Cryptocurrency Price Predictor

Submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Technology
(Computer Science)

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of:
Mr. S. Rakesh Kumar

Name of Students:
Arjun Agarwal
19SCSE1010096
Karan Kumar
19SCSE1140028
Group ID: BT3309

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING, GALGOTIAS UNIVERSITY, GREATERNOIDA, INDIA

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"Cryptocurrency Price Predictor"** in partial fulfilment of the requirements for the award of the Bachelor of Technology submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of 2021 -22, under the supervision of Asst. Prof. S. Rakesh Kumar, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Arjun Agarwal 19SCSE1010096
Karan Kumar 19SCSE1140028

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Asst. Prof. S. Rakesh Kumar

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Arjun Agarwal (19SCSE1010096) and Karan Kumar (19SCSE1140028) has been held on__________________a Nd his/her work is recommended for the award of B.Tech.

Signature of Examiner(s)

Signature of Project Coordinator **Signature of Dean**

Date:

ACKNOWLEDGEMENT

We express our sincere gratitude and thanks to our university i.e., Galgotias University for providing us the opportunity to work on **“Cryptocurrency Price Predictor”.** We take this opportunity to acknowledge all the people who have helped us whole heartedly in every stage of this project.

No project is created by an individual. Many people have helped us to create this project and each of their contribution has been valuable. We respect and thank Mr. S. Rakesh Kumar, Assistant Professor, School of Computing Science and Engineering, Galgotias University, for giving us an opportunity to do the project work and providing us all support and guidance which made us complete the project on time. We are extremely grateful to him for providing such a nice support and guidance.

We are thankful and fortunate enough to get constant encouragement and support from our teachers, friends and parents.

Contents

Abstract

Cryptocurrency is a new sort of asset that has emerged as a result of the advancement of financial technology and it has created a big opportunity for researches. Cryptocurrency price forecasting is difficult due to price volatility and dynamism. Around the world, there are hundreds of cryptocurrencies that are used. This paper proposes three types of recurrent neural network (RNN) algorithms used to predict the prices of three types of cryptocurrencies, namely Bitcoin (BTC), Litecoin (LTC), and Ethereum (ETH). The models show excellent predictions depending on the mean absolute percentage error (MAPE). Results obtained from these models show that the gated recurrent unit (GRU) performed better in prediction for all types of cryptocurrencies than the long short-term memory (LSTM) and bidirectional LSTM (bi-LSTM) models. Therefore, it can be considered the best algorithm. GRU presents the most accurate prediction for LTC with MAPE percentages of 0.2454%, 0.8267%, and 0.2116% for BTC, ETH, and LTC, respectively. The bi-LSTM algorithm presents the lowest prediction result compared with the other two algorithms as the MAPE percentages are: 5.990%, 6.85%, and 2.332% for BTC, ETH, and LTC, respectively. Overall, the prediction models in this paper represent accurate results close to the actual prices of cryptocurrencies. The importance of having these models is that they can have significant economic ramifications by helping investors and traders to pinpoint cryptocurrency sales and purchasing. As a plan for future work, a recommendation is made to investigate other factors that might affect the prices of cryptocurrency market such as social media, tweets, and trading volume.

Introduction

Cryptocurrency is a virtual or digital currency used in financial systems. It is secured by cryptography that makes it impossible to be counterfeited or double-spent. Furthermore, it is not issued from a central authority or central banks, and it is decentralized virtual currencies that can be converted via cryptographic procedures and this make it distinguishable from traditional currencies. The other feature is that it is created by technology called blockchain, which is an extremely complex, and aims to storing data that makes it difficult or impossible to alter, hack, or defraud the system. Bitcoin has begun to carve out a niche for itself, which may either help cryptocurrencies to gain widespread acceptance or be the major cause of their demise. Cryptocurrencies are still in their infancy, and it is difficult to predict whether they will ever be widely used in global markets or not. The most prominent cryptocurrency, Bitcoin, was established in 2009 and for more than two years was the sole Blockchain-based cryptocurrency. Today, however, there are over 5000 cryptocurrencies and 5.8 million active users in the cryptocurrency industry. Because of its intrinsic nature of mixing encryption technology with monetary units, Bitcoin has recently gotten a lot of attention in the disciplines of economics, cryptography, and computer science. Blockchain (BC), the technology that underpins the Bitcoin cryptocurrency system, is widely seen as critical in providing the backbone for assuring greater security and privacy in a variety of other fields, including the Internet of Things (IoT) ecosystem. It is mainly a digital ledger of transactions that is distributed across the entire network of computer systems on the blockchain. The blockchain consists of two fundamental components; the first one is a transaction, and the second is a block. The transaction represents the action triggered by the participant, and the block is a data collection that records the transaction and additional details such as the correct sequence and creation timestamp. Blockchain have a signaling system (BloSS) of multi-domain, blockchain-based, cooperative DDoS defense system in which each autonomous system (AS) joins the defensive alliance. Reference reveals that the effects of networks on competition in the nascent cryptocurrency market over a period of time regarding exchange rates among cryptocurrencies depends on two aspects: competition among different currencies and competition among exchanges. There are hundreds of cryptocurrencies, but Bitcoin is the most popular one as it is a stubborn competitor and did not emerge out of the cryptocurrency competition track. As a result, it has become the dominant cryptocurrency. The authors of describe the competition between cryptocurrency as "healthy competition" and suggests that new technology and security innovation. The authors of reveal that Bitcoin and national currencies show volatility shock transmission, while economic policy uncertainty has little effect. The authors of investigate the interaction between big data and cryptocurrency. One of the most appealing marketplaces for financial speculation is the cryptocurrency market, which means that deceptive activities have flourished via social media. Many people have reaped a lot of profits through speculation in the digital markets, but every investment process suffers from many hidden risks and some investors, particularly those with a high-risk tolerance, are interested in investing in cryptocurrency. Therefore, market analysts and speculators rely on prediction. With variations in predictive power per cryptocurrency, machine learning and artificial intelligence algorithms are moderately appealing. Low-volatility cryptocurrencies are more predictable than high-volatility ones. There is evidence that the usefulness of different information sets varies between machine learning algorithms, implying that prediction is likely to be much more complicated when a set of machine learning algorithms is used. Despite the widespread use of cryptocurrencies for various types of purchases and transactions around the world, there is no consistent opinion on the definition of cryptocurrency or its legal status. Furthermore, the aforementioned situation exacerbates challenges in criminal investigations of cryptocurrency-based money laundering. As a result, law enforcement organizations are having difficulty pinpointing criminals' identities and proving that they have committed a crime. Focusing on Bitcoin pricing is similar to stock pricing: none of the risk variables that explain stock price movements apply to cryptocurrencies. Furthermore, traditional macroeconomic variables such as currency rates, commodity prices, and macroeconomic factors that affect other assets have little to no impact on most cryptocurrencies. As a result of the cryptocurrency market's surge in 2017, various governments across the world have begun to move toward standardizing and overseeing digital money. Because of the security of blockchain technology and their economic environment, people have become more confident in using Bitcoin. Although the blockchain provides a high security ecosystem, the research area surrounding the legality of cryptocurrency cannot be isolated from the people who utilize cryptocurrencies for illicit purposes. The legality of cryptocurrencies has been the subject of numerous debates. The authors of discuss the perspectives and the nature of cryptocurrencies in terms of monetary features, legal considerations, economic considerations, and Sharia considerations. Based on the perspective and characteristics of traditional currency, cryptocurrency does not satisfy the characteristics of a currency from an economic standpoint. There are hundreds of cryptocurrencies in digital markets, but Bitcoin is the most popular and is affected and interacted with by external influences such as the news, social media, and small cryptocurrencies that have a limited market share, which are often not taken into account from investors and traders. Due to the strong relationships between cryptocurrencies, the smaller ones have become a source of shocks that can positively or negatively affect other cryptocurrencies. The authors of reveal that gold as an independent currency can be used as a good hedging instrument to decrease the risk related to unexpected movement in the cryptocurrency market. Cryptocurrency prices are difficult to forecast due to price volatility and dynamism. Around the world there are hundreds of cryptocurrencies that clients use. In this paper, we focus on three of the most popular ones. As a result, the paper aims to achieve the following by using deep leaning algorithms, which can discover hidden patterns from data, integrate them, and create far more efficient predictions:

* Presenting a comprehensive study of the various existing schemes to predict the prices of BTC, ETH, and LTC cryptocurrencies.
* Using AI algorithms such as LSTM, bi-LSTM, and GRU to accurately predict the prices of cryptocurrencies.
* Utilizing long short-term memory (LSTM), a deep learning algorithm, and Fbprophet, which is an auto machine learning algorithm, for prediction.
* Evaluating the proposed hybrid models using evaluation matrices such as RMSE and MAPE for Bitcoin, Ethereum, and Litecoin.

