International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 9 Issue 4 Series III || Apr.2020 || PP 14-24

An Enhanced Cryptocurrency Prediction Model Using Sentiment Analysis and Machine Learning

Onuodu Friday Eleonu[1], Ufio Augustine Nkemjika[2]
[1]*Department of Computer Science,University of Port Harcourt, Nigeria*
[2]*Department of Information Technology, National Open University, Nigeria(NOUN)*
[1]*gonuodu@gmail.com*

ABSTRACT:*Cryptocurrencyrefers to a form of digital or virtualcurrencywhichisused as a legal tender for exchange of goods and services. Cryptocurrencyis one of the fastestgrowing concepts in recent times which have attracted the attention of researchersfromvariousfield of endeavor, rangingfromacademics, Finance, Commerce etc. Bitcoin is a major example of Cryptocurrency. It isalso one of the mostcommonlyusedCryptocurrency for transactions. Several Bitcoin Predictionmodels are currently in use for the forecasting of the future price of the Bitcoin, yetthere are still open problems of inefficient predictions due to the use of poorpredictiontools for the prediction. In thiswork, an enhancedCryptocurrencyPrediction Model has been developed to eliminate the inefficiencies of the Bitcoin prediction System. The workproduced a predictionaccuracyUsing Long Short Term Memory (LSTM) as a machine learningapproach and Sentiment Analysis of data from Twitter and Facebook as opposed to the previouslydevelopedmodels. Tools such as VADER wereused for processing the data gottenfrom the web to remove the redundant data. The Structured System Analysis and Design Methodology (SSADM) wereadopted in thisapproach. The system wasimplementedusing Python and Java programminglanguages on the Jupyter notebook IDE The result shows thataccurate Bitcoin Price Prediction can nowbe made using large datasets and data from the social media and other web sites. This workcouldbebeneficial to Forex traders, to Investors, to Scientists, to Financial Analysts, and to people in differentresearchcommunitiesthattransactwithvirtualcurrencies*

KEYWORDS*Prediction, Market, Stock, Mining, Cryptography, Time Series, Sentiment*

Date of Submission: 16-04-2020 Date of Acceptance: 01-05-2020

I. INTRODUCTION

Crypto currency is a concept whose popularity has grown aggressively since its introduction by David Chaum in the 1980s, and has attracted a lot of attention to itself both in industry and academia. Its research spans across several fields such as economics, science, finance etc. It is currently at the forefront of financial development due is its continuous growth the market capitalization which rose up to $800 Billion in January, 2018. Cryptocurrency is a compound name for electronic transaction which involves the use of virtual currencies sometimes called digital money. Some types of Cryptocurrencies are Bitcoin (this is the most popular), Litcoin, Etherum, Ripple etc.

Some financial institutions have discouraged people from investing in Cryptocurrency with a claim that it is not a recognized legal tender, these institutions warn that great risks such as unenforceability of Cryptocurrency are associated with Cryptocurrency transactions, and so advise the public against transacting with Cryptocurrencies. Despite the many different views of people and sectors on the usefulness or danger of Cryptocurrency, the rate of subscribers to the system keeps increasing at a high rate daily.

The Cryptocurrency market is one characterized by high value change, high market data availability, smaller capitalization therefore a ledger was developed to regulate the activities of all owners of Bitcoin, and this ledger is referred to as block chain technology, it is a decentralized system and is managed by the Bitcoin owners and not by private sectors such as banks or other financial institutions. This ledger also allows for safe transfer of values between users. Though the Cryptocurrency market is a fast growing one, it is faced with a challenge of not being able to fundamentally value Cryptocurrency in the same way that stock is valued. However, since prices are highly driven by speculations, the most practical way to value each Cryptocurrency will be to predict the future price on the basis of the data gotten from the previous price data [1].

Over the years, Time Series have been used as a means to predict the future value of Cryptocurrency, however our study will focus on the use of a hybrid model involving the use of Sentiment analysis and Machine learning for the Cryptocurrency prediction and the Bitcoin will be used as the case study for the prediction. The use of the Sentiment analysis and the Machine learning model will produce a prediction which will be applicable in the new and emerging Cryptocurrency market.

This research discusses the problem associated with the present prediction system, some of these problems include: volatility, complexity, and inefficiency. We are however going to address the inefficiency of prediction models which makes them not applicable to the new and emerging Cryptocurrency market using efficient tools.

1.1 AIM AND OBJECTIVES

The aim of this research is to develop an enhanced Cryptocurrency prediction model using sentiment analysis and machine learning:

i) design an internet crawler to fetch news from twitter which will then be used for sentiment analysis.
ii) train the model that will be able to predict the future crypto-currency value using the data from March 2018 to October 2019.
iii) implement with Python and JAVA programming languages.
iv) compare results with existing

II. RELATED WORKS

Alessandretti et al [2] presented on Anticipating Cryptocurrency Prices Using Machine Learning, proposed an approach to test the hypothesis that the inefficiency of the Cryptocurrency market can be exploited to generate abnormal profit. To do this they analyzed daily data of 1,681 Cryptocurrencies including Bitcoin with the data gotten from 300 exchange market platforms for the period between October 2015 and November 2018 using Machine Learning methods. Their result showed that simple trading strategies when combined with state of the art Machine Learning algorithms over performs other standard benchmarks. Their result also shows that nontrivial but ultimately simple step to step mechanism will help to forecast the short term evolution of Cryptocurrency market. Some of the Machine learning techniques they employed include the gradient boosting decision trees, Long Short Term Memory Recurrent Neural Network. They did not however exploit the existence of different prices on different exchanges, the consideration of which could open the way to significantly higher returns on investments. They did not also consider the impact of public opinion on market behavior (Sentiment analysis) for better prediction of Cryptocurrencies especially Bitcoin.

