Predictive Analytics in Cryptocurrency using Neural networks: A Comparative Study

Agha Salman Khan, Peter Augustine

*Abstract***:** *This paper is concerned with assessing different neural network based predictive models. Each of these predictive models has one goal and that is to predict the price of a cryptocurrency, Bitcoin is the cryptocurrency taken into consideration. The models will be focusing on predicting the USD equivalent value of bitcoin using historical data and live data. The neural network models being assessed are a Convolutional Neural Network, and two variations of the Recurrent Neural Network that are Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). The goal is to observe the validation loss of each model and also the time it takes to train or epoch for each training set which basically just determine its efficiency and performance. The results that are achieved are almost what was expected as LSTM outperforms CNN but the when we take a look at GRU, it is at par with LSTM.However, CNN is quicker at training or creating epochs and the validation loss is acceptable and not too high but it looks so when it is compared with the Recurrent Neural Networks such as Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU).*

*Index Terms***:** *Convolutional Neural Network (CNN), Cryptocurrency,Gated Recurrent Unit (GRU), Long Short Term Memory (LSTM).*

I. INTRODUCTION

One of the most important aspects of our lives, the global economy has been immensely influenced by the internet in which transactions between people in different parts of the world can be carried out online. We have seen different forms of currency develop over time starting from the period where commodities were exchanged as currencies to the popular fiat currency to the recent introduction of internet-based currencies. There are basically two forms of internet-based currencies, electronic money and virtual money the difference is that virtual money is defined as a traditional currency unit like AED, USD, Euro, etc. that is stored electronically to purchase goods and products online while virtual money has its own currency units that can be used to buy both online and offline products or services. The newest generation of virtual currencies is known as "Cryptocurrency". In 2008 Nakamoto published a paper called "Bitcoin: A peep-to-peer Electronic Cash System" this paper described a cash system that was completely decentralized and maintained A peer-to-peer relationship based ledger called blockchain and in 2009 Nakamoto released a software called bitcoin which was the first ever decentralised cryptocurrency, there were few attempts made before to achieve a cryptographic electronic cash system but none of them was truly decentralized like bitcoin. The creation of bitcoin dawned on a new type of currencies which also later encouraged the creation of other such cryptocurrencies, as of August 2018 the number of cryptocurrencies available to us is approximately 1600 and is rapidly growing. Cryptocurrency can be bought or sold using real or virtual currency according to specific exchange rates the same way one would exchange real currency [1].

Prediction of other financial markets such as stock has already been researched at length. Cryptocurrency, on the other hand, presents a totally different but an interesting picture as cryptocurrency market is still in its transitory stage it presents high volatily in the market. Bitcoin is still the leading cryptocurrency because it was the first ever and also because of its consistent growth over the years. Apart from its transitory state of the market, there is another interesting paradigm to Bitcoin, which is its open nature. Bitcoin operates on a peer-to-peer, trustless and a decentralised system in which an open ledger called the blockchain is maintained on which all the transactions are posted. Before Bitcoin or cryptocurrency, this type of transparency was never heard of in any other financial market. Traditional prediction methods especially time series prediction methods or models primarily rely on linear assumptions that are it requires data that can be broken down into a trend, seasonal and noise to be effective. These types of models and methodologies are only effective for tasks where seasonal effects are present like sales forecasting. Unfortunately, bitcoin lacks seasonality or trend which makes it highly volatile and renders these methods useless. Due to the complexity of the task deep learning makes for an interesting technology to be considered based on its performance in similar areas and especially artificial neural network models in deep learning This is because there is no definitive algorithm when it comes to artificial neural network, the model builds one as it keeps learning from the data which helps immensely with volatile data [2]. This paper focuses on assessing artificial neural networks model that predicts the price of a cryptocurrency. Bitcoin is the cryptocurrency being considered here since it was the first and is the most popular cryptocurrency.

Agha Salman Khan*, Department of Computer Science, CHRIST (Deemed to be University),Bangalore, Karnataka 560029,India
Dr. Peter Augustine, Department of Computer Science, CHRIST (Deemed to be University),Bangalore, Karnataka 560029,India

The artificial neural networks considered are a Convolutional Neural Network, Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU), LSTM and GRU are variants of Recurrent Neural Networks. The models are compared by their validation loss.

II. LITERATURE REVIEW

A. Related work

Sean McNally: This paper is concerned with predicting the price of Bitcoin and it achieves this using an implementation of a Bayesian optimised recurrent neural network (RNN) and Long Short Term Memory (LSTM) network and also compared with ARIMA which is a popular time series furcating model but as expected the deep learning model outperforms ARIMA with an accuracy of 52% and an RSME of 8% [2]. Isaac Madan, Shaurya Saluja and Aojia Zhao: The paper is divided into two phases, in the first phase all the models and algorithms are applied to the Bitcoin data to see which of them would perform efficiently and would have fewer errors while classifying. The Binomial Generalized Linear Model (GLM) performs with precision although the Random Forest Model is more accurate it is less precise. SVM increased the rate of errors because of which binomial GLM and Random Forest were selected for the second phase. In the second phase, the GLM and Random Forest time series models are used to find the pros and cons of having a big or a small window to determine which would be more efficient to implement in the automation. The big window would be more efficient to implement because the only drawback is missing a short-term hike or burst in price but the overall gain is much more in a larger window. And the paper also concludes with an observation that the Random Forest model gives higher accuracy when compared to GLM because Random forests use nonparametric trees hence outliers and linear separability of data are not involved [3]. Brandon Ly, Divendra Timaul et al.: This paper is concerned with creating a deep learning model that predicts the price of Bitcoin. The paper uses trial and error method in creating the algorithm. It creates different models all with a different combination of optimizers and activation function to find a combination that would result in a deep learning model that would predict the price of bitcoin. The resulting model produced results that aren't perfectly accurate but still, the model has predictive capabilities [4]. Fedor Lisovskiy: this paper uses LSTM networks to predict cryptocurrency prices. In this paper, they train two LSTM networks one solely using cryptocurrency price data and the other using both the cryptocurrency price data and sentiment features. After 1000 epochs of training the network trained just with price data had an accuracy of 49.2% and the network with price data and sentiment features had an accuracy of 52% [5]. Bruno Spilak: this paper introduces a Neural Network framework that provides a deep machine learning solution to the price prediction problem. The framework is built with three instances a Multilayer Perceptron (MLP), a simple RNN and an LSTM. This paper also shows how LSTM is useful for trend prediction as it achieved a high prediction accuracy on cryptocurrency data [6]. Matthew Chen, Neha Narwal and

Mila Schultz: This paper talks about price prediction in Ethereum and to reach the goal of predicting a number of methods and models were assessed and out of them ARIMA outperformed Random Forest, SVM, Naïve Bayes and RNN this is mainly because the data used was time series data and unstructured with price features that are not likely to repeat. The reason for the underperformance of the neural networks is that it may not have run sufficient iterations to facilitate convergence to the global minima of their objective functions. Regardless of all facts considered all the methods scored an accuracy score of 50% and above with ARIMA at the top with 61.7% [7]. Sneha Gullapalli: to perform prediction this paper has used temporal neural networks like Time-Delay Neural Network (TDNN) and RNN on historical data. The models are compared by computing various measures like r (Pearson's correlation coefficient), MSE and NMSE on the continuation of time series, held out for validation. The results show that TDNN is a more accurate and efficient as it takes lesser time to train, has fewer errors and also has a higher Pearson's correlation coefficient (r) [8].

B. Convolutional Neural network

Image classification and text recognition are some of the common applications of Convolutional Neural Networks. When it comes to capturing invariant patterns, Convolutions are a very powerful tool. When we are trying to identify an animal like a rat or a dog, the position of its whiskers in the picture hold very less value compared to their presence in the picture similarly when classifying a document as a legal document, we focus more on identifying the presence of legal jargons than the positions. similarly, there might be some patterns that occur periodically like certain autocorrelation structures that could be handled by convolutions and there are some specialized convolutional architectures that could be considered for a tie series prediction task [9],[10].

C. Recurrent Neural Network

One of the biggest drawbacks of traditional neural networks for a while was the ability to interpret sequences of inputs which depended on each other for information and context. This information can be anything it could be previous words in a sentence that might give some sort of context that would help in predicting the next word or it could be some temporal information a sequence that would allow some sort of context o the time-based elements of that sequence. In other words, each time a traditional neural network takes in independent data vectors they have no place or memory to store them because they lack the concept of memory and this is a huge set back when it comes to a task that requires memory. The use of a simple feedback type approach for neurons in the network was basically an early attempt to tackle this problem where the output was fed-back as input to provide context on the last seen input.

These neural networks with a feedback type approach were called Recurrent Neural Networks (RNNs). But These RNNs worked only to a certain extent, they had a flaw that when used in any scenario it would lead to a problem called the Vanishing Gradient Problem. Whenever the Back-propagation method is used that is calculating gradients of loss in regards to the weight while moving backwards, the gradient keeps getting smaller and smaller. This means that the neurons in the Earlier layers learn very slowly as compared to the neurons in the later layers in the Hierarchy. The Earlier layers in the network are slowest to train. Earlier layers are important as they are expected to learn and detect the simple patterns and are actually the building blocks of our Neural Network and if they produce inaccurate results, then how is it expected of the next layer and the entire neural network to perform adequately and generate accurate outcomes as it results in training process taking a lot of time and accuracy of predictions decreasing. This is what a vanishing gradient problem does to a Neural Network model and because of this issue RNNs are poorly suited in most real-world problems [11],[12],[13],[14].

D. Recurrent Neural Network

Just like the neurons in RNN, the neurons in LSTM also keep a context of memory within their pipeline to allow for tackling sequential and temporal problems without the issue of the vanishing gradient affecting their performance. But still, they usually face the problem of overfitting which is why it has to go through regularization [11].

E. Gated Recurrent Unit

GRUs are also an improvised version of the standard recurrent neural network. GRU uses an update gate and reset gate to solve the vanishing gradient problem of the RNN. The update gate and reset gate are basically two structures of data or vectors which decide what data or information should be passed to the output. The speciality of GRU is that they can be trained to keep information from the past, without washing it through time or removing information which is irrelevant to the prediction [15].

III. METHODOLOGY

In this paper, we aim to study different artificial neural networks and their performance on their ability to predict cryptocurrency data and this will be determined by their validation loss. Artificial Neural networks are considered here because of the volatility in the cryptocurrency data and the lack of trends in them which makes linear assumptions and methods useless. Since there is no trend or seasonal pattern in the data set it becomes extremely difficult to devise an algorithm or conditions that would help in the prediction. Hence, we use artificial neural networks which do not have a fixed algorithm for its interpretations or predictions. They, in turn, learn on each iteration and focus towards improving the validation loss which is a summation of the errors made for each example in training or validation sets. CNN, LSTM and GRU are the different variants of artificial neural networks being considered because each of them has their unique method of going about the process which brings in its certain advantages and disadvantages. We will also use each of the networks with a different activation function to see which

combination is more efficient.

A. Activation Function

We tested each model and combined it with different activation functions. The main purpose of activation function is to convert the input signal of a node in a neural network to an output signal they basically just introduce the much-needed non-linear properties to the neural network without an activation function converting the signals the model would just be a simple regression function [12].

B. Rectified Linear Units (ReLu)

Rectified Linear Units has become very popular among the last few years. (1) is the mathematical form of the function and one can observe it is a relatively very simple and yet very efficient. Sometimes the simplest methods prove to be the most efficient and the most used. ReLU avoids the Vanishing gradient problem and also rectifies it as a result almost all the deep learning methods use it nowadays. But, it also has its limitations, a ReLU can only be sued in the hidden layer of the neural network model and another problem with the ReLU is that sometimes gradients can be fragile while training which can result in them dying. Simply put, a ReLU could sometimes result in a dead neuron [12].

$$R(x) = max(0, x) \qquad (1)$$

C. Leaky Rectified Linear Units (Leaky ReLu)

To fix the problem of dying neurons Leaky ReLU was introduced which uses a small slope to keep the updates alive [12].

D. Hyperbolic Tangent function (Tanh)

(2)is the mathematical form of the function. Its output is centred 0 because its range is between -1 and 1. This function makes optimization a lot easier and hence is always preferred over the sigmoid function. But, it suffers from vanishing gradient problem [12].

$$f(x) = 1 - exp(-2x) = 1 + exp(-2x) \qquad (2)$$

IV. RESULT

The Table I has the results of every model and its variations. Where in each model the number of layers were changed and also the activation functions were changed to see which is the most optimized and efficient model

A. CNN

The most efficient CNN model used here is a three-layer CNN model and uses leaky RELU as its activation function and uses Mean Squared Error (MSE) as its loss function and the state-of-the-art Adam as its optimizer and also uses a dropout layer to avoid overfitting. It takes about 2 secs per epoch and has a validation loss of 0.00022.

Table I: Results of different models

Model	Layers	Activation	Validation Loss	Inverted	Secs/epoch
CNN	3	ReLU	0.00025	183682	2
CNN	3	Leaky ReLU	0.00022	90664	2
LSTM	1	tanh + ReLU	0.00007	28721	45
LSTM	1	tanh + Leaky ReLU	0.00004	17436	45
GRU	1	tanh + ReLU	0.00004	19739	40
GRU	1	tanh + Leaky ReLU	0.00004	17546	40

B. CNN

The most efficient CNN model used here is a three-layer CNN model and uses leaky RELU as its activation function and uses Mean Squared Error (MSE) as its loss function and the state-of-the-art Adam as its optimizer and also uses a dropout layer to avoid overfitting. It takes about 2 secs per epoch and has a validation loss of 0.00022.

C. LSTM

The most efficient LSTM Network used here is a single-layer LSTM Network and uses tanh and leaky RELU as its activation functions and uses Mean Squared Error (MSE) as its loss function and the state-of-the-art Adam as its optimizer, it uses a dropout layer to avoid overfitting but also uses Regularization. It takes 45 secs per epoch and has a validation loss of 0.00004. The reason we use tanh + ReLU is to make optimization easier which is provided by tanh and avoid vanishing gradient problem with the help of ReLU. The advantage of using tanh + Leaky ReLU is that Leaky ReLU it fixes the problem of dead neurons faced by ReLU.

D. GRU

The most efficient GRU Network used here is a single-layer GRU Network and uses tanh and leaky RELU as its activation functions and uses Mean Squared Error (MSE) as its loss function and the state-of-the-art Adam as its optimizer, it uses a dropout layer to avoid overfitting but also uses Regularization. It takes 40 secs per epoch and has a validation loss of 0.00004. The reason we use tanh + ReLU is to make optimization easier which is provided by tanh and avoid vanishing gradient problem with the help of ReLU. The advantage of using tanh + Leaky ReLU is that Leaky ReLU it fixes the problem of dead neurons faced by ReLU.

V. CONCLUSION

The study on the neural networks used, to predict cryptocurrency price have some expected and some unexpected results. When we study the results, we come to observe that the activation function plays an important role. As we can clearly see ReLU is the not the best performing activation function when compared to Leaky ReLU and tanh + Leaky ReLU. When we come to observe neural networks as a whole the GRU network stands out as it performs at par with the LSTM network but with a bit more efficiency as it only takes 40 seconds per epoch as compared to the 45 seconds per epoch of the LSTM network. CNN models are the fastest that can be trained as they only take 2 seconds per epoch but have a higher validation loss when compared to LSTM and GRU. The best model among the ones tested is the LSTM with tanh and Leaky ReLU as the activation function as GRU with tanh and Leaky ReLU is almost at par with LSTM but is slightly more efficient, but the CNN 3 layered model is far less time consuming and a bit more effective at capturing local temporal dependency of data. The CNN model apart from being quicker than LSTM and GRU also has another advantage that is it does not suffer from the problem of overfitting like LSTM and GRU although this problem is solvable through the method of Regularization but it is still time added to the total executing time . The CNN model is not explored to its maximum neither are the LSTM and GRU models as part of future work we could measure their accuracy and also compare the performance of each model by changing the number of layers used in the network.

REFERENCES

1. Jeremiah Bohr, Masooda Bashir. Who Uses Bitcoin?An exploration of the Bitcoin community. (US): Twelfth Annual Conference on Privacy, Security and Trust (PST); 2014.
2. Sean McNally. Predicting the price of Bitcoin using Machine Learning. (IE): School of Computing National College of Ireland; 2016.
3. Isaac Madan, Shaurya Saluja, Aojia Zhao. Automated Bitcoin Trading via Machine Learning Algorithms. (US): 2014.
4. Brandon Ly, Divendra Timaul, Aleksandr Lukanan, et al. Applying Deep Learning to Better Predict Cryptocurrency Trends. (US): 2017.
5. Fedor Lisovskiy. Using LSTM Networks to Predict Cryptocurrency Prices: 2017.
6. Bruno Spilak. DEEP NEURAL NETWORKS FOR CRYPTOCURRENCIES PRICE PREDICTION. (DE): 2018.
7. Matthew Chen, Neha Narwal, Mila Schultz. Predicting Price Changes in Ethereum. (US): 2017.
8. Sneha Gullapalli. Learning to predict cryptocurrency price using artificial neural network models of time series. (US): 2016.
9. Eddy~J. Time Series Forecasting with Convolutional Neural Networks - a Look at WaveNet [Internet]. Jeddy92.github.io. 2018 [cited 2018 Nov 16]. Available from:https://jeddy92.github.io/JEddy92.github.io/ts_seq2seq_conv
10. Pokharna H. The best explanation of Convolutional Neural Networks on the Internet! [Internet]. Medium. 2018 [cited 2018 Nov 16]. Available from: https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8
11. Aungiers J. Time Series Prediction Using LSTMDeep Neural Networks [Internet]. Altumintelligence.com. 2018 [cited 2018 Nov 16]. Available from: https://www.altumintelligence.com/articles/a/Time-Series-Prediction-Using-LSTM-Deep-Neural-Networks
12. Singh Walia A. The Vanishing Gradient Problem – Anish Singh Walia – Medium [Internet]. Medium. 2018 [cited 2018 Nov 16]. Available from: https://medium.com/@anishsingh20/the-vanishing-gradient-problem-48ae7f501257
13. Pokharna H. For Dummies — The Introduction to Neural Networks we all need ! (Part 1) [Internet]. Medium. 2018 [cited 2018 Nov 16]. Available from: https://medium.com/technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-c50f6012d5eb
14. Pokharna H. For Dummies — The Introduction to Neural Networks we all need ! (Part 2) [Internet]. Medium. 2018 [cited 2018 Nov 16]. Available from: https://medium.com/technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-part-2-1218d5dc043
15. Kostadinov~S. Understanding GRU networks – Towards Data Science [Internet]. Towards Data Science. 2018 [cited 2018 Nov 16]. Available from: https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be
16. Singh Walia A. Activation functions and it's types-Which is better? [Internet]. Towards Data Science. 2018 [cited 2018 Nov 16]. Available from: https://towardsdatascience.com/activation-functions-and-its-types-which-is-better-a9a5310cc8f

AUTHORS PROFILE

Agha Salman Khan Post Graduate Scholar from Department of Computer Science, CHRIST (Deemed to be University),Bangalore, Karnataka 560029,India.

Dr PETER AUGUSTINEMCA, MPhil, PhD.Associate Professor, Department of Computer Science, CHRIST (Deemed to be University),Bangalore, Karnataka 560029,India.

Study of Phishing Attacks and Preventions, International Journal of Computer Applications, 18-Apr-2017. Comparative Study of Marching Cubes Algorithms for the Conversion of 2D image to 3D, International Journal of Computational Intelligence Research, 31-Mar-2017. Efficient Brain Tumor Classification using PCA and SVM, International Journal of Research in Engineering, IT and Social Sciences, 31-Mar-2017. Bad effects of urbanization and lifestyles, Population Health improvements using Analytics, improvements using Analytics, International Research Journal of Engineering and Technology, 31-Mar-2017. Comparative Study of Clustering algorithms using R, International Journal of Research in Engineering, IT and Social Sciences, 28-Feb-2017. CBIR: IMAGE INDEXING AND RETRIEVAL FOR DIGITAL IMAGES, INTERNATIONAL JOURNAL OF RESEARCH IN IT, MANAGEMENT AND ENGINEERING, 28-Feb-2017. Performance Evaluation of Parallel Genetic Algorithm for Brain MRI Segmentation in Hadoop and Spar. Indian Journal of Science and Technology, 22-Dec-2016

Enhancing the Efficiency of Parallel Genetic Algorithms for Medical Image Processing with Hadoop, International Journal of Computer Applications, 10-Dec-2014.Leveraging Big Data Analytics and Hadoop in Developing India's Healthcare Services, International Journal of Computer Applications, 12-Mar-2014.