓ Appropriate design of deep learning models interms of net-work parameters is imperative to their success. The three main Options available when choosing how to select parameters for deep learning models are random search, grid search and Heuristic search methods such as genetic algorithms.

3. Random Forest:

✓ Random Forest belongs to the supervised learning technique.
✓ It can be used for both Classification and Regression problems in ML.
✓ It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

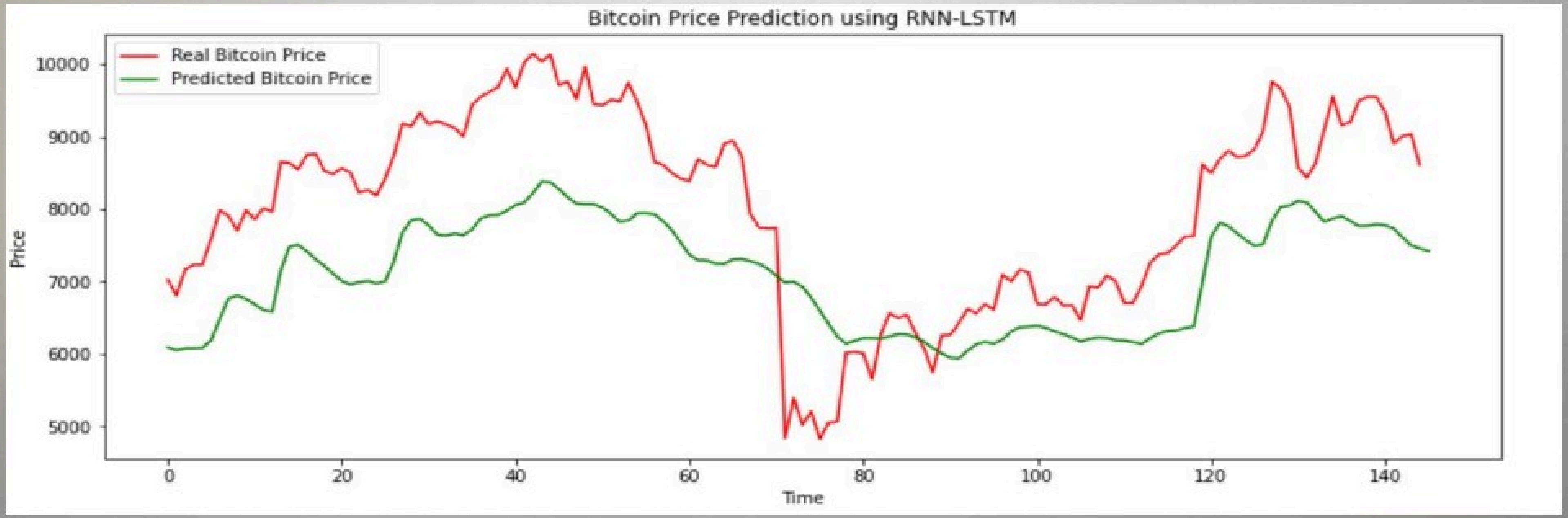
4. Long Short-Term Memory (LSTM):

- ✓ Long short-term memory networks are an extension of recurrent neural networks, which basically extend the memory. Therefore it is well suited to learn from important experiences that have very long time lags in between.

- ✓ LSTMs enable RNNs to remember inputs over a long period of time. This is because LSTMs contain information in a memory, much like the memory of a computer. The LSTM can read, write and delete information from its memory.
- ✓ LSTM you have three gates: input, forget and output gate.

5. Artificial Neural Networks:

- ✓ (ANN) are part of supervised machine learning where we will be having input as well as corresponding output present in our dataset. Our whole aim is to figure out a way of mapping this input to the respective output. ANN can be used for solving both regression and classification problems.

