International Journal of Scientific & Engineering Research Volume 11, Issue 3, March-2020
1542
ISSN 2229-5518

Crypto Currency Market Prediction

Jyotirmai Kaushal , Devansh Sharma , A. Meena Priyadharsini

Abstract-The realm of cryptographic money has become exponentially over the previous decade, with the most quick advances found in the previous hardly any years as an ever increasing number of gatherings around the globe perceive the benefit of holding computerized resources on the web. Insights bolster this explanation where, around 1,500 quests about Bitcoin alone is recorded every hour. Therefore, numerous individuals are starting to turn out to be increasingly mindful and tolerating of the idea of advanced monetary standards, and merchants specifically try to know how they can make gainful crypto-coin exchanges and speculations. Albeit various research ventures have been attempted to create frameworks that can viably anticipate value developments in the digital money showcase, they show huge productivity holes, which this paper further investigates. The creators at that point endeavor to gain from past examinations and develop an increasingly all encompassing way to deal with a prescient cost model for the digital money advertise. This spotlights on evaluating key factors that influence the instability of the market – open discernment, exchanging information, noteworthy cost information, and the interdependencies among Bitcoin and Altcoins - and how they can be best used from a mechanical viewpoint by applying opinion investigation and AI methods, to build the proficiency of the procedure.

Index Terms: ANN , Cryptocurrency , Data Exchange , LSTM , Machine Learning , Price Prediction , SVM.

Catchphrases AI, prescient models, conclusion examination.

I. Introduction

In the present profoundly marketed worldwide scene, the interest for a progressively open and straightforward vehicle of cash has developed quickly . As the world pushes ahead with new progressions in innovation, so too has the domain of money related trade advanced with the idea of computerized cash (or digital forms of money) following the arrival of the principal virtual money in 2009, Bitcoin. The idea of digital currency centers around having quicker and progressively secure money related exchanges on the web. The innovation used to empower this is known as the Blockchain, where there is no focal gathering to check exchanges and rather the entire framework is decentralized, henceforth making it significantly more secure. This demonstrates altogether more beneficial than the present framework used to check and lead fiat cash exchanges, where the capacity to effortlessly mishandle their utilization in exchanges has prompted probably the greatest money related embarrassments of the century. One such model is the Libor Scandal of 2016 where banks controlled loan costs for greater overall revenues.

When taking a gander at the development and accomplishment of computerized monetary forms, information from the world's first digital currency overview makes four primary inferences. In the first place, cryptographic money showcase capitalization has expanded fundamentally because of the becoming stronger and acknowledgment of crypto-coins as a computerized resource; the market top lies at USD 592 billion with the Bitcoin advertise cost at over USD 17,000 . Second, since 2009 Bitcoin has remained the market head for cypto-coins, with a predominance of over 55% in the cryptographic money advertise (which comprises of an aggregate of 1360 computerized coins) (as of December 2017); this bears witness to its solid impact over the conduct of Altcoins. Third, out of every single invested individual (merchants, diggers and financial specialists), brokers are the biggest partner bunch who draw in the most with the cryptographic money market to benefit from purchasing and selling advanced resources in online digital money trades. What's more, fourth, the extraordinary unpredictability of the market makes it dangerous for merchants to hold or exchange their benefits beneficially. Every one of these perceptions delineate the dynamism of the market and feature the requirement at a digital currency cost expectation framework.

II. LIMITATIONS OF CURRENT STUDIES

Attempts thus far to construct such a model has faced significant limitations. The biggest limitation is the fact that current research has been heavily restricted to a few more popular crypto-coins in the market - Bitcoin, the market leader, closely followed by Ethereum, Dash, Monero, Ripple and Litecoin. But as of today, 1360 other Altcoins exist in the market (as of December 2017). At the same time, the cryptocurrency market is highly unstable and experiences periods of extreme volatility which often makes it difficult to predict behavioral patterns. Past studies only take into consideration one or two market variables when attempting to predict the price of crypto-coins, failing to account for all factors that may affect the market. These two observations result in predictive models with limited accuracy, and the problem of limited access to information in the global cryptocurrency market continues to persist.

The predictive model suggested by the authors aims to be more holistic in nature. It takes into consideration multiple factors affecting the market, and applies a range of technological methodologies, tools and techniques, in order to provide an accurate prediction so that users will be able to better benefit through investing, trading or mining cryptocurrencies more effectively.