Chhatwani [3] presented on Predicting Bitcoin Price Fluctuations by Analyzing Global Currency Patterns & Sentiments, proposed a study to analyze Bitcoin prices using machine learning and Sentiment analysis. He analyzed the impact global currencies such as the US Dollar foreign exchanges on Bitcoin prices and whether Bitcoin has the stability to dethrone other global currencies and become the single medium of transaction. With the proposed model, he was able to achieve a near to accurate prediction. Although he achieved this but he could not stream Bitcoin data continuously as a result, he had to add guessed price data for some days.

Valencia et al [4] proposed Price Movement Prediction of Cryptocurrency Using Sentiment Analysis and Machine Learning. They proposed the usage of common machine Learning tools such as the Neural Networks, Support Vector machines and the Random Forest, and other available social media data (Twitter) for predicting the price movement of Bitcoin, Litcoin, Etherium, Ripple Cryptocurrencies. Their work proved the possibility of using these tools to accurately predict the direction of price movement for the emerging Cryptocurrency market. However, they could not produce a high quality sentiment analysis due to the fact that they did not filter the comments gotten from twitter which contained a lot of redundant data.

Spiklak [5] proposed Deep Neural Networks for Cryptocurrencies Price Prediction, proposed a Neural Network framework to produce Deep Machine Learning solution to the prediction problem. He employed the three main instants of the framework: Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), and Multi Layer Perceptron (MLP) to carry out this task. He studied the models and used them to predict the direction of major Cryptocurrencies with a rolling window regression method. From his study the summarized that the LSTM has the best accuracy in predicting the directional movement of most Cryptocurrencies. However his strategy could not analyze and understand the financial cost of misclassification using deep neural network which would have led to a better prediction.

Jones [1] presented Feedforward Machine Learning Systems to Improve Cryptocurrency Pricing Models, proposed a study to use historical Cryptocurrency trade data and Machine Learning models to generate prices that are better than market prices. He used an autoregressive linear model, several baseline machine learning models, and systems of machine learning models to implement the work. He tested accuracy of each model in predicting the future 1 hour percent change in price of ten (10) different Cryptocurrencies. However, he could only predict the price change for a very short period due to the fact that his data was very narrow.

Gullapalli [6] presented learning to predict Cryptocurrency price using artificial neural network models of time series. They proposed a study to predict the daily price, particularly the daily high and closing price, of the Bitcoin which will play a vital role in making trading decisions. To perform the prediction, He trained temporal neural networks such as Time-Delay Neural Networks (TDNN) and Recurrent Neural Networks

(RNN) on historical time series i.e. past prices of Bitcoin over several years. He implement TDNNs and RNNs using the NeuroSolutions artificial neural network (ANN) development environment to build predictive models and evaluated them by computing various measures such as the MSE (mean square error), NMSE (normalized mean square error), and r (Pearson's correlation coefficient) on a continuation of the training data from each time series, held out for validation. However, he could not perform multi-step ahead prediction of Bitcoin price thereby increasing the scope of experiments.

Bai et al [7], presented a Collective Cryptocurrency Up/Down Price Prediction Engine. They proposed a study that analyses the problem of predicting whether the price of twenty one 21 most popular Cryptocurrencies (from the data gotten from coinmarketcap.com) will go up or down on particular day using data up to the previous day. They developed an algorithm (C2P2 algorithm) for the implementation. This algorithm was the first algorithm to consider the fact that the price of a Cryptocurrency might not only depend on historical prices, sentiments, global stock indices, but also on the prices and predicted prices of other Cryptocurrencies. The algorithm predicted the prices of all the Cryptocurrencies at once and not one coin at a time. However, their algorithm could not predict the price of Cryptocurrency for more than one day.

Kaushik et al [8] defined sentiment analysis as a process of automatic extraction of features by mode of notions of other people about specific product, services or experience. He stated that it is the process of using machine learning algorithm to analyze the emotions behind data from people. Sentiment analysis is also called Opinion mining because it deals largely on the opinion of others. These opinions affect the business, person or entity it talks about to a very great extent. The opinion of the people may be in form of a judgment, mood or evaluation of the writer, but it is generally categorized under negative and positive comments. In recent times the internet has become a major source of opinions. Sentiment analysis is carried out in two major ways: the first is the direct opinions which involve making positive or negative comments about a particular thing, and the second is through comparison which involves comparing two things that are similar.

Juarez and Manzanilla [9] carried out a review to analyze and compare different Bitcoin prediction models. Their emphasis was based on a model known as the ARIMA (Autoregressive integrated mobile average) model; this model is a combination of the pure autoregressive series (AR) and pure moving averages. They also added other hybrid models which use neural networks to complete the method. With the combination of these models at different levels they were able to predict the value of Bitcoin for a given sample period. They observed that when the models were combined there was a 70% positive result as opposed to the 30% result gotten from using the models separately. They were not however able to determine the factors that contribute to the volatility of the Bitcoin exchange rate, the correlation of Bitcoin with other currencies was not also analyzed.