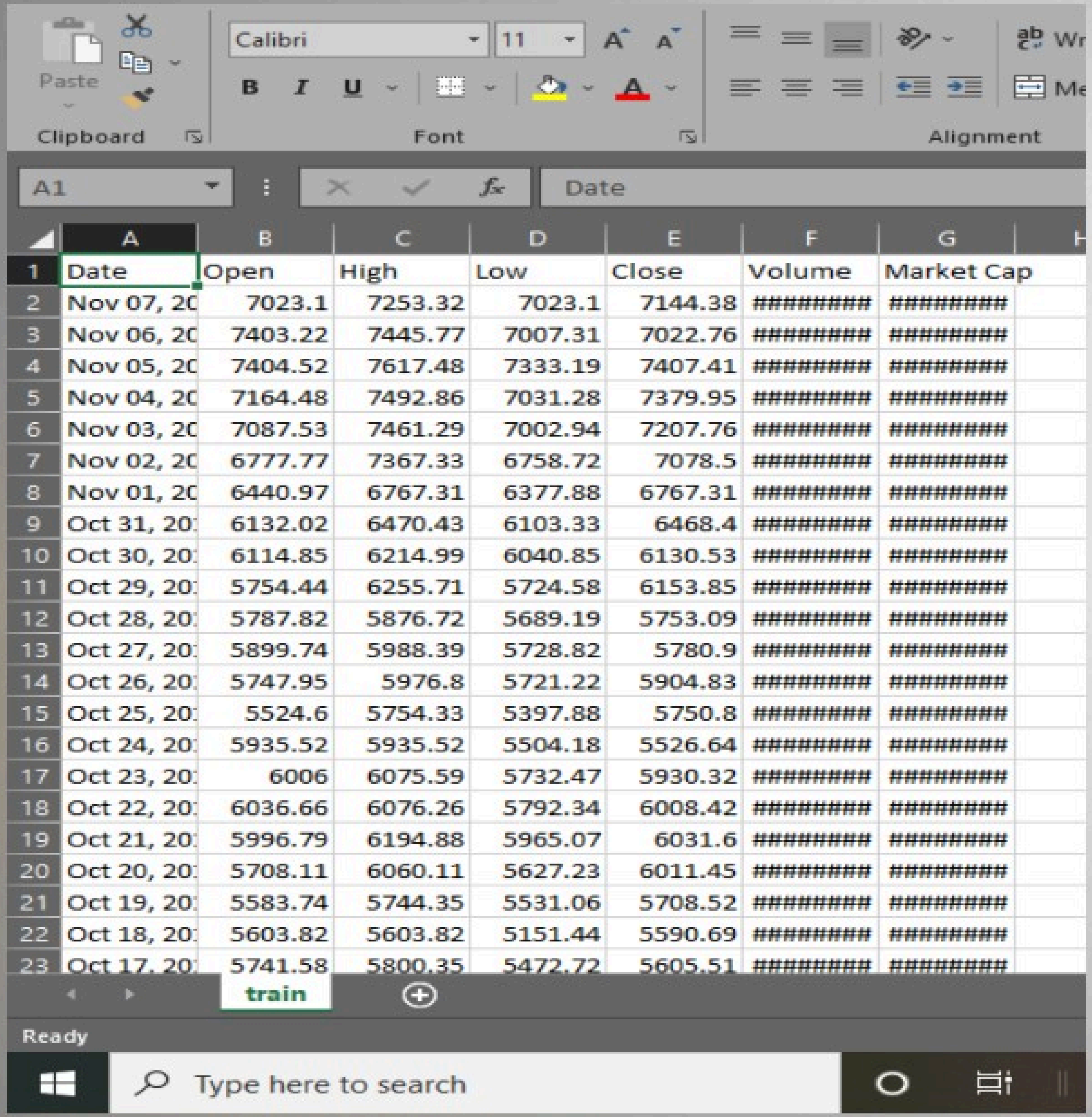


6. RNN

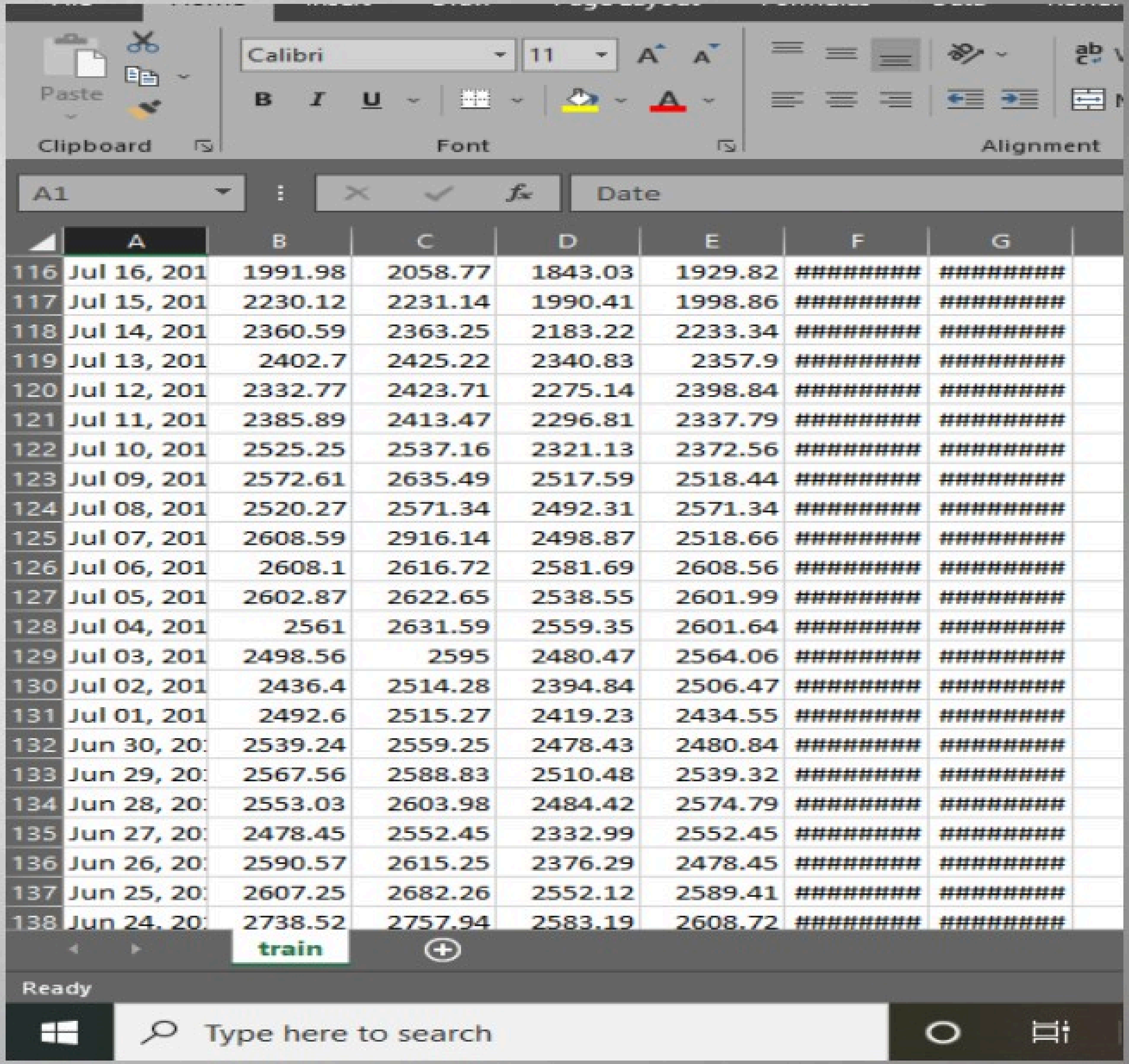
✓ The first parameter to consider was the temporal length window. As suggested by supporting literature [34] these type of networks may struggle to learn long term dependencies using gradient based optimisation. An autocorrelation function (ACF) was run for the closing price time series to assess the relationship between the current closing price and previous or future closing prices

BITCOIN PRICE:

FORMAT OF DATA:



	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Volume	Market Cap
2	Nov 07, 20	7023.1	7253.32	7023.1	7144.38	#########	#########
3	Nov 06, 20	7403.22	7445.77	7007.31	7022.76	#########	#########
4	Nov 05, 20	7404.52	7617.48	7333.19	7407.41	#########	#########
5	Nov 04, 20	7164.48	7492.86	7031.28	7379.95	#########	#########
6	Nov 03, 20	7087.53	7461.29	7002.94	7207.76	#########	#########
7	Nov 02, 20	6777.77	7367.33	6758.72	7078.5	#########	#########
8	Nov 01, 20	6440.97	6767.31	6377.88	6767.31	#########	#########
9	Oct 31, 20	6132.02	6470.43	6103.33	6468.4	#########	#########
10	Oct 30, 20	6114.85	6214.99	6040.85	6130.53	#########	#########
11	Oct 29, 20	5754.44	6255.71	5724.58	6153.85	#########	#########
12	Oct 28, 20	5787.82	5876.72	5689.19	5753.09	#########	#########
13	Oct 27, 20	5899.74	5988.39	5728.82	5780.9	#########	#########
14	Oct 26, 20	5747.95	5976.8	5721.22	5904.83	#########	#########
15	Oct 25, 20	5524.6	5754.33	5397.88	5750.8	#########	#########
16	Oct 24, 20	5935.52	5935.52	5504.18	5526.64	#########	#########
17	Oct 23, 20	6006	6075.59	5732.47	5930.32	#########	#########
18	Oct 22, 20	6036.66	6076.26	5792.34	6008.42	#########	#########
19	Oct 21, 20	5996.79	6194.88	5965.07	6031.6	#########	#########
20	Oct 20, 20	5708.11	6060.11	5627.23	6011.45	#########	#########
21	Oct 19, 20	5583.74	5744.35	5531.06	5708.52	#########	#########
22	Oct 18, 20	5603.82	5603.82	5151.44	5590.69	#########	#########
23	Oct 17, 20	5741.58	5800.35	5472.72	5605.51	#########	#########



	A	B	C	D	E	F	G
116	Jul 16, 201	1991.98	2058.77	1843.03	1929.82	#########	#########
117	Jul 15, 201	2230.12	2231.14	1990.41	1998.86	#########	#########
118	Jul 14, 201	2360.59	2363.25	2183.22	2233.34	#########	#########
119	Jul 13, 201	2402.7	2425.22	2340.83	2357.9	#########	#########
120	Jul 12, 201	2332.77	2423.71	2275.14	2398.84	#########	#########
121	Jul 11, 201	2385.89	2413.47	2296.81	2337.79	#########	#########
122	Jul 10, 201	2525.25	2537.16	2321.13	2372.56	#########	#########
123	Jul 09, 201	2572.61	2635.49	2517.59	2518.44	#########	#########
124	Jul 08, 201	2520.27	2571.34	2492.31	2571.34	#########	#########
125	Jul 07, 201	2608.59	2916.14	2498.87	2518.66	#########	#########
126	Jul 06, 201	2608.1	2616.72	2581.69	2608.56	#########	#########
127	Jul 05, 201	2602.87	2622.65	2538.55	2601.99	#########	#########
128	Jul 04, 201	2561	2631.59	2559.35	2601.64	#########	#########
129	Jul 03, 201	2498.56	2595	2480.47	2564.06	#########	#########
130	Jul 02, 201	2436.4	2514.28	2394.84	2506.47	#########	#########
131	Jul 01, 201	2492.6	2515.27	2419.23	2434.55	#########	#########
132	Jun 30, 20	2539.24	2559.25	2478.43	2480.84	#########	#########
133	Jun 29, 20	2567.56	2588.83	2510.48	2539.32	#########	#########
134	Jun 28, 20	2553.03	2603.98	2484.42	2574.79	#########	#########
135	Jun 27, 20	2478.45	2552.45	2332.99	2552.45	#########	#########
136	Jun 26, 20	2590.57	2615.25	2376.29	2478.45	#########	#########
137	Jun 25, 20	2607.25	2682.26	2552.12	2589.41	#########	#########
138	Jun 24, 20	2738.52	2757.94	2583.19	2608.72	#########	#########

RESULTS:

```
Epoch 1/50
5/5 [==============================] - 10s 17ms/step - loss: 0.3927
Epoch 2/50
5/5 [==============================] - 0s 16ms/step - loss: 0.2764
Epoch 3/50
5/5 [==============================] - 0s 15ms/step - loss: 0.1519
Epoch 4/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0438
Epoch 5/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0132
Epoch 6/50
5/5 [==============================] - 0s 14ms/step - loss: 0.0234
Epoch 7/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0061
Epoch 8/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0091
Epoch 9/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0150
Epoch 10/50
```

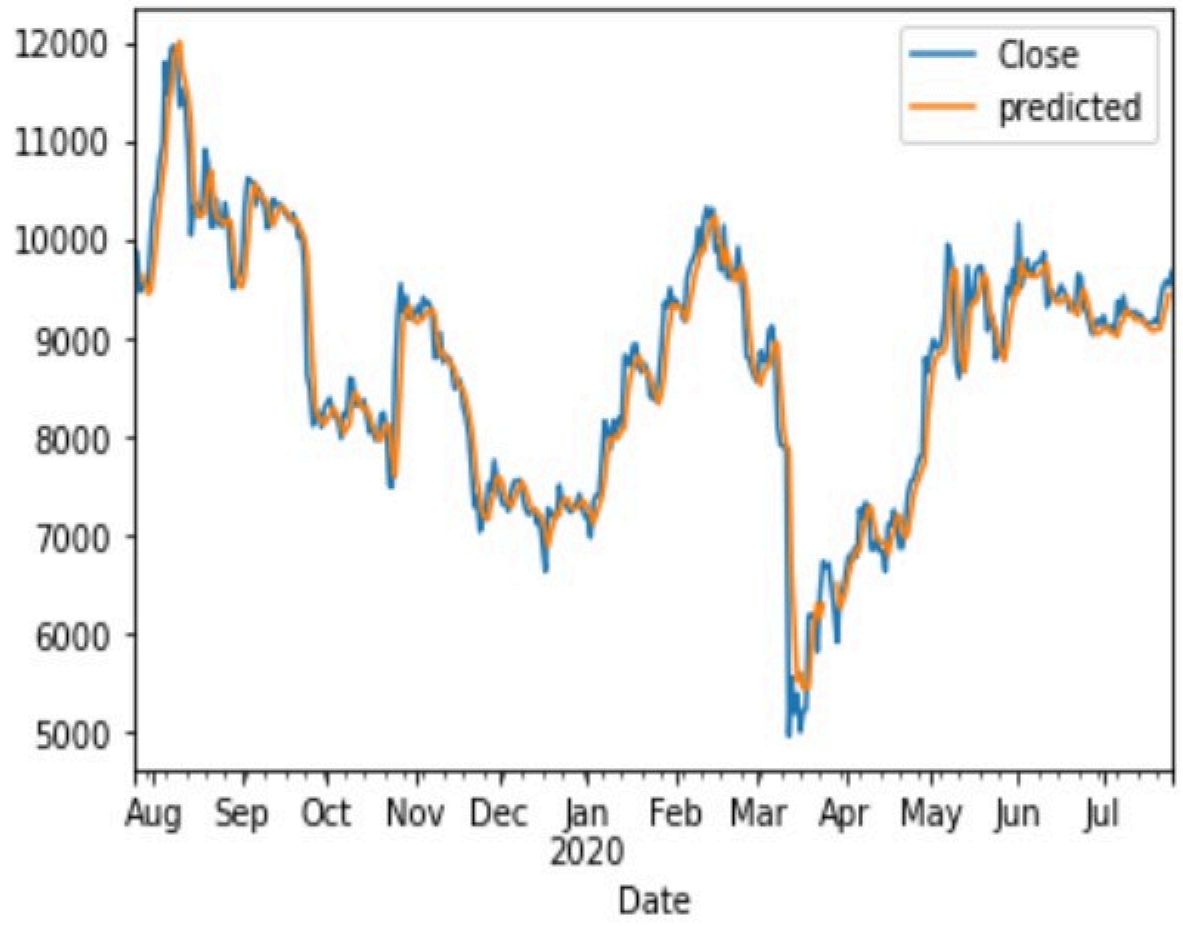
```
5/5 [==============================] - 0s 17ms/step - loss: 0.0113
Epoch 11/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0067
Epoch 12/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0054
Epoch 13/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0048
Epoch 14/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0051
Epoch 15/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0059
Epoch 16/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0054
Epoch 17/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0046
Epoch 18/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0042
Epoch 19/50
5/5 [==============================] - 0s 18ms/step - loss: 0.0041
Epoch 20/50
```

```
5/5 [==============================] - 0s 18ms/step - loss: 0.0030
Epoch 32/50
5/5 [==============================] - 0s 18ms/step - loss: 0.0030
Epoch 33/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0030
Epoch 34/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0030
Epoch 35/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0029
Epoch 36/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0029
Epoch 37/50
5/5 [==============================] - 0s 18ms/step - loss: 0.0029
Epoch 38/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0029
Epoch 39/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0029
Epoch 40/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0029
Epoch 41/50
```

```
5/5 [==============================] - 0s 18ms/step - loss: 0.0029
Epoch 42/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0028
Epoch 43/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0028
Epoch 44/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0028
Epoch 45/50
5/5 [==============================] - 0s 19ms/step - loss: 0.0028
Epoch 46/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0028
Epoch 47/50
5/5 [==============================] - 0s 15ms/step - loss: 0.0028
Epoch 48/50
5/5 [==============================] - 0s 16ms/step - loss: 0.0028
Epoch 49/50
5/5 [==============================] - 0s 18ms/step - loss: 0.0028
Epoch 50/50
5/5 [==============================] - 0s 17ms/step - loss: 0.0027
Train Score: 380.47 RMSE
```

In [7]:
```
ans.plot()
```

Out[7]: <AxesSubplot:xlabel='Date'>



CONCLUSION

Bitcoin offers an efficient means of transferring money over the internet and is controlled by a decentralized network with a transparent set of rules, thus presenting an alternative to central bank-controlled fiat money. There has been a lot of talk about how to price Bitcoin, and we set out here to explore what the cryptocurrency's price might look like in the event it achieves further widespread adoption. However, it is useful to back up a step. Bitcoin and other digital currencies have been touted as alternatives to fiat money.

REFERRED:

[1] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2018

[2] M. Briere, K. Oosterlinck, and A. Szafarz, “Virtual currency, tangible ` return: Portfolio diversification with bitcoins,” Tangible Return: Portfolio Diversification with Bitcoins 2019.

[3] I. Kaastra and M. Boyd, “Designing a neural network for forecasting financial and economic time series,” Neurocomputing, 2018

[4] H. White, “Economic prediction using neural networks: The case of ibm daily stock returns,” in Neural Networks, IEEE International Conference ,2019

[5] C. Chatfield and M. Yar, “Holt-winters forecasting: some practical issues,” The Statistician, pp. 2019


Thank
you