III.Factors Used Predict The Price of Cryptocurrencies

A. Open Perception

As examined in the past areas, individuals are starting to comprehend and utilize digital currencies all the more much of the time. The measure of prattle online has additionally expanded fundamentally in the course of recent years, offering ascend to the expanded fame of the digital currency advertise. In this way, a fascinating connection between Bitcoin costs and the measure of news online has been distinguished in examine led by D'Alfonso et al. (2016). The examination's discoveries clarified that the expanded measure of talk online considerably affected the cost; right now, talk about Bitcoin lead to an expansion in its value . To additionally affirm this point, Google patterns show that for Etherum, Bitcoin's greatest rival, there is a very high corelationship between its expansion in cost and the measure of google look through online.

B. Exchanging Data

With the development of cryptographic forms of money and the overall population beginning to acknowledge its advantages, numerous online trades have been set up so individuals everywhere throughout the world can put resources into digital forms of money or exchange their computerized resources request to increase a benefit. With time the quantity of clients on trades has expanded colossally. With this expansion, partners show a solid enthusiasm for exchanging data as this data can be utilized to settle on educated choices about exchanges or ventures. Because of the interest for data, mainstream trades, for example, Bittrex and Poloniex now open information to their clients through API's. This is an unmistakable pointer that immediate partners are keen on advanced data about the conduct of computerized monetary standards, and consequently adds to the requirement at a cost expectation framework.

C. Noteworthy Price Data

In the realm of securities exchanges, individuals use charts of noteworthy value information to distinguish social patterns and examples of monetary forms so as to attempt to foresee their future cost. Research completed by Patel et al. (2015) study the systems used to distinguish value examples to make value expectations of a couple of chosen organizations on the Indian Stock Market . Along these lines, designs have additionally been distinguished on cryptographic money value charts which proposes this could likewise be an important credit to consider while foreseeing the cost of digital currencies. The memorable costs of cryptographic forms of money can be acquired through online trades as clarified in the past area

D. Interdependencies among Bitcoin and Altcoins

Research did by Ciaian et al. (2016) endeavor to recognize factors that can make an interdependency among Bitcoin and Altcoins. Because of Bitcoin's strength (55%) in the market, it is being exchanged off to purchase Altcoins. To additionally expand on this point, 68% of all Altcoins are obtained through Bitcoin while just 14% of Altcoins are purchased straightforwardly through the US dollar. Another factor that is considered is the similar price developments of Bitcoin on Altcoins. This is because the vast majority of Altcoins are largely clones of Bitcoin with minor changes in parameter values (different block times, currency supplies or issuance schemes). A visual inspection of Bitcoin and Altcoin price indices show a similar price drop in both cryptocoins in the same period of time; this trend is again observed when Altcoins follow an increase in Bitcoin prices. The research states that the similar behavioral patterns could be as a result of the fact that Altcoins are following the price of the market leader, Bitcoin.

IV. Areas of Sentiment Analysis Applicability

Sentiment analysis, also known as “opinion mining”, is the process of digitally understanding the polarity of text. Polarity in this context refers to the ability to understand if the specific text expresses a positive or negative opinion. Determining this fact allows a system to process what a particular individual’s sentiment is. Therefore, this technique can be used to understand what kind of opinion people have with regard to a certain topic, given that the data provided to be analyzed is of the same topic.

The world we live in today is highly globalized, and people of all ages have access to social media which they use to keep in touch with others, as well as to express their opinions on specific topics of interest. The increase in use of social media has brought in many social media networks such as Facebook, the largest social media network where there are over 1,200 million active users per day, and Twitter, which produces over 500 million tweets per day. These statistics prove that there is a great deal of information which can be used to in order understand what exactly people feel about a particular topic. Taking a closer look at the statistics from Twitter, around 1500 tweets about Bitcoin has been recorded per hour. The massive number of posts/tweets, and the number of users they come from, makes social media a great data source from which to gather primary information about Bitcoin and Altcoins, which the proposed system can then use to better understand through the application of sentiment analysis.

V. AREAS OF MACHINE LEARNING APPLICABILITY

Machine learning is a method of getting systems to act in a certain manner without explicitly programming it to do so. This technique is used in areas like self-driving cars, speech recognition and also prediction systems like stock prediction systems. Having said that, machine learning will have to be applied in multiple areas and using a variety of techniques in order for this approach to be a success. This section discusses and justifies how machine learning techniques will be applied to make the predictive model a success.

A. AI Applied to Trading Information

Studies show that exchanging data is a dependable factor that can be utilized for expectations, particularly when applying AI as there are existing connections previously found inside exchanging data, for example, between the opening value, shutting cost and exchanging volumes. AI strategies get on these connections and base their forecasts out of these components. Greaves and Au (2015) utilized exchanging data from online trades to anticipate the cost of Bitcoin. At first the examination depended on making expectations utilizing data from the blockchain itself. Having no achievement right now study proceeds by concentrating on exchanging data. The investigation explains on a fascinating relationship where the cost increments when Bitcoin is sold for more than it is purchased at. The legitimization for this influence is that the cost of Bitcoin increments because of an expansion sought after for the coin .