Jain et al [10] developed a Cryptocurrency forecasting method using Tweets Sentiment Analysis; their task was to predict a two hour price change of Cryptocurrencies specifically Bitcoin and Litcoin Using the Multi-Linear Regressive model. Although they were able to predict the price of Bitcoin Litcoin in two hours using the feeds from twitter, yet they were not able to achieve user credibility, popularity of user, user network are some other social factors that can be considered to measure the price of cryptocurrency and for increasing the accuracy of the price prediction model.

Shirakawa and Korwatansakul [11] accessed the effectiveness of government institution and de jure financial openness of the policy makers in pursuing further financial development by allowing the use of Cryptocurrency. They carried out their estimation using a method called Cross-Sectional ordered probit using institutional and macroeconomic data drawn from sources such as the Chinn-lto index, World Bank's World Wide Governance indicators etc. Their results show that a certain level of institutional quality may be necessary before opening up to the new forms of technology such as Cryptocurrency. They reported that though they effectively carried out their research, yet they were not able to conduct a systematic investigation of the policy, economic and institutional factors influencing policy choice.

III. OUR APPROACH

We have successfully reviewed a number of related research papers on the different techniques and models which has been used to predict the price of Cryptocurrencies particularly the Bitcoin in the paragraphs above. But our main focus on a related work by Chattwani [3]. They looked at the prediction of Cryptocurrency fluctuations using Sentiment Analysis and Machine Learning for four Cryptocurrencies (Bitcoin). They extended the application of financial time series forecasting method with machine learningand sentiment analysis techniques to alternative Cryptocurrency markets.

For the sentiment Analysis a method known as the VADER (Valence Aware Dictionary and Sentiment Reasoner) was used to analyze the data gotten from the social media (Twitter) which contained the opinion of people about the prices of the Cryptocurrencies. VADER was chosen because it performed extremely well for their purposes in independent benchmarks, its open source and is validated by human and is specifically attuned to twitter data.

For the Machine learning model they used the Multilayer Perceptron (MLP) which is a type of the neural network, the Support Vector Machine (SVM), and the Random Forests (RF). All the models used are supervised machine learning models. They were used to train the system based on the previous data collected with the previous predictions, and from time series analysis over a period of time. However the authors were only able to use contents from the Twitter social platform which led to a myopic view of the opinions of users and low accuracy rate of the opinion analysis. Also the Neural Network models used are not as effective as the LSTM which is regarded as a more effective approach. In other words, we intend to improve on the work of Chattwani (2019), by collecting social data from Twitter and Facebook and using the LSTM model for the machine learning model.

3.1. EXISTING SYSTEM

The existing system is a Bitcoin fluctuation prediction system using Twitter sentimentsanalysing and machine learning. The data for the research was gotten from several sources which include Yahoo Finance, Twitter and Web Blogs. The data was further checked for authenticity in order to ensure that there were no redundant or useless data values present which could affect the end results in any manner. Initially, the Bitcoin Price data was gotten from Yahoo Finance and it was crosschecked with the one gotten from Coinbase to remove any discrepancies. This data when applied in the prediction produced wrong results (predictions), therefore a shift was made to the QuandlAPI to fetch three datasets from three sources Kraken, CoinBase, BitStamp which proved to be more efficient since the results now obtained were more correct.

The data for the sentiment analysis was collected from twitter, tweets from different twitter handles were collected by using Twitter API's for python. The data was collected by writing Python codes and automating the feature of reading text and running prediction algorithms on them. Tweepy, a twitter library written in Python was used to search out all related tweets using specific key words. To filter the tweets which were gathered (since data generated from these sources contain some irrelevant data) in order to collect only the useful data, they used the regular expressions library of python (the 're' library for removing the regular expression) and the retweet library ('RT' for removing the retweets). The TextBlob (A twitter API) was used to analyze the data which has been cleaned in order to deduce the sentiment. TextBlob an algorithm based on Naïve Bayes easily predicted the sentiment scores of the data that was provided to it, however due to some inefficiencies the TextBlob, the machine learning technique (Recurrent Neural Network) was used for the sentiment classification.

The prediction process for Bitcoin involved using Machine Learning techniques for running a neural networks-based prediction model for predicting the next day price. The accuracy of the prediction depended on the quality of the dataset accurate or near-to-accurate prediction. Their training datasets were selected from 20th May 2014 to 14th January 2018. The machine learning process was implemented on a framework called Keras, which is a neural network API running on top of TensorFlow. The system was implemented using the Python Programming Language.

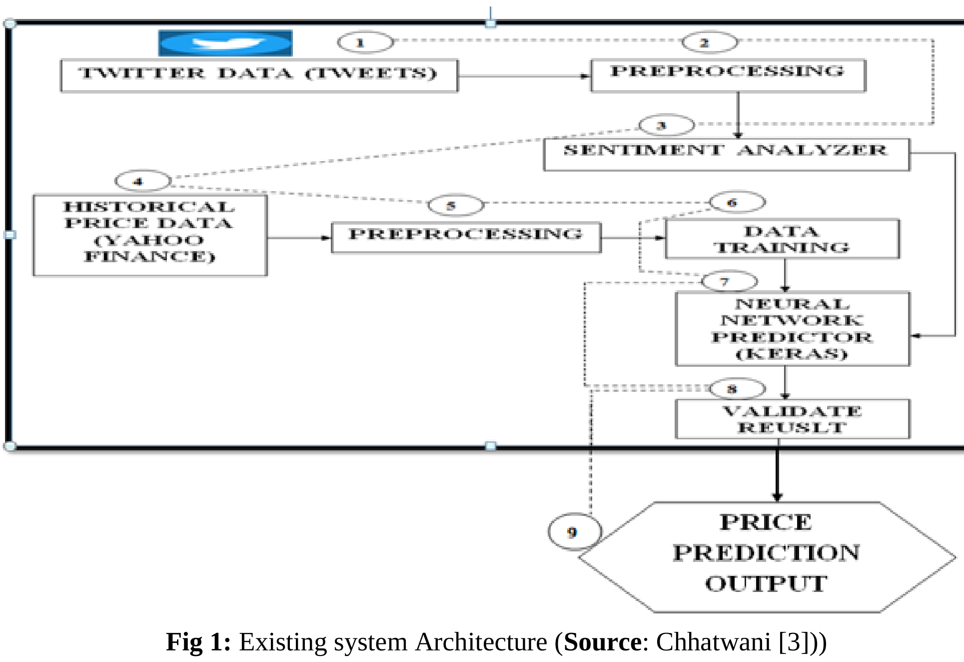

Fig 1: Existing system Architecture (**Source**: Chhatwani [3]))

Twitter: Twitter is used as the source of web opinion data. This choice was made because twitter has a very large number of users all around the world who actively render thoughts and opinions on certain issues (political, financial etc). Some of these users are Bitcoin traders, which makes it the most suitable platform to mine Bitcoin sentiment data.

Preprocessing: Twitter comes with certain API's which can process data gotten from it by removing the redundant data and retweets, usually referred to as noisy data. The data are filtered with the twitter filter to produce quality data which will yield positive sentiment analysis.Sentiment Analyzer: Machine learning techniques were used to classify the data. The data was classified as positive, negative or neutral sentiment and the entire classification process was supervised learning. Keras was used to train the dataset in order to create a Bag of Words and testing dataset built from gathering data from Twitter. TextBlob is a Naïve bayes algorithm which predicts the sentiment scores of the data. It also performs the classification into positive, negative, neutral, based on a certain cutoff value.

Collection of Historical Price Data (Bloomberg, Yahoo finance, Coin Desk etc.): Web crawlers are programs which collect data from sites. Using a particular keyword, it collects all data related that keyword and makes it available for the user. Examples of these crawlers include the Google crawlers, and some user defined crawlers which are designed by the user himself to suite his mining needs.

Data Training: Quandl datasets was chosen due to the volatile changes in Bitcoin prices and yahoo finance not keeping up with the updated values in their datasets. A good training dataset and testing dataset can help to achieve an objective with greater accuracy compared to a poorly created dataset, this is another reason Quandl datasets was chosen.

Neural Network Predictor: KERAS is a neural network API framework which runs on TensorFlow. It is less expensive in performance and computation than most frameworks. Along the line, due to some inefficiency of the trained data, they switched to the Quandl's REST API which proved more accurate and productive.

Validate Result: During this phase, the result from the prediction is being checked against the actual price of the Bitcoin for the period of the prediction.

3.1.1. Disadvantages of the Existing System

i. The existing system worked with a limited stream of Bitcoin price data which would have produced a better prediction model.

ii. The medium for social media opinion mining (Twitter) contained a lot of redundant and noisy data, so a lot of time was required for cleaning up the data. This could have been avoided if more than one social media was sourced.

3.2. PROPOSED SYSTEM

Our proposed improvement of the existing system is an enhanced Cryptocurrency prediction model using sentiment analysis and machine learning, as further illustrated in Fig 2. Since the price or value of the Bitcoin is influenced by several factors, opinions, market value, governmental laws etc. it was necessary to use a hybrid method for the prediction of the price prediction. A major attribute of our proposed system is the existence of an additional social media platform for opinion mining and also the use of the Long Short Term memory supervised machine learning model, which will produce a better Bitcoin price prediction or forecast. Facebook is the second most populated or most visited social media platform for users all around the world. LSTM was also chosen because it's best for remembering long term sequences, and analyzes sequential datacorrectly. This also implies that a good number of Bitcoin traders, investors and critics are part of the users of this platform. Therefore opinions on the topic of Bitcoin price fluctuation and general news about Bitcoin prices will be widely expressed and discussed on this platform. And the cost of accessing Face is relatively cheaper than the cost of accessing most other platform, this is because most times you require zero data to access Facebook (Facebook for free). Combining Twitter and Facebook for mining sentiment data using Facebook and Twitter crawlers to legally scrape comments will yield a wider pool of opinions or sentiments which will be useful for predicting the correct price of the Bitcoin. Tweepy, Twitter's API for Python and Facebook API for Python will be used to extract all related data from these social media sources using the keyword "BITCOIN". The Python regular expression library ('re') and the retweet library ('RT') will be used to remove noise from the data scraped from Twitter and Facebook. It will remove duplicate data, irrelevant data, based on a given criteria.

For the Bitcoin price data, Kafka's API from Quandl will be used to scrape the data from Kraken, Coin Base and BitStamp dataset sources. This is because these sources keep up with all the fluctuations of Bitcoin data, and so we will not record any missing data because there was a continuous streaming of the price data.

For the machine learning approach, we adopted LSTM (Long Short Term Memory), a supervised learning approach based on Neural Network which recorded 90% classification accuracy for prediction and time series analysis. It was compared to other supervised machine

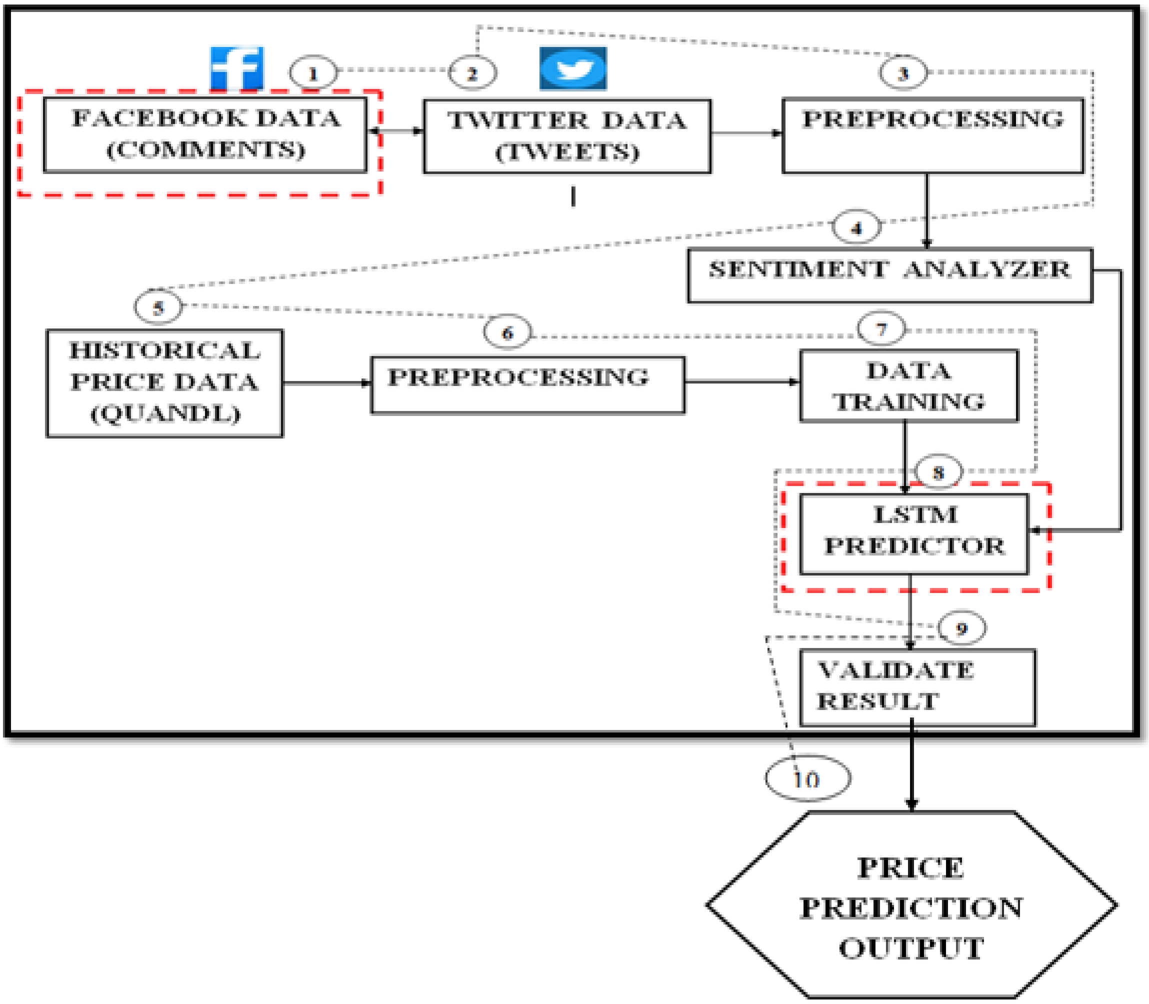

Fig.2. Architecture of the Enhanced Cryptocurrency Prediction Model Using Sentiment Analysis and Machine Learning (Proposed System)

learning models such as the Support Vector machines, ARIMA and the Recurrent Neural network (RNN), and came out as the best approach for prediction. The LSTM model will be used the historical price data collected from the sources mentioned above in order to produce an accurate close to accurate prediction result. The LSTM will also be used during the sentiment analysis to classify the sentiment as positive, negative and neutral based on a given score.

3.2.1. Advantages of the Proposed System

i) A user friendly model that makes it easy for traders to conveniently forecast the price and the fluctuation direction of the Bitcoin, in order to make wise investment decisions.
ii) This model is fast and saves time required to forecast the price of the Bitcoin.
iii) The prediction results obtained from this model are correct due to the use of correct datasets for the training of the system. The prediction result is also easy to understand by the users through the use of simple pictorial representation.
iv) The fluctuations or inconsistence of the Bitcoin price data does not affect the efficiency of this model. This means that no matter the direction of the price data the model will still accurately predict the price.
v) Since the data for the System training were legally collected, governmental laws does not hinder or affect the accuracy of the model.

Facebook Platform: The Facebook platform is an additional data source which was employed because of the number of active users of the application. Facebook is an American Online Social Media and Networking Service which can be accessed from the internet through internet connectivity and with almost every gadget (phones, laptops, pads, etc.)

LSTM Predictor: The Long Short Term Memory is a supervised machine learning model that is used for both sentiment analysis (Used for classification) and time series analysis (used to train the history price data). This makes it a very efficient tool to be used for prediction of Bitcoin prices.

3.3. ALGORITHM

STEP ONE:

IMPORT MODULES (Pandas, Keras, Numpy, Utils, Matploetc)

STEP TWO:

INITIALIZE WEB SCRAPPING TOOL FOR TWITTER AND FACEBOOK (For twitter use TextBlob, for facebook use utils)

STEP THREE:

CARRY OUT TEXT PREPROCESSING (clean up the data gotten from TWITTER, FACEBOOK and QUADL)

STEP FOUR:

CARRY OUT SENTIMENT ANALYSIS USING THE TWITTER/FACEBOOK DATA

STEP FIVE:

CARRY OUT TIME SERIES EVALUATION.

STEP SIX:

TRAIN THE DATA USING LONG SHORT MEMORY (LSTM)

STEP SEVEN:

MERGE THE DATA GOTTEN FROM THE SOCIAL MEDIA PLATFORMS AND THE PRICE DATA (QUANDL).

STEP EIGHT:

MAKE A PREDICTION AND PLOT TE CHART TO SHOW THE PRECTED PRICE AND THE ACTUAL PRICE USING MATPLOT.

IV. RESULTS AND DISCUSSION

Quandl dataset was used as the source of the data. Quandl contains several historical and price data ranging from Stock to Forex, Crypocurrencies etc. It was chosen because it contains the most current and accurate price data due to the fact that it is updated regularly. The data was downloaded from the Quandl website for the Bitcoin price data from 2018 to November 11th 2019.

The Prediction will be for the price of the Bitcoin in the future, the dataset was also collected on an hourly basis in order for the data training process to be efficient. The price data however contained empty price cells, this means that some the price data for some hours were not recorded or there was no price increase or decrease for that hour. This data was cleaned up and the empty data rows was removed and the accurate time and date format that will allow for the accurate data training, processing and prediction.

Date_Time	Price
10/30/2019 0 :00 : 00	6123 .21
10/30/2019 1 :00 : 00	6131 .35
10/30/2019 2 :00 : 00	6114 .17
10/30/2019 3 :00 : 00	6153 .11
10/30/2019 4 :00 : 00	6151 .09
10/30/20195 :00 : 00	…………
10/30/20197 :00 : 00	…………
10/30/20198 :00 : 00	6151 .23
10/30/20199 :00 : 00	………....

Fig.3. Bitcoin Price data from Quandl showing the First Ten (5) Rows

Date_Time	Price
2019-10-30 00.00.00	6123 .21
2019-10-30 01.00.00	6131 .35
2019-10-30 02.00.00	6114 .17
2019-10-30 03.00.00	6153 .11
2019-10-30 04.00.00	6151 .09
2019-10-30 08.00.00	6151 .23

Fig.4. Cleaned up Bitcoin Price data

Date time	Tweets
Thu Nov 09 17:43:40	RT @Forbes: The failure of the accurate pred
Thu Nov 09 17:43:41	RT @Mindstatex: Lots of love from the Bit
Thu Nov 09 17:43:42	RT @Fernando Huama: warning built in ru

Fig. 5 Tweets from Twitter on Bitcoin

Fig.5 shows the Data scrapped from the twitter platform. The data from twitter contains some symbols "@", "RT", "#", "url", "hash tags". The data contains one positive, one negative and one neutral tweet. These symbols cannot be processed by the sentiment analyzer VADER; therefore the tweets were cleaned up using the twitter preprocessor. Table 4.4 shows the data from twitter which has been processed and the redundant data removed in order to produce a suitable data set for sentiment analysis.

Date time	Tweets
2019-11-09 17:43:40	Forbes: The failure of the accurate prediction....
2019-11-09 17:43:41	Mindstatex: Lots of love from the Bitcoin Co...
2019-11-09 17:43:42	Fernando Huama: warning built in rules can...

Fig.6 Cleaned up twitter Data

Out[3]:

	DateTime	Sentiment_score	Price
0	2017-10-31T05:00:00.000-04:00	0.092263	6158.76
1	2017-10-31T06:00:00.000-04:00	0.103770	6105.90
2	2017-10-31T07:00:00.000-04:00	0.111691	6094.36
3	2017-10-31T08:00:00.000-04:00	0.082134	6125.13
4	2017-10-31T09:00:00.000-04:00	0.089370	6165.00

Fig.7. Final Sentiment analysis output merged with Bitcoin Price Data

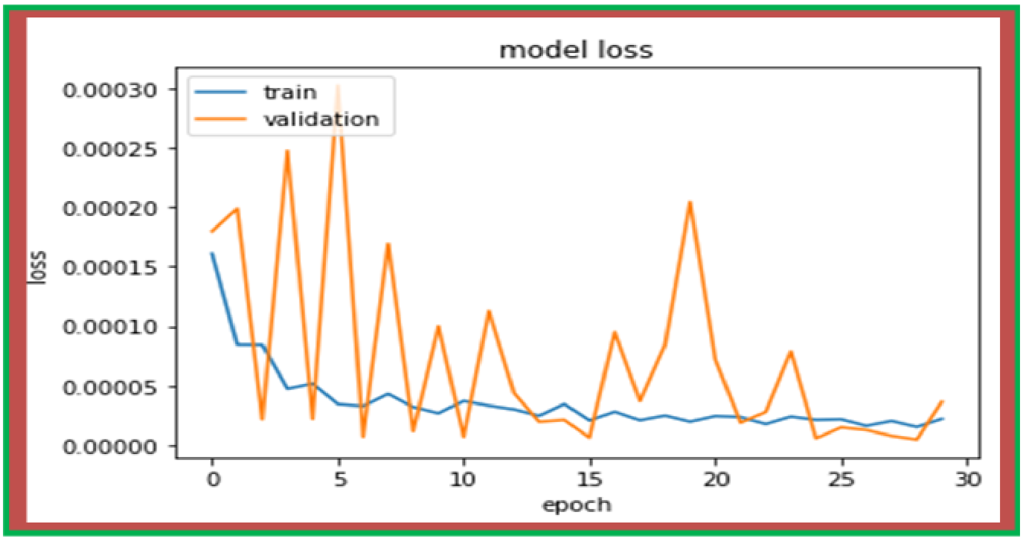

Fig.8. Data training Output

The training sample had a size of 1150 samples and only 290 were validated. The time loss and the value loss were computed and the chart was plotted using this parameters. From the chart above, it could be observed that the data training has a great degree of accuracy because the training curve has minimal losses as the training proceeds, and the validation of the training is high. The history data was used to train the system because we are using a supervised learning method where a system is trained with past experiences or data to enable it make or take correct future actions. After a comparison using some parameters such as the time efficiency, cross platform adaptability, cost efficiency, between the existing system and the proposed system, the overall performance of the proposed system was higher than that of existing system. The overall performance evaluation score of the existing system was 68% and that of the proposed system is 90%, which makes it a better prediction model than the existing system.

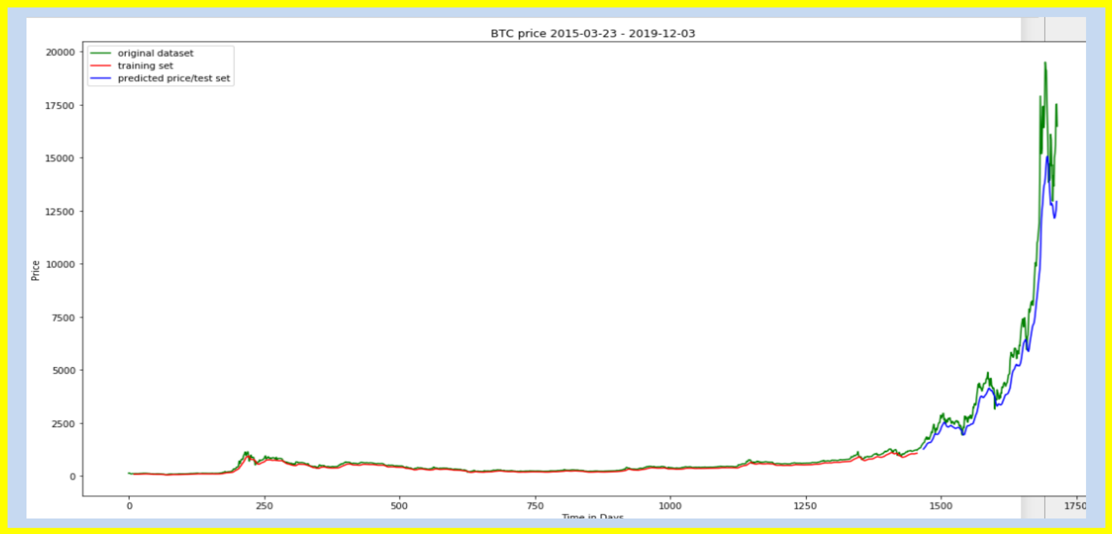

Fig.9. Bitcoin Prediction output

The final prediction result in Fig.9 shows the future price of Bitcoin curve and the actual price curve. The blue curve represents the forecasted price and the green line represents the actual price. It can also be observed that the forecasted price is very close to the actual price. It represents more than 85% prediction accuracy, which is better result than other predictions previously carried out. Matplot was used to plot this chart. Matplot is a library in Python which is used to plot graphs and charts. The Y axis represents the price (feature) of the Bitcoin, and the X axis represents the time interval in hours. The actual price curve is represented by the "inv_y", and the forecast curve is represented with the "inv_yhat".

Table 1: Existing system Performance Analysis (68%)

S/N	PARAMETERS	ASSESSED PERFORMANCE RATE (%)
1	Speed (S)	10
2	Time Saving (TS)	14
3	Cost Benefit Analysis (CBA)	12
4	Cross Platform Adaptability (CPA)	17
5	Model Efficiency (ME)	15

Table 2: Proposed system Performance Analysis (90%)

S/N	PARAMETERS	ASSESSED PERFORMANCE RATE (%)
1	Speed (S)	15
2	Time Saving (TS)	18
3	Cost Benefit Analysis (CBA)	15
4	Cross Platform Adaptability (CPA)	23
5	Model Efficiency (ME)	19

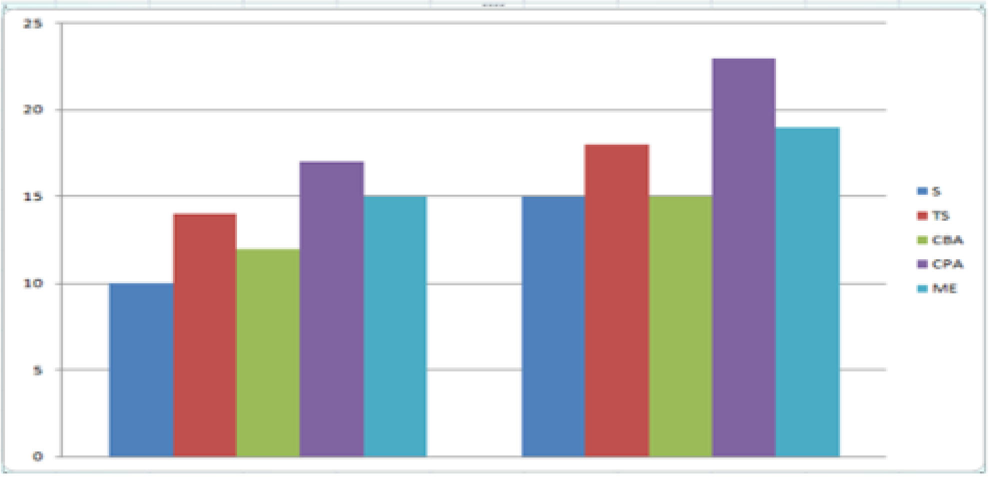

Fig.10 Performance Evaluation Chart of the Proposed System against the Existing System

V. CONCLUSION AND FUTURE WORK

The study concluded on the need to strengthen Price forecast through the use of more efficient tools to bring about a better prediction result. Bitcoin is not a new term; it has become a very major research topic in various sectors such as academia, finance, trade and industry, etc. The study also recorded on of the highest percentage of accuracy in its prediction, from those which has been carried out by other researchers. This work has also solved the problem of inefficiency of the prediction models, by developing a better model which predicts prices that are applicable to the current market trend. This work has recorded a high level of correctness and fulfillment; however, there is still room for improvements in the future for better results and use. One of the areas we will improve on is the use of an even larger data set and more social media platforms, and the development of a model that can predict more Cryptocurrencies, e.g. Etherium, B-money etc. using the hybrid model.

REFERENCES

[1] S. Jones, Feedforward Machine Learning Systems to Improve Cryptocurrency Pricing Models, Washington University in St. Louis, 1-26, 2017.
[2] L. Alessandretti, A. ElBahrawy, L. M. Aiello, and A.Baronchelli, Anticipating Cryptocurrency Prices Using Machine Learning, 1-17, https://doi.org/10.1155/2018/8983590. 2018.
[3] K. Chhatwani, Predicting Bitcoin Price Fluctuations by Analyzing Global Currency Patterns and Sentiments, 1-5, 2019.
[4] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning, Entropy 2019, 21, 589; doi:10.3390/e21060589, 1-12. 2019.
[5] B. Spiklak, Deep Neural Networks forCryptocurrencies Price Prediction, ISSN: 586460, 1-67, 2018.
[6] S. Gullapalli, Learning to predict cryptocurrency price using artificial neural network models of time series, 1-37, 2016.
[7] C. Bai, T. White, L. Xiao, V.S. Subrahmanian, and Z. Zhou, C2P2: A Collective Cryptocurrency Up/Down Price Prediction Engine, 1-6. 2019.
[8] A. Kaushik, and S. Naithani, A Study on Sentiment Analysis: Methods and Tools, *International Journal Of Science and Research*, 4(12), 2015, 287-292.
[9] D. Olvera-Juarez and E. Huerta-Manzanilla, Forecasting bitcoin pricing with hybrid models: A review of the literature, *International Journal of Advanced Engineering Research and Science*, 6(9), 2019, 161-164.
[10] A. Jain, S. Tripathy, H. Dharwivedi, P. Saxena, Forecasting The Price of Cryptocurrency Using Sentiment Analysis, 11th int. Conf. Contemp. Comput. Noida, 208, pp.1-7. Doi:10.1109/IC3.2018.8530659.
[11] J. Bernadette, and R. Shirakawa, Cryptocurrency Regulations: Institute and Financial openness, ABDI Working Paper Series. [Online]. No.978. https://www.adb.org/publications/cryptocurrency-regulations-institutions-financial-openness, 2019.

Onuodu Friday Eleonu,etal. "An Enhanced Cryptocurrency Prediction Model Using Sentiment Analysis and Machine Learning." *International Journal of Engineering Science Invention (IJESI)*, Vol. 09(04), 2020, PP 14-24.