The main idea behind these models is to achieve a reliable prediction model that investors can rely on based on historical cryptocurrency prices. Moreover, the paper aim is to answer the following research questions: ‘How can machine learning algorithms help investors and decision makers to predict cryptocurrency prices?’ and ‘What is the best model for predicting future cryptocurrency prices?’

Literature Survey

Machine learning (ML) is a type of artificial intelligence that can predict the future based on past data. ML-based models have various advantages over other forecasting models as prior research has shown that it not only delivers a result that is nearly or exactly the same as the actual result, but it also improves the accuracy of the result. Examples of machine learning include neural networks (NN), support vector machines (SVM), and deep learning. The authors of demonstrate that incorporating cryptocurrency into a portfolio improves its effectiveness in two ways. The first is to reduce the standard deviation, and the second is to provide investors with more allocation options. The best cryptocurrency allocation was reported to be in the range from 5% to 20%, depending on the risk tolerance of the investor. The results show that the ML ensemble technique can be used to anticipate Bitcoin values. The decision-making process needs to make the appropriate decision at the right time, reducing the risks associated with the investment process. In, a hybrid cryptocurrency prediction system based on LSTM and GRU is presented, focusing on two cryptocurrencies, Litecoin and Monero. The authors of use minute-sampled Bitcoin returns over 3 h periods to aggregate RV data. A variety of machine learning methods, including ANN (MLP, GRU, and LSTM), SVM, and ridge regression, were used to predict future values based on past samples, which are compared to the heterogeneous auto-regressive realized volatility (HARRV) model with optimized lag parameters. The findings show that the suggested system correctly predicts prices with high accuracy, indicating that the method may be used to forecast prices for a variety of cryptocurrencies. The authors of employ the traditional support vector machine and linear regression methods to forecast Bitcoin values. This research takes into account a time series prediction made up of everyday Bitcoin closing prices for the creation of Bitcoin prediction models. The authors utilize powerful artificial intelligence frameworks, including a fully linked artificial neural network (ANN) and a long short-term memory (LSTM) recurrent neural network, and they discovered that ANN relies more on long-term history, whereas LSTM relies more on short-term dynamics, implying that LSTM is more efficient at extracting meaningful information from historical memory than ANN. The study in on Bitcoin daily price prediction with high-dimensional data reveals that logistic regression and linear discriminant analysis achieve an accuracy of 66%. On the other hand, surpassing (a sophisticated machine learning algorithm) outperforms the benchmark results for daily price prediction, with statistical techniques and machine learning algorithms having the greatest accuracies of 66% and 65.3%, respectively. The study in examines the use of neural networks (NN), support vector machines (SVM), and random forest (RF). The findings demonstrate that machine learning and sentiment analysis may be used to anticipate cryptocurrency markets (with Twitter data alone being able to predict specific coins) and that NN outperforms the other models. In, the LSTM model is used to predict and find methods for forecasting Bitcoin on the stock market through Yahoo Finance that may predict a result of more than 12,600 USD in the days after the prediction. Due to the importance of the development of a robust and reliable method for predicting cryptocurrency prices, researchers have focused on more innovative models. In, both linear and non-linear time-series components of the stock dataset were used for forecasting using the hybrid model. In the non-linear time series forecast, CNN and Seq2Seq LSTMs were successfully coupled for dynamic modeling of short- and long-term dependent patterns. The study in focused on social factors, which are increasingly being utilized for online transactions throughout the world, by using a multi-linear regression model and that analyzes two big capital market cryptocurrencies, BTC and LTC. The authors of found that the R2 scores were 44% for LTC and 59% for BTC. Ref. used two different LSTM models (a standard LSTM model and an LSTM with an AR model). This study presented a forecasting framework, using an LSTM model to forecast Bitcoin daily prices. The study in found that the model with AR was better than LSTM with an RMSE of 247.33. Researchers compared three different models (ARIMA, LSTM and GRU) for predicting BTC's price. The experimental outcomes showed that ARIMA achieved the best performance with a MAPE of 2.76% and RMSE of 302.53. The study in presented two types of prediction models constructed using Bayesian optimized RNN and LSTM to predict the price of BTC. The study revealed that LSTM showed better performance and achieved an accuracy of 52% and RMSE of 8%. The investment process mainly depends on the historical price of a cryptocurrency. One of the most important strategies that the investor depends on is building Markov chains. This strategy consists of multiple decision trees that are used to identify the cryptocurrency that is estimated to provide a greater return when sold and then comparing the estimation with the actual figure. Due to the importance of prediction in the investment process that many people depend on to earn revenue, this paper focuses on three models that can predict future cryptocurrency prices using machine learning algorithms and artificial intelligence approaches to achieve accurate prediction models with the aim of helping investors.

Materials And Methods

To achieve the aims of this paper, we trained three distinct models for three different forms of cryptocurrency price prediction using historical cryptocurrency prices. Then, in order to evaluate the suggested schemes' performances, we compare the accuracy of our proposed model to that of current models by following five stages:

1. Collecting historical cryptocurrency data;
2. Data exploration and visualization;
3. Training three types of models;
4. Testing the models;
5. Extracting and comparing the results.

In this section, we present and compare three types of algorithms—long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (Bi-LSTM)—to predict the price of three types of cryptocurrencies based on historical data—Bitcoin (BTC), Litecoin (LTC) and Ethereum (ETH). Figure 1 shows the methodology of processing the dataset. It starts with data collection, then the data visualization process is used to illustrate and explore the data's behavior and distribution and the relationship between the cryptocurrencies.
Next, the models are trained with 80% of the collected dataset. The training dataset is from 22 January 2018 until 22 October 2020 and the testing dataset (20% of the data) is form 22 October 2020 until 30 June 2021. Then, after training the models we tested them. Then, we extracted and compared the results and selected the best model depending on the daily closing price.

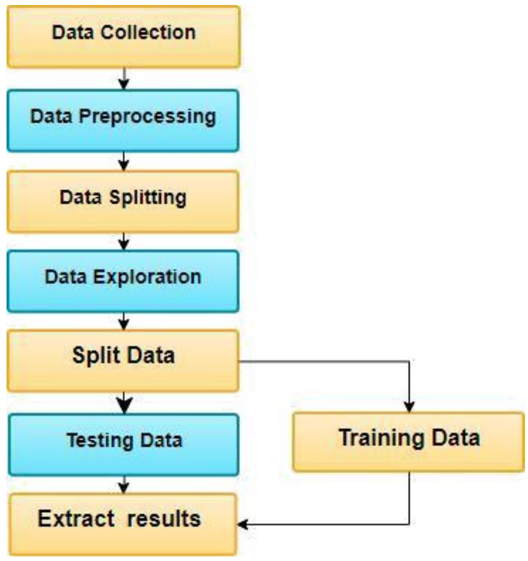


Figure 1. Methodology of processing data and model selection

Figures 2–4 illustrate the training and testing dataset for every targeted cryptocurrency. We can see that the price for each currency roughly increased and decreased together along the time-series.

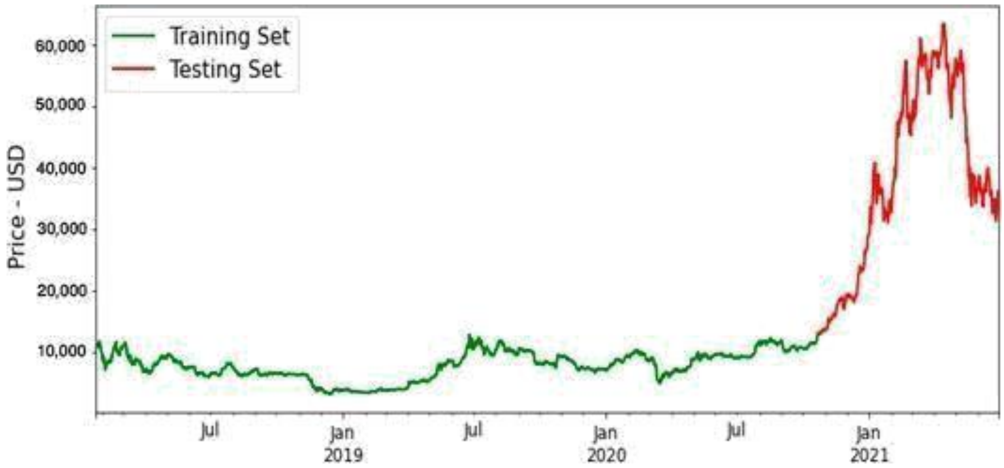


Figure 2. Training and testing dataset for BTC.

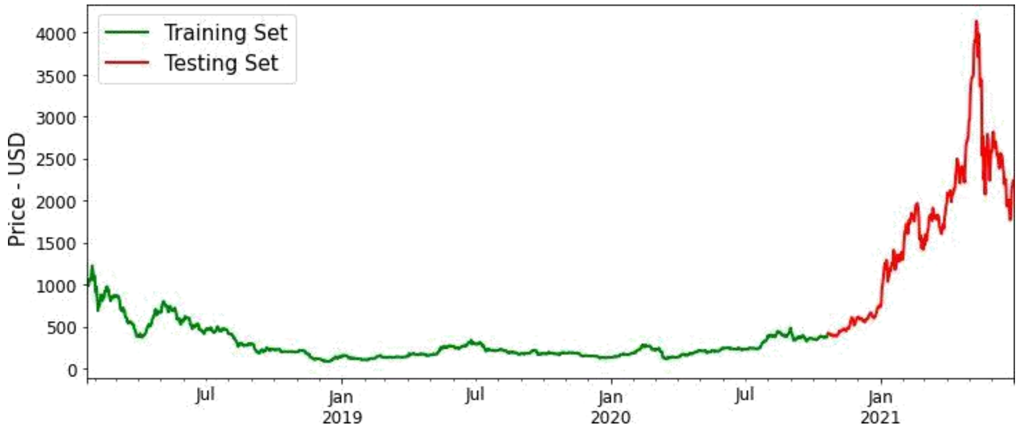


Figure 3. Training and testing dataset for ETH.

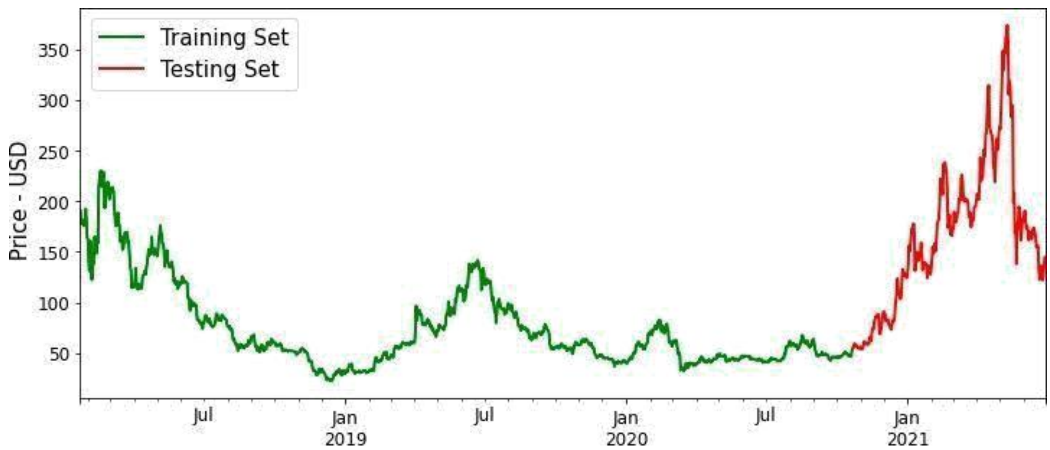


Figure 4. Training and testing dataset for LTC.

Figure 4 shows the LTC closing price within the targeted collected dataset. Its how's the closing price increased gradually until the end of 2020, when the price increased suddenly, it reaches a high of 63,381 USD in a top of peak of time series. Figure 3 shows the ETH closing price within the targeted collected dataset. It demonstrates that the closing price increased gradually until the end of 2020, then the price increased suddenly, reaching a high of 4140 USD. Figure 4 shows the LTC closing price within the targeted collected dataset. It illustrates that the closing price increased gradually until the end of 2020, then the price increased suddenly, reaching a high of 373.64 USD.

Training Data

Training data is an extremely large dataset that is used to teach a machine learning model. For supervised ML models, the training data is labelled. The data used to train unsupervised ML models is not labelled. The idea of using training data in machine learning programs is a simple concept, but it is also very foundational to the way that these technologies work. The training data is an initial set of data used to help a program understand how to apply technologies like neural networks to learn and produce sophisticated results. It may be complemented by subsequent sets of data called validation and testing sets. Training data is also known as a training set, training dataset or learning set.

Test Data

Test Data in Software Testing is the input given to a software program during test execution. It represents data that affects or affected by software execution while testing. Test data is used for both positive testing to verify that functions produce expected results for given inputs and for negative testing to test software ability to handle unusual, exceptional or unexpected inputs.

Poorly designed testing data may not test all possible test scenarios which will hamper the quality of the software.

Random Forest Regressor

Every decision tree has high variance, but when we combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that particular sample data and hence the output doesn't depend on one decision tree but multiple decision trees. In the case of a classification problem, the final output is taken by using the majority voting classifier. In the case of a regression problem, the final output is the mean of all the outputs. This part is Aggregation.

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as **bagging**. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.

Cryptocurrency

Cryptocurrency is a digital payment system that doesn't rely on banks to verify transactions. It's a peer-to-peer system that can enable anyone anywhere to send and receive payments. Instead of being physical money carried around and exchanged in the real world, cryptocurrency payments exist purely as digital entries to an online database describing specific transactions. When you transfer cryptocurrency funds, the transactions are recorded in a public ledger. Cryptocurrency is stored in digital wallets.

Cryptocurrency received its name because it uses encryption to verify transactions. This means advanced coding is involved in storing and transmitting cryptocurrency data between wallets and to public ledgers. The aim of encryption is to provide security and safety.

The first cryptocurrency was Bitcoin, which was founded in 2009 and remains the best known today. Much of the interest in cryptocurrencies is to trade for profit, with speculators at times driving prices skyward.

How Does Cryptocurrency Work

Cryptocurrencies run on a distributed public ledger called blockchain, a record of all transactions updated and held by currency holders. Units of cryptocurrency are created through a process called mining, which involves using computer power to solve complicated mathematical problems that generate coins. Users can also buy the currencies from brokers, then store and spend them using cryptographic wallets. If you own cryptocurrency, you don't own anything tangible. What you own is a key that allows you to move a record or a unit of measure from one person to another without a trusted third party. Although Bitcoin has been around since 2009, cryptocurrencies and applications of blockchain technology are still emerging in financial terms, and more uses are expected in the future. Transactions including bonds, stocks, and other financial assets could eventually be traded using the technology.

Fbprophet

Time series forecasting is usually a complex task because the structure of already univariate data often contains many unobserved factors. Standard models such as ARIMA, or filters, e.g., Kalman Filter are complex models that often need tweaking which requires a rigorous understanding of the underlying theory. Practitioners with good domain knowledge but little statistics know-how want to make use of machine learning and forecasting methodologies to inform their business decisions. So, a number of software packages and libraries attempt to bridge this gap by offering automated solutions. The aim of this article is to investigate a promising library by Facebook, called Prophet. The promise of Prophet is to generate forecasts automatically at scale.

Results

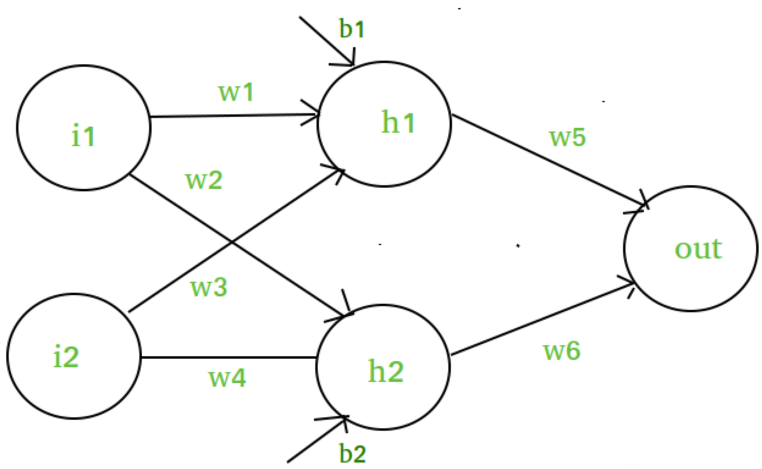
Unfortunately, the hopes that an engine can do reliable forecasts at scale without any interaction were too high, even under laboratory circumstances. After some tweaking, however, the library performs well and recognizes all the elements that were put into the test data. So, with having a little knowledge on the domain, the library produces useful results

Neural Networks

Neural networks are artificial systems that were inspired by biological neural networks. These systems learn to perform tasks by being exposed to various datasets and examples without any task-specific rules. The idea is that the system generates identifying characteristics from the data they have been passed without being programmed with a pre-programmed understanding of these datasets.

Neural networks are based on computational models for threshold logic. Threshold logic is a combination of algorithms and mathematics. Neural networks are based either on the study of the brain or on the application of neural networks to artificial intelligence. The work has led to improvements in finite automata theory.

“To put it in the coding world, a neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. Neural networks can adapt to changing input; so, the network generates the best possible result without needing to redesign the output criteria.”



CNN
(Convolution Neural Networks)

It is assumed that the reader knows the concept of Neural networks. When it comes to Machine Learning, Artificial Neural Networks perform really well. Artificial Neural Networks are used in various classification tasks like image, audio, words. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolution Neural networks. In this blog, we are going to build a basic building block for CNN. Before diving into the Convolution Neural Network, let us first revisit some concepts of Neural Network. In a regular Neural Network, there are three types of layers:

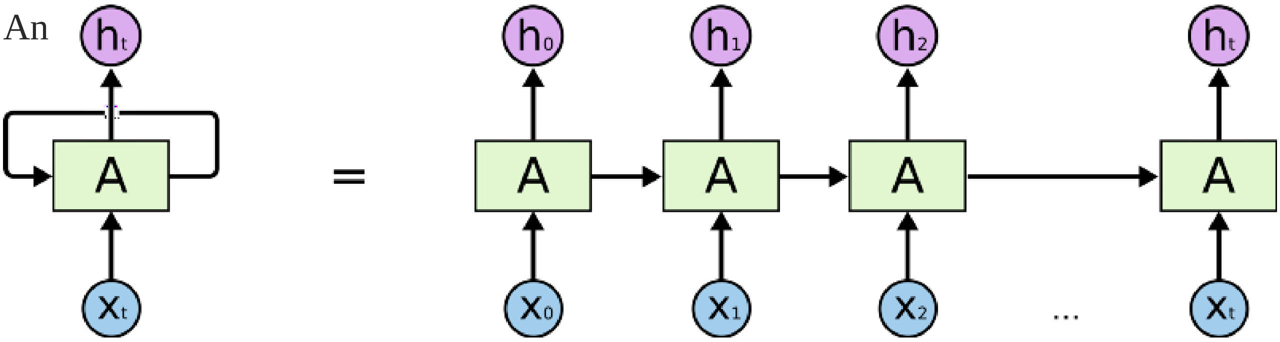
- **Input Layers:** It’s the layer in which we give input to our model. The number of neurons in this layer is equal to the total number of features in our data (number of pixels in the case of an image).
- **Hidden Layer:** The input from the Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by the addition of learnable biases followed by activation function which makes the network nonlinear.
- **Output Layer:** The output from the hidden layer is then fed into a logistic function like sigmoid or SoftMax which converts the output of each class into the probability score of each class.

The data is then fed into the model and output from each layer is obtained this step is called feedforward, we then calculate the error using an error function, some common error functions are cross-entropy, square loss error, etc. After that, we backpropagate into the model by calculating the derivatives. This step is called Backpropagation which basically is used to minimize the loss.

RNN
(Recurrent Neural Network)

In traditional neural networks, all the inputs, and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words (as the next word will depend on your previous input). Example you watch a movie and in between you stop to predict the ending, it will depend on how much you have seen it already and what context has come up yet. Similarly, RNN remembers everything. It solves this problem of traditional neural network with a hidden layer.

An

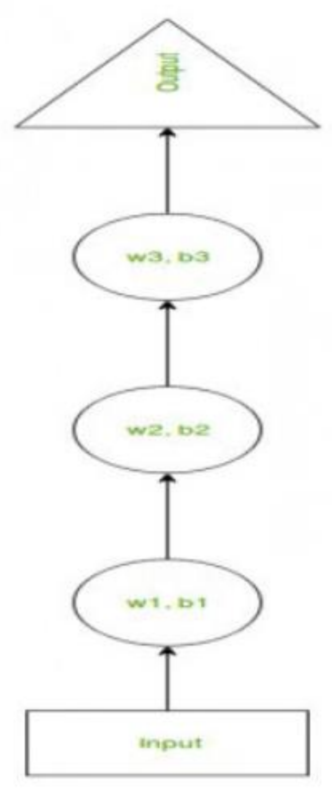


RNN remembers each and every information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short Term Memory.

Working of RNN

RNN creates the networks with loops in them, which allows it to persist the information. This loop structure allows the neural network to take the sequence of input.

- RNN converts the independent activations into dependent activations by providing the same weights and biases to all the layers, thus reducing the complexity of increasing parameters and memorizing each previous outputs by giving each output as input to the next hidden layer.
- Hence these three layers can be joined together such that the weights and bias of all the hidden layers is the same, into a single recurrent layer.

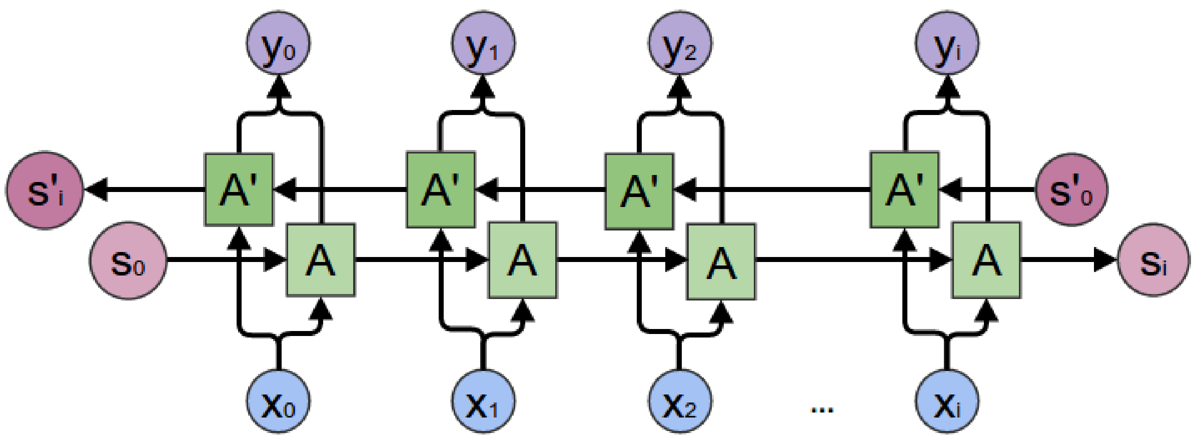


Bi-LSTM

(Bi-directional long short-term memory)

Bidirectional recurrent neural networks (RNN) are really just putting two independent RNNs together. This structure allows the networks to have both backward and forward information about the sequence at every time step

Using bidirectional will run your inputs in two ways, one from past to future and one from future to past and what differs this approach from unidirectional is that in the LSTM that runs backward you preserve information from the future and using the two hidden states combined you are able in any point in time to preserve information from both past and future.



Dataset

The analyzed dataset was collected from, an open-access website. It consists of one .csv file separated into three sheets; the first sheet for Bitcoin (BTC), the second for Litecoin (LTC), and the last sheet for Ethereum (ETH). The recorded prices in the dataset were collected on a daily basis from 1 January 2018 to 30 June 2021. In this research, we used time-series data with 1277 records. Table 1 illustrates the dataset specification of the targeted cryptocurrency and Figure 10 shows sample data from the dataset.

Table 1. Dataset specification.

Variable Name	Variable Description	Data value
Date	Date of Observation	Date
Open	Opening price on the given day	Number
High	High price on the given day	Number
Low	low price on the given day	Number
Close	close price on the given day	Number

Sample of the data from BTC dataset

	Date	Open	High	Low	Close
0	2018-01-22	11,348	11,869	10,051	10,264
1	2018-01-23	10,264	11,358	9,972	10,984
2	2018-01-24	10,986	11,474	10,497	11,208
3	2018-01-25	11,213	11,711	10,889	11,246
4	2018-01-26	11,241	11,609	10,321	10,899

Sample of the data from ETH dataset

	Date	Open	High	Low	Close
0	2018-01-22	1,052.22	1,090.99	913.24	1,000.5
1	2018-01-23	1,000.12	1,025.0	910.0	985.96
2	2018-01-24	984.1	1,067.0	957.9	1,063.77
3	2018-01-25	1,063.77	1,107.67	1,000.01	1,048.0
4	2018-01-26	1,048.0	1,079.95	982.0	1,051.07

Sample of the data from LTC dataset

	Date	Open	High	Low	Close
0	2018-01-22	190.93	195.85	165.1	180.01
1	2018-01-23	180.01	187.39	165.25	177.32
2	2018-01-24	178.01	186.13	173.0	180.99
3	2018-01-25	181.18	185.0	175.0	179.05
4	2018-01-26	179.05	182.7	165.37	175.39

Figure 5. Screenshot showing a sample of the data from the BTC, ETH, and LTC dataset.

Results

This section shows the results obtained from long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (bi-LSTM) algorithms using three types of popular cryptocurrency: BTC, ETH, and LTC. For each model, the results are illustrated in Tables 2–4. The model that gives the lowest RMSE and MAPE is considered the best model. Based on this criterion, all of the models applied to three types of currencies can be considered good models but the GRU was found to be the best of the three. The RMSE of the GRU model is the lowest. Thus, GRU is more capable of predicting long-term dependencies as compared to LSTM and bi-LSTM. This is due to the dependency on past figures. 11–19 illustrate the comparisons between the actual and the predicted results. Simulation results from those models indicate that there are few occasions where the forecast results different from actual result.

Table 2. BTC models results.

Model	**RMSE**	**MAPE**
LSTM	410.399	1.1234%
bi-LSTM	2927.006	5.990%
GRU	174.129	0.2454%

Table 3. ETH models result.

Model	**RMSE**	**MAPE**
LSTM	59.507	1.5498%
bi-LSTM	321.061	6.85%
GRU	26.59%	0.8267%

Table 4. LTC model results

Model	RMSE	MAPE
LSTM	3.069	0.8474%
bi-LSTM	4.307	2.332%
GRU	0.825	0.2116%

Figure 6. Actual and predicted price of BTC using the LSTM model.

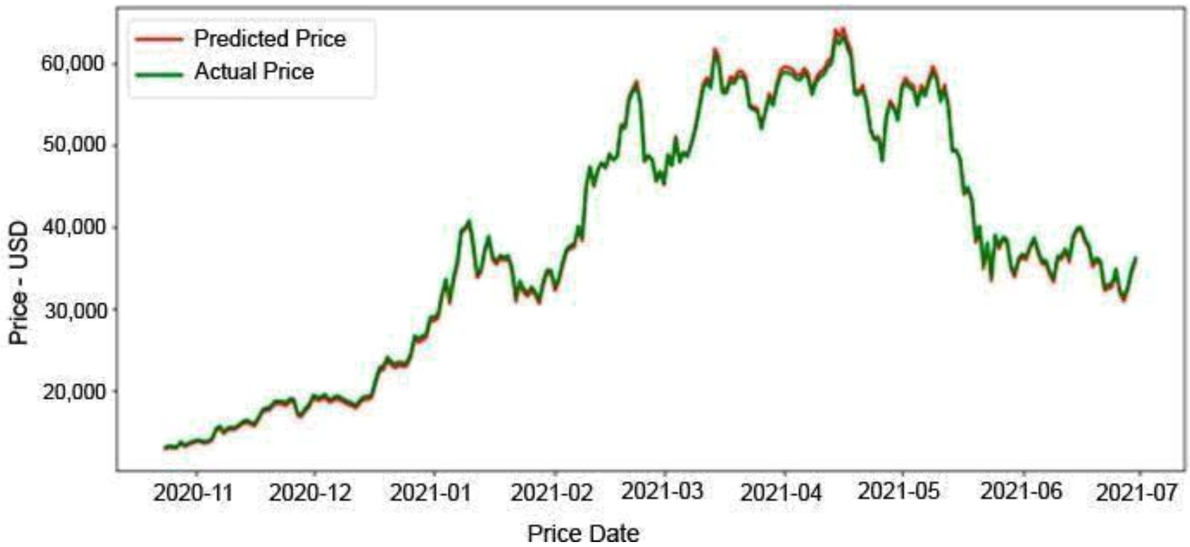


Figure 7. Actual and predicted price of BTC using the GRU model.

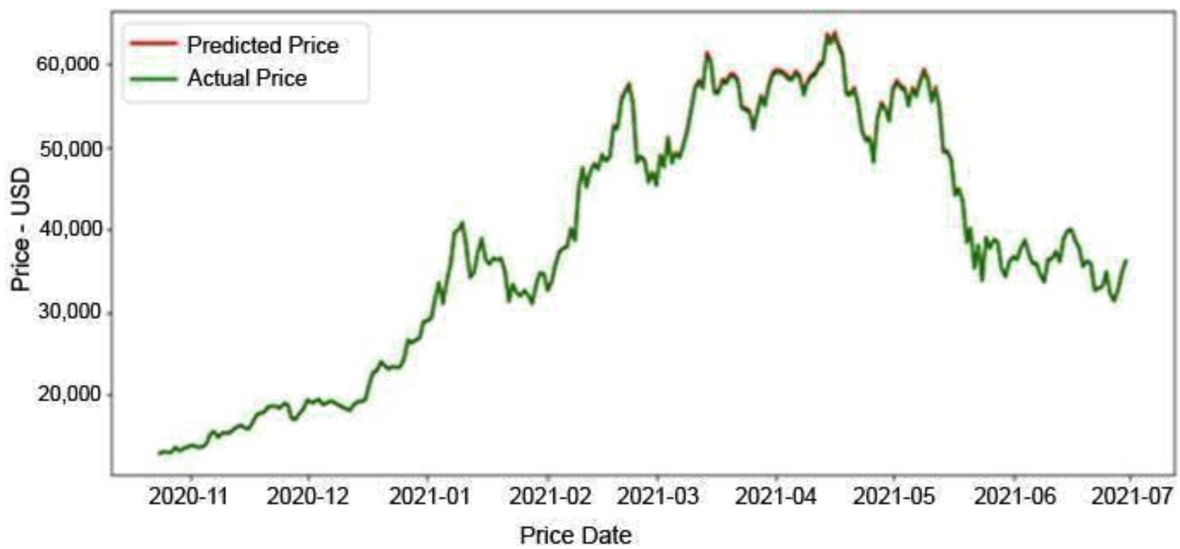


Figure 8. Actual and predicted price of BTC using the bi-LSTM model

Figure 9. Actual and predicted price of ETH using the LSTM model.

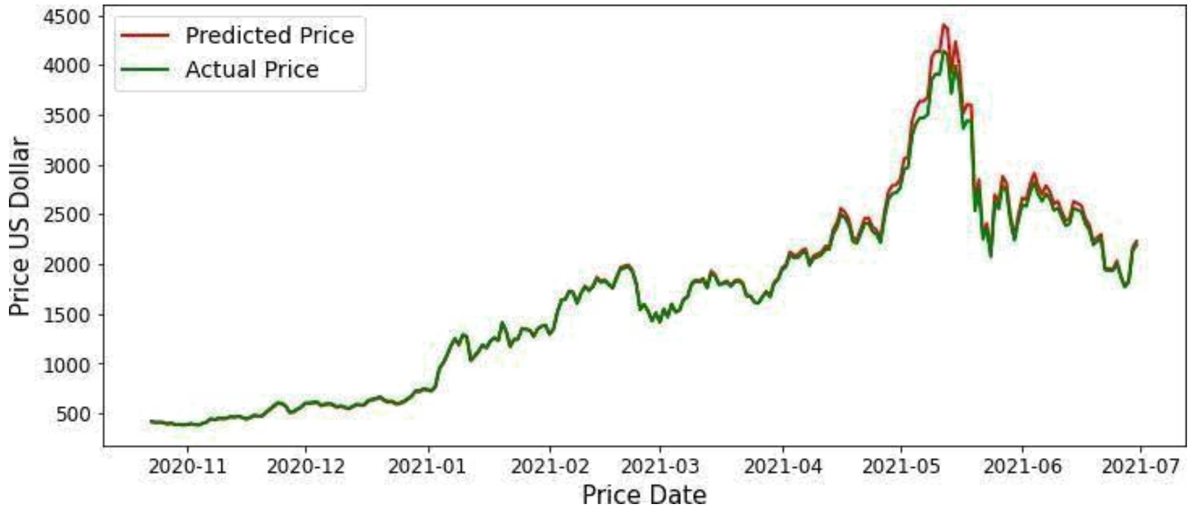
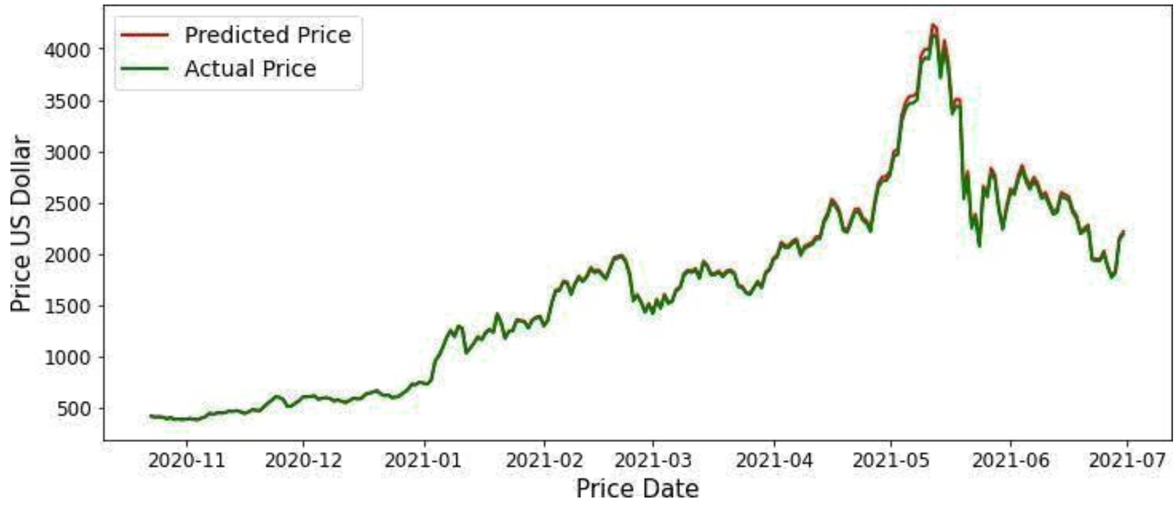


Figure 10. Actual and predicted price of ETH using the GRU model



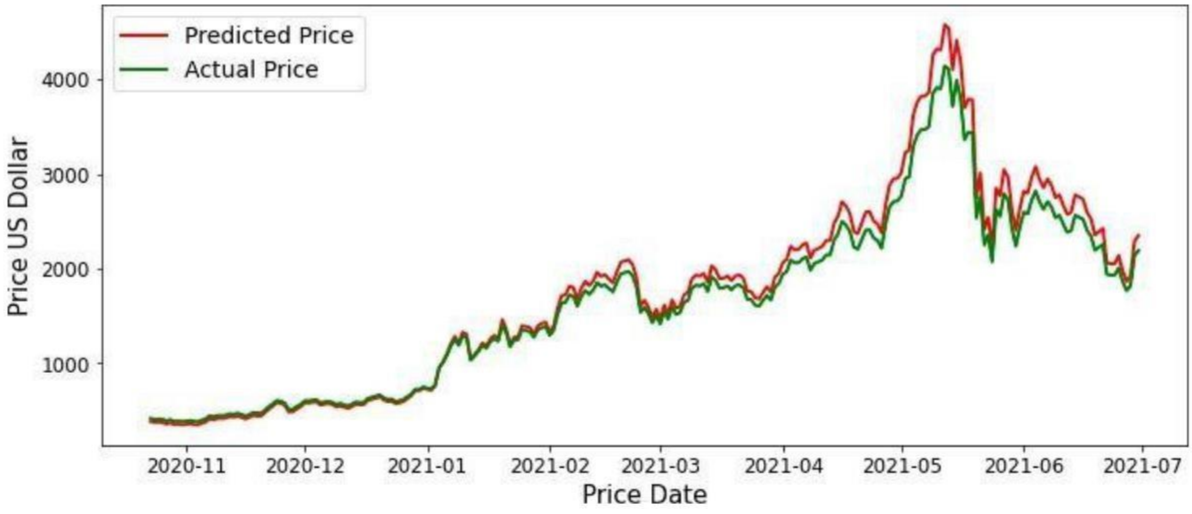


Figure 11. Actual and predicted price of ETH using the bi-LSTM model.

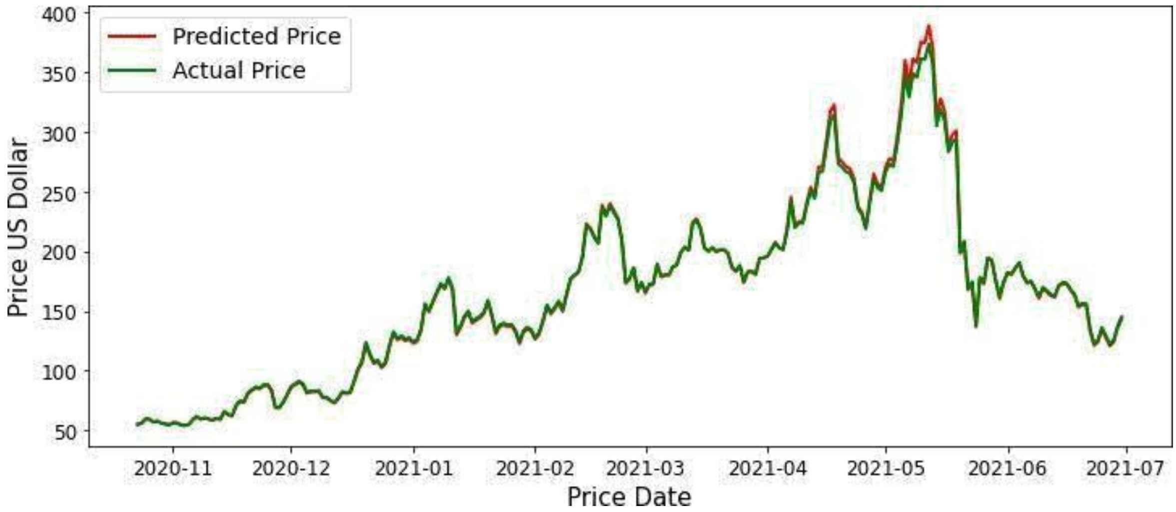


Figure 12. Actual and predicted price of LTC using the LSTM model.

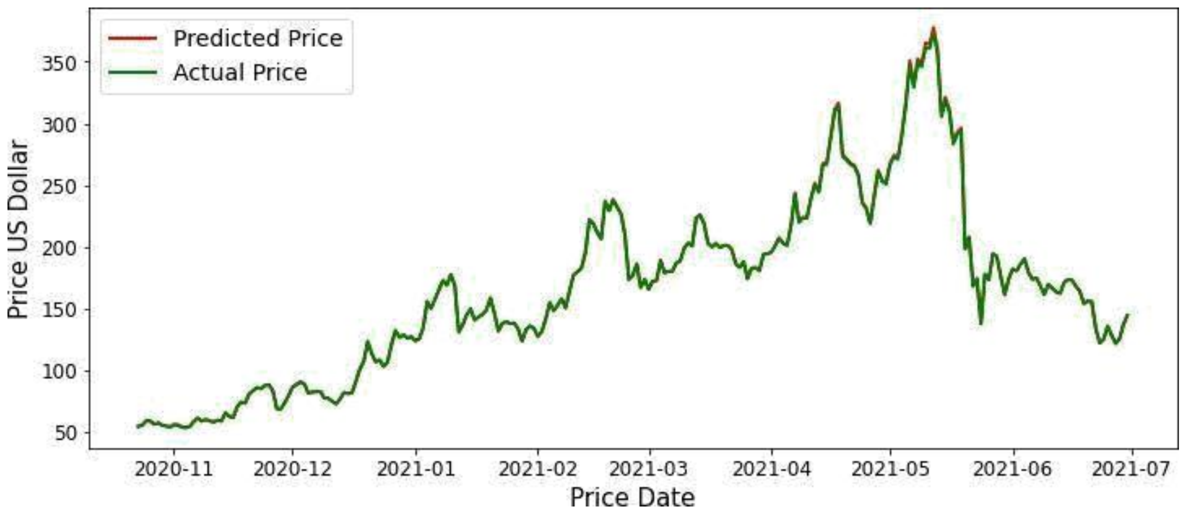


Figure 13. Actual and predicted price of LTC using the GRU model

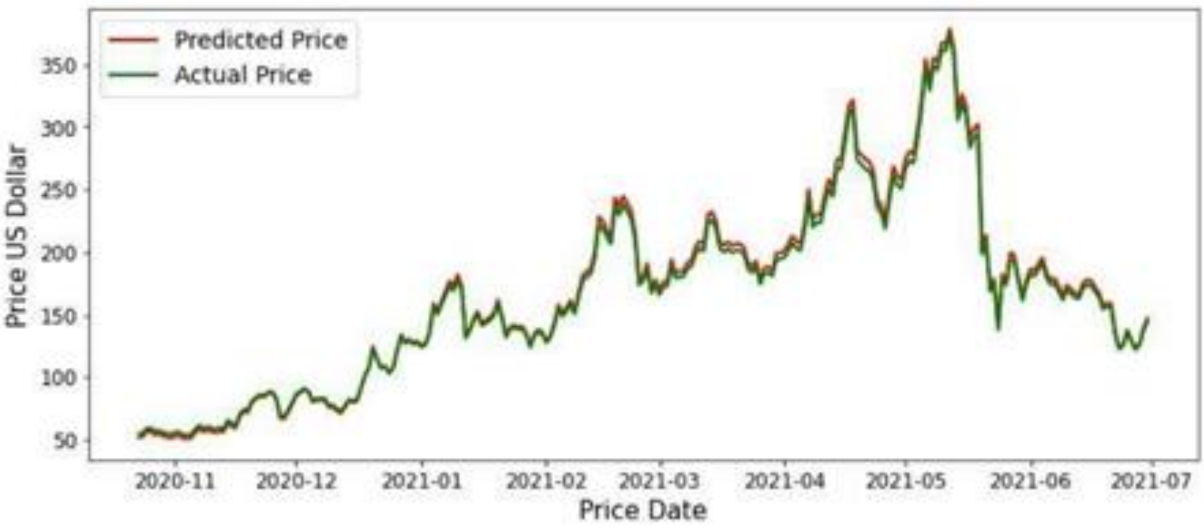


Figure 14. Actual and predicted price of LTC using the bi-LSTM model.

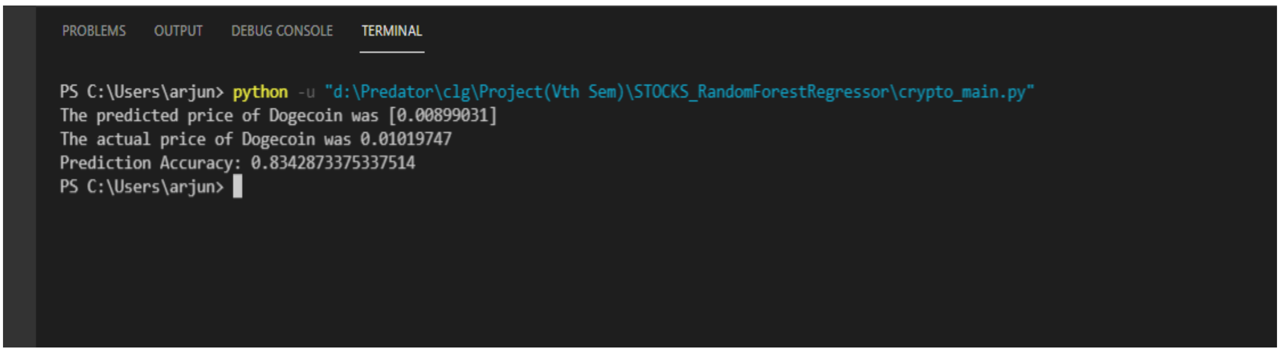
Implementation

Code:

```
#Libraries
import numpy as np
import pandas as pd
import datetime as dt
from sklearn.model_selection import train_test_split
#pdata
df= pd.read_csv("D://Predator//clg//Sem_V//Project(Vth
Sem)//STOCKS_RandomForestRegressor//doge.csv")
#set index=date
df= df.set_index(pd.DatetimeIndex(df['Date'].values))

#Get close price
df = df[['Close']]

#Variable to store number of days into the predicted future
prediction_days = 1

#Column to store predicted price
df['Prediction'] = df[['Close']].shift(-prediction_days)

#Create independent dataset
X = np.array(df.drop(['Prediction'],1))

#Remove n+1 rows of data , n=prediction prediction_days
X = X[:len(df)- prediction_days - 1]

#Creating dependent dataset 'Y'
y = np.array(df['Prediction'])

#All values of y except for last n+1 rows
y = y[:- prediction_days -1]

#Split data into training and testing dataset (80-20)
```

```
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

#ML model
from sklearn.ensemble import RandomForestRegressor
forest = RandomForestRegressor(n_estimators = 2, random_state = 587)
forest.fit(x_train, y_train)

#Get validation data
#Varible to store all except last n rows of dataset
temp_df= df[:-prediction_days]

#Variable to store independent price Values
x_val = temp_df.tail(1)['Close'][0]

prediction = forest.predict([[x_val]])

#Print the price of Dogecoin for next n days
print('The predicted price of Dogecoin was', prediction)

#Actual Values
print('The actual price of Dogecoin was', temp_df.tail(1)['Prediction'][0])

#Accuracy
print("Prediction Accuracy:",forest.score(x_test, y_test))
```

Output



Dataset Used

```
Close, High, Low, Open, Date, Volume
0.32596167,0.34537626,0.26706741,0.30372583,18-04-2021,14518479246
0.31387909,0.35443295,0.31387909,0.35443295,17-04-2021,51222207135
0.20423106,0.21321115,0.135264,0.13884566,16-04-2021,48628881409
0.18121614,0.18988116,0.1204675,0.12246324,15-04-2021,21925194741
0.12365066,0.14820844,0.0935555,0.09612397,14-04-2021,25764023464
0.09381815,0.09552192,0.07115828,0.07136712,13-04-2021,8339483884
0.07174924,0.07687516,0.06980632,0.07544494,12-04-2021,3230769945
0.07476455,0.08056264,0.06398123,0.06442087,11-04-2021,7400166193
0.06456375,0.06595793,0.06170916,0.0624922,10-04-2021,1597175691
0.06282469,0.06544822,0.06151311,0.06294961,09-04-2021,1378202689
0.06275467,0.06339491,0.05953554,0.05990077,08-04-2021,1396999897
0.05998606,0.06922612,0.05702443,0.06565343,07-04-2021,3587258446
0.06532548,0.06709487,0.05881927,0.06105184,06-04-2021,3482576404
0.06124639,0.06237953,0.05727446,0.05832825,05-04-2021,2169122204
0.05796695,0.05926601,0.05595678,0.05659677,04-04-2021,1459641978
0.05684885,0.06100261,0.056656,0.05904109,03-04-2021,1670328534
0.05816582,0.06351434,0.05772932,0.06263085,02-04-2021,2592342909
0.06269749,0.07269484,0.05393406,0.0541567,01-04-2021,7171562899
0.05382154,0.05467041,0.05159113,0.05409486,31-03-2021,1096408037
0.05415312,0.05577132,0.05368689,0.05423208,30-03-2021,1172420263
0.05412032,0.05482573,0.05322695,0.05378618,29-03-2021,935395380
0.05366048,0.05489438,0.05302035,0.05442855,28-03-2021,845528074
0.05439612,0.05546945,0.0533413,0.05391508,27-03-2021,1120517740
0.05399785,0.05417176,0.05137011,0.05147511,26-03-2021,916314140
0.0514273,0.05278387,0.04977967,0.05185251,25-03-2021,1422087889
0.05196777,0.05686561,0.0512119,0.05367787,24-03-2021,1205263147
0.05378236,0.0564532,0.05291241,0.05525609,23-03-2021,1106461281
0.05502405,0.05792486,0.05454301,0.05734335,22-03-2021,933237559
0.05820051,0.06030622,0.05735369,0.05940642,21-03-2021,993662153
0.05927771,0.0612362,0.05825638,0.05871008,20-03-2021,1204608326
0.05839433,0.05995476,0.05647184,0.05753608,19-03-2021,1151382976
0.05746123,0.05929792,0.05699875,0.05790745,18-03-2021,1121222009
0.05783967,0.05906958,0.05648292,0.05871269,17-03-2021,1102033556
0.05871913,0.05921463,0.05544777,0.05704185,16-03-2021,1687859532
0.05713665,0.05999565,0.05506701,0.05841818,15-03-2021,2019305500
0.05845138,0.06373488,0.05822253,0.06247542,14-03-2021,3102189648
0.06238338,0.06264506,0.05467378,0.05533197,13-03-2021,2290213218
0.05548929,0.05737538,0.0543784,0.05607465,12-03-2021,1593502436
```

0.05586889,0.05696565,0.05433628,0.05597245,11-03-2021,1600519636
0.05594808,0.05877163,0.05445055,0.05794486,10-03-2021,1930072213
0.05771727,0.0628582,0.05568817,0.06259073,09-03-2021,2818481688
0.0626891,0.06300107,0.05163913,0.05209469,08-03-2021,4255275379
0.05192368,0.05236554,0.05040476,0.05095989,07-03-2021,1158465912
0.05095869,0.05261121,0.04915794,0.04954042,06-03-2021,1799358982
0.04961294,0.05115163,0.04813848,0.0501206,05-03-2021,1560292850
0.05015958,0.05134181,0.04786645,0.05053988,04-03-2021,1520320325
0.05057487,0.05230228,0.05004059,0.05057289,03-03-2021,1391397434
0.05057092,0.0526653,0.04929801,0.05070368,02-03-2021,1660524281
0.05060855,0.05244589,0.0478945,0.0482923,01-03-2021,1886983460
0.04822056,0.05040364,0.04487177,0.05017584,28-02-2021,1839162728
0.05023359,0.05217653,0.04928306,0.05069824,27-02-2021,1662314243
0.05064865,0.05276839,0.04862964,0.05025379,26-02-2021,2107545525
0.05182493,0.08658661,0.05033149,0.0676742,25-02-2021,3167309723
0.05678525,0.0763299,0.0464904,0.05437796,24-02-2021,5235136033
0.04729293,0.05382878,0.03915901,0.05345923,23-02-2021,2714399930
0.05366332,0.06089581,0.04664169,0.05594688,22-02-2021,3844184995
0.0560674,0.05968982,0.05331433,0.05442133,21-02-2021,2979501668
0.05432262,0.0608289,0.0512397,0.05511103,20-02-2021,3658347274
0.05506991,0.06014898,0.05376497,0.0587721,19-02-2021,3354855193
0.05877316,0.06386411,0.04807801,0.04961023,18-02-2021,5126372295
0.04960345,0.05529813,0.04794349,0.05368956,17-02-2021,2568634442
0.05366451,0.0600262,0.05084604,0.05674478,16-02-2021,3016222561
0.05664133,0.06429976,0.04732127,0.06252206,15-02-2021,4887989360
0.06261793,0.06685109,0.0558228,0.06655764,14-02-2021,3861038976
0.06651141,0.07184294,0.06581713,0.0697764,13-02-2021,2490260048
0.06985453,0.07309939,0.06106761,0.06979741,12-02-2021,3965088014
0.0697794,0.07568091,0.06816313,0.07298642,11-02-2021,3614824612
0.07276916,0.08170613,0.06672456,0.07012964,10-02-2021,6690509505
0.070174,0.08398014,0.06381103,0.07891811,09-02-2021,6529498556
0.07884683,0.08594246,0.06376713,0.07864008,08-02-2021,11646581650
0.07817614,0.08672813,0.05336059,0.0576198,07-02-2021,14438831338
0.05757215,0.05868454,0.04455635,0.04689723,06-02-2021,5883240063
0.04674722,0.05445979,0.04304259,0.05283404,05-02-2021,5350433805
0.0528768,0.05934694,0.03564621,0.03705526,04-02-2021,12682773624
0.03689334,0.03972905,0.03084683,0.03143633,03-02-2021,2890203050
0.03139988,0.03564648,0.02894687,0.03502225,02-02-2021,2620447218
0.03503255,0.04388077,0.03278009,0.03677874,01-02-2021,5279870624
0.03690705,0.04824828,0.02659953,0.02821995,31-01-2021,8236688157
0.02835622,0.05084036,0.02196525,0.04621153,30-01-2021,7447995023
0.04622248,0.08545541,0.03041121,0.03600898,29-01-2021,21802011544
0.03593646,0.03696566,0.007458,0.00758092,28-01-2021,19081256527

0.00743915,0.00826988,0.00726364,0.00825772,27-01-2021,239330227
0.00825886,0.0084884,0.00795087,0.00836737,26-01-2021,222970028
0.00837588,0.00892022,0.00814725,0.00873975,25-01-2021,243976263
0.00874434,0.0089892,0.0084835,0.00858033,24-01-2021,256469982
0.00856993,0.00885278,0.00832264,0.0084996,23-01-2021,267712080
0.0085116,0.00880135,0.00760666,0.00825175,22-01-2021,298292627
0.00821918,0.00908386,0.00800904,0.00901319,21-01-2021,306332141
0.00905033,0.00917221,0.00850419,0.00904552,20-01-2021,328933247
0.00905283,0.00968077,0.00896399,0.00917316,19-01-2021,330851291
0.00917383,0.00933251,0.00893011,0.0091381,18-01-2021,303884548
0.00912927,0.00941223,0.0087121,0.00922891,17-01-2021,352598223
0.00923104,0.0095723,0.00901043,0.00932127,16-01-2021,361490258
0.00933659,0.00985892,0.00831604,0.00939947,15-01-2021,440841735
0.00938748,0.01011938,0.00831267,0.00867299,14-01-2021,456226639
0.00864403,0.00870476,0.00776521,0.00810076,13-01-2021,296982799
0.00809177,0.00914406,0.00773281,0.00875607,12-01-2021,361883602
0.00879706,0.00987698,0.00677026,0.00979872,11-01-2021,626672424
0.00981436,0.01088934,0.00900523,0.01017289,10-01-2021,469749163
0.01019747,0.01084667,0.00955574,0.0097649,09-01-2021,510048283

Conclusion

Three types of machine learning algorithm are constructed and used for predicting the prices of three types of cryptocurrencies—BTC, ETH, and LTC. Performance measures were conducted to test the accuracy of different models as shown in Tables 2–4. Then, we compared the actual and predicted prices. The results show that GRU outperformed the other algorithms with a MAPE of 0.2454%, 0.8267%, and 0.2116% for BTC, ETH, and LTC, respectively. The RMSE for the GRU model was found to be 174.129, 26.59, and 0.825 for BTC, ETH, and LTC, respectively. Based on these outcomes, the GRU model for the targeted cryptocurrencies can be considered efficient and reliable. This model is considered the best model. However, bi-LSTM represents less accuracy than GRU and LSTM with substantial differences between the actual and the predicted prices for both BTC and ETH. The experimental results show that:

- The AI algorithm is reliable and acceptable for cryptocurrency prediction.

GRU can predict cryptocurrency prices better than LSTM and bi-LSTM but overall, all algorithms represent excellent predictive results. In future work, we will investigate other factors that might affect the prices of the cryptocurrency market, and we will focus on the effect that social media in general and tweets in particular can have on the price and trading volume of cryptocurrencies by analyzing tweets using natural language processing techniques and sentiment analysis

Results for BTC

The accuracies of these models for BTC cryptocurrency are tabulated in Table 2. The MAPE of the GRU model is the lowest with a value of 0.2454 and the RMSE is 174.129. Therefore, GRU is more capable of predicting BTC trends than LSTM or bi-LSTM, with a small difference between it and the LSTM model. Figures 11–13 show a visual representation, comparing the actual and predicted values of the training dataset of the three models for BTC.

Results presented in Figure 11 compare the actual and LSTM-predicted price of BTC. MAPE of the GRU model is the lowest with a value of 0.2454 and the RMSE is 174.129. Therefore, GRU is more capable of predicting BTC trends than LSTM or bi-LSTM, with a small difference between it and the LSTM model. Figures 11–13 show a visual representation, comparing the actual and predicted values of the training dataset of the three models for BTC. The graph shows that the predicted and the actual price is approximately the same over the entire interval. This model is considered the second-best model. The mean absolute. Results presented in Figure 11 compare the actual and LSTM-predicted price of BTC. percentage error for the prediction model (MAPE) of BTC for LSTM is 1.1234%, and the root mean square error (RMSE) is 410.399.

Statistical analysis of the data indicates that the predicted price has a mean value of 38,173.258 USD, a maximum value of 64,358.805 USD,
and a minimum value of 12,775.013 USD, whereas the actual price has a mean value of 38,249.388 USD, a maximum value of 63,380.999 USD, and a minimum value of 12,941.0 USD. The mean difference between the mean values of the actual and the predicated prices is 76.13 USD.

Results for ETH

The accuracies of these models for ETH cryptocurrency are tabulated in Table 3. The mean absolute percentage error for the GRU model is the least with a value of 0.8267and a root means square error of 26.59. Therefore, GRU proved to be the best predictor compared to LSTM and bi-LSTM for ETH. Figures 14–16 show the visual representation of the comparison between the actual and the predicted values of the training dataset of the three models for the ETH.
LSTM model for ETH, it represents that the difference between the predicted and the actual price is very small as red and green curves moving over each other’s over the whole period of time of Figure 14. This model is considered the second-best model. The mean absolute percentage error prediction model of ETH for the LSTM model is 1.5489% and the root mean square error is 59.507. Statistical analysis of the data indicates that the predicted price has a mean value of 1663.1392 USD, a maximum value of 4399.33 USD, and a minimum value of 379.41837 USD, whereas the actual price has a mean value of 1636.7091
USD, a maximum value of 4140.0 USD, and a minimum value of 383.35 USD.
The mean difference between the mean values of the actual and the predicated prices is 26.43 USD.

Results for LTC

The accuracy of the models for the LTC cryptocurrency are shown in Table 4. The mean absolute percentage error of the GRU model is the lowest with a value of 0.2116 and a root mean square error of 0.825. Therefore, GRU proved to be most capable for prediction as compared to LSTM and bi-LSTM for LTC. Figures 17–19 show the visual representation of the data by comparing the actual and the predicted values of training dataset of the three models for LTC. The results in Figure 17 show the comparison between the actual and the predicted price of the LSTM model for LTC. They show that the difference between the predicted and the actual price is very small with a mean absolute percentage error of 0.8474%, and a root mean square error of 3.069. Statistical analysis of the data indicates that the predicted price has a mean value of 166.16

USD, a maximum value of 388.59 USD, and a minimum value of 53.95 USD, whereas the actual price has a mean value of 165.68 USD, a maximum value of 373.64 USD, and a minimum value of 53.64 USD. The mean difference between the mean values of the actual and the predicated prices is 0.48 USD.

References

1. https://towardsdatascience.com/cryptocurrency-price-prediction- using-deep-learning-70cfca50dd3a
2. https://www.researchgate.net/publication/345262720_Bitcoin_Price_Prediction_and_Analysis_Using_Deep_Learning_Models
3. https://arxiv.org/abs/1911.11819
4. https://link.springer.com/chapter/10.1007/978-3-030-49186-4_9
5. https://www.sciencedirect.com/science/article/pii/S2405918821000027
6. https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039&context=datasciencereview
7. https://www.researchgate.net/publication/339401014_Fundamental_Research_Questions_and_Proposals_on_Predicting_Cryptocurrency_Prices_using_DNNs
8. https://iopscience.iop.org/article/10.1088/1757- 899X/928/3/032007/meta
9. http://202.62.95.70:8080/jspui/bitstream/123456789/12673/1/1NH16CS719.pdf
10. https://medium.com/@raghavaggarwal0089/bi-lstm-bc3d68da8bd0
11. https://towardsdatascience.com/forecasting-at-scale-with-facebook-prophet-a-case-study-on-its-merits-limitations-b24a694dd7c7
12. **https://www.researchgate.net/profile/Y-Dimopoulos/publication/222442208_Application_of_neural_networks_to_modeling_nonlinear_relationships_in_ecology/links/5ab3b165458515ecebeff1cb/Application-of-neural-networks-to-modeling-nonlinear-relationships-in-ecology.pdf**