B. AI Applied to Historic Price Data

By utilizing notable value information, AI calculations can endeavor to recognize value designs. This would enable a prescient framework to foresee costs all the more precisely. A comparable strategy has been utilized by Patel et al. (2015) to foresee cost lists for four organizations recorded on the Indian Stock Market. So as to pick up results with high exactness, the framework they structured utilized ten specialized parameters in the informational collection which was utilized to prepare and anticipate costs. To additionally improve precision different AI calculations were utilized. This examination utilizes a pattern deterministic information planning layer which controls constant information to discrete qualities (+1 or - 1), as per the outcomes show an expansion in precision. Aftereffects of the test show that the calculation Naïve Bayes played out the best with a precision of 90.19%. An examination completed by Rowland (2014) additionally utilized this way to deal with anticipate the cost of Bitcoin as it were. This framework utilized a group casting a ballot framework together with two AI calculations which ran synchronously so as to augment on the precision of the forecasts. The group learning approach creates different classifiers which are later joined. This methodology attempts to upgrade precision by killing mistakes produced by a solitary classifier.

C. AI Applied to the Data Results of Sentiment Analysis

Deciding the feeling of client posts alone isn't sufficient to think of a genuine value forecast. For this to be done viably, AI methods should be utilized. Initial, a check of negative and positive tweets from the applicable timespan will be gotten from the database. The quantity of negative and positive tweets goes about as the informational index which will permit the AI program to comprehend the assessment of the clients. Utilizing this data, the framework would now be able to be prepared to foresee the price change which will then allow us to make a prediction on the future price of cryptocurrencies.

VI.MACHINE LEARNING TECHNIQUES

Machine learning technologies are being constantly developed over time to bring many new algorithms to life. While some algorithms can be used to produce results with high accurate levels when addressing one type of problem, they can also produce a low accuracy when attempting to address a problem of a different nature. Hence it is important to first assess the properties, strengths and weaknesses of different available algorithms in order to understand and determine which algorithm forms the best fit for this system.

A. Artificial Neural Networks

Artificial Neural Network (ANN) is one of the most commonly used machine learning techniques. Its structure is based on two layers called the input and output layers. Apart from these layers, an unknown number of nodes can exist in a layer known as the hidden layer. The hidden layer enables the algorithm to have a strong learning capability. One of the biggest advantages this algorithm holds is its capability to have a finite number of layers and yet still be able to accurately predict values for any continuous function. This property is referred to as the universal approximator [4]. Due to this property, the algorithm also faces an issue know as overfitting. This is when the algorithm detects a considerable amount of non-existent relationships which could lead to a major drop in accuracy .

B. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model. The algorithm defines classes by separating data using clear boundaries while considering errors caused

ISSN 2229-5518

during the process. In certain real-world situations where boundaries need to be defined in a more complex manner, an implementation called the kernel implementation, also known as the non-linear kernel, is used . This algorithm has the ability to automatically find optimum parameters for high levels of accuracy . However, it must be noted that this algorithm might use a very high amount of computational power, as well as time, to produce results.

C. Random Forest

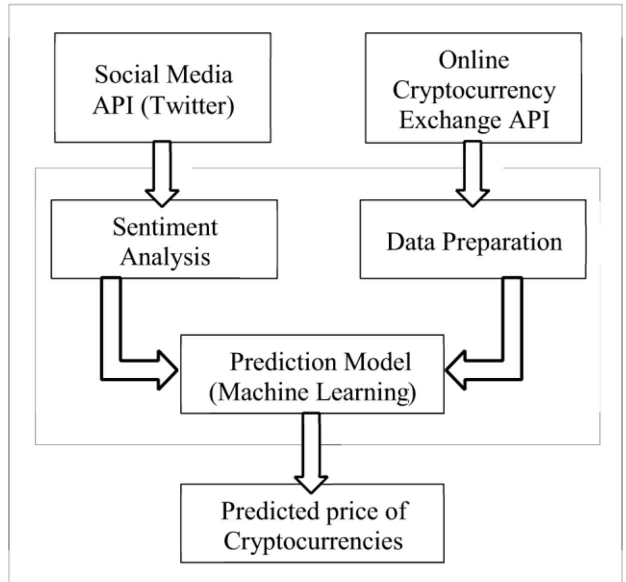
This algorithm is based on the logic of decision trees. The algorithm generates 'n' number of trees, thereafter each node selects three random features. When the data is analyzed multiple outputs are generated, and the final output is chosen by selecting the output that went through the most number of trees . This algorithm is fairly easy to implement as it does not require too much fine-tuning in order to reach high limits of accuracy. However, this technique uses up a lot of memory depending on the classification at hand.

D. Naïve Bayes

Naïve Bayes is a simple and efficient algorithm. It is based on the assumption that any one of the attribute identified is completely unrelated to each other attribute; this gives the algorithm the ability to ignore irrelevant attributes. Although this assumption does not hold true in all real-world situations, it has been proven to reach high accuracy levels in certain realworld scenarios. As each feature is defined into the prediction model there are no issues in scaling it, which is extremely advantageous. However, if a new feature needs to be considered then a new prediction model needs to be built from scratch .

VII.THE PROPOSED SYSTEM

In the previous sections of this paper the methods and techniques of the proposed system have been discussed. The proposed system aims to capture the main factors that affect the price of cryptocurrencies, and employ various machine learning techniques on the data, in order to make an accurate prediction of the price of cryptocurrencies. A high-level diagram of the proposed system is depicted below.



1: High Level Diagram of the Proposed System.

A. Data Capture and Data Preparation

One of the factors that affect the price of cryptocurrencies is public perception of cryptocurrency stakeholders on the health of the market. This factor is captured through social media API's, Twitter to be precise. As Twitter's API exposes all tweets based on a certain topic, only the needed tweets can be filtered out and gathered. Thereafter, sentiment analysis is performed in order understand the public perception of Twitter users who identify with the concept of cryptocurrency. Unfortunately, this technique cannot be applied to all cryptocurrencies as most online chatter is limited to a few more popular coins like Bitcoin, Etherum and Litecoin. Due to this reason, there is a confinement where adequate information on different digital forms of money can't be accumulated to make a precise enough expectation. So as to conquer this issue, opinion examination will be applied uniquely to tweets identified with Bitcoin. This is additionally expounded underneath in Section B 'Forecast of Bitcoin Prices'. When the slant of a tweet is resolved, the tweet will be spared in the framework's database alongside its conclusion with the goal that this information can be effectively reused for additional exploration purposes later on.
The information expected to catch the elements of exchanging data, noteworthy costs and data identified with the value interdependency among Bitcoin and Altcoins can be gotten through online trades. At the point when the information is gathered, it will be balanced with the goal that the framework can utilize this data to prepare AI calculations which will at that point foresee costs into what's to come. The balanced informational index will likewise be spared in the framework's database as this data can be utilized to prepare the framework again in future to adjust to new patterns or exam-

ples. The upkeep of information will be a booked procedure so the framework can be as modern as could reasonably be expected.

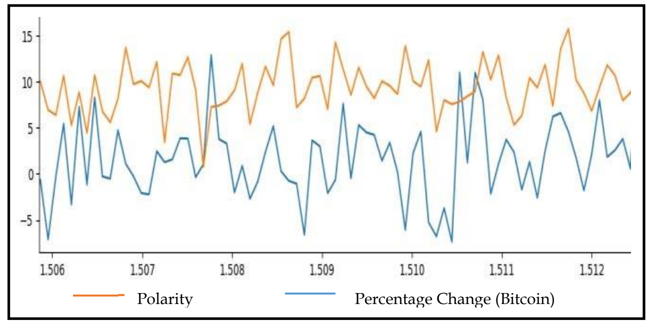
B. Forecast of Bitcoin Prices

So as to foresee the cost of Bitcoin three variables are considered; the part of open observation, as talked about beforehand, will be caught through Twitter API's and the database will contain the tweets and their individual assessments. The key data that is required will be a tally of negative and positive presents mapped on its time succession. The other two factors that will be considered for this expectation is the exchanging data and the memorable cost of cryptocoins. The dataset will contain qualities, for example, the rate change and exchanging volume, which can be gotten from the exchanging data when at first spared in the database. The dataset ought to likewise have information from a large enough timespan with the goal that the AI calculations will have the option to distinguish the value designs. When the informational index comprises of all the fundamental information, a chose AI method can be applied so as to acquire a precise expectation.

C. Expectation of Altcoin Prices

Perhaps the most concerning issue looked in past investigations is the confinement of information online about more up to date or less well known digital forms of money. In spite of the fact that this factor can't be legitimately caught here, data dependent on the expectation of Bitcoin, for example, the rate change of the new value forecast, can be utilized to say something regarding the value forecast of Altcoins. Considering data about the adjustment in cost of Bitcoin while foreseeing the costs of Altcoins is useful as there is an interdependency between the value variances of Bitcoin and that of Altcoins. Along these lines, including this information will no doubt make another relationship which will thus help increment the precision of the expectation. Thus, the interdependency among Bitcoin and Altcoins would then be able to be utilized to anticipate the costs of Altcoins and hence goes about as a substitute for the factor of open recognition which can't be utilized for altcoins.D. Initial Evaluation and Testing

To make sure that the captured tweets are useful and that they have an effect on the market, the polarity and the actual percentage change of the price change in Bitcoin was compared repetitively while applying various filters. 2 represents a comparison of the polarity and the percentage change of price in Bitcoin for period of time with high volatility on a daily basis.



2: Comparison between polarity and percentage change in Bitcoin

When evaluating the above , it is clear that in some instances there is a clear relationship whereas in other instances this relationship isn’t as clearly visible. Observations also show there is a delayed reaction to a spike or drop in polarity, which suggests that in certain periods the market takes time to respond to the online chatter. As this system is employing a neural network, its learning capability helps it to reduce the impact on accuracy due to these factors. A fact that should be remembered is that the final prediction of the system is based on the frequency of tweets and trading information in addition to the feature discussed above.

Three main coins have been used in order to test the accuracy of the system. The data from predictions made on the past 3 months were considered for each coin. The table below summarizes the reasons as to why these specific coins were used to testing purposes, as well as their relevant accuracy levels.

TABLE 1: REASONS AND ACCURACY OF CRYPTOCURRENCY USED FOR TESTING

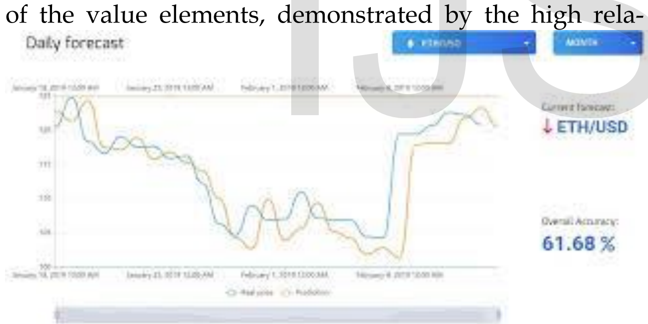
Cryptocurrency	Reason	Accuracy
Bitcoin	The most popular cryptocurrency. This coin is needed in order for other coin predictions to happen.	85%
Ethereum	An altcoin which is comparatively popular and Bitcoins biggest competitor. Used to test how the system will handle altcoins.	93.33%
Bitcoin Cash	A fairly new cryptocurrency. Used to monitor how the system will handle coins with small data sets	70%

Results show that that predictions of Ethereum has the highest accuracy which can be a result of including the predicted percentage price change of Bitcoin as one of the factors in the predictive model. Bitcoin also has a comparatively high accuracy. Overall, the predictive model did perform better in time periods of less volatile market fluctuations. Bitcoin cash, on the other hand, displayed the lowest accuracy level of 70%, which is reflective of the limited size of the dataset.

VIII. Results and Discussion

A. ANN Estimate of time Series Memory

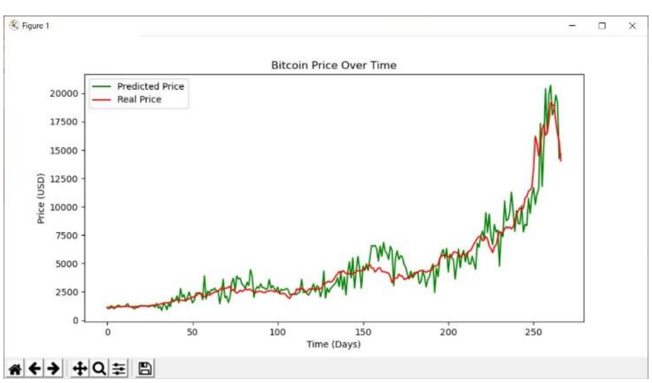
For the initial segment of ANN model, we utilize an ANN model to anticipate the cost of Bitcoin one day into the future utilizing five unique lengths of memory: 7, 14, 21, 30 and 60 days. To quantify the blunder between the information and model, we utilize the mean square mistake and the connection. The consequences of our demonstrating tests are indicated in Fig. 2. It shows that the cost of cryptographic forms of money displays a long haul self-clarify include. By learning the full history of the earlier month, the ANN model forecast of Ethereum is to a great extent improved, contrasted and those momentary cases (blue bars in Fig. 2). Besides, concerning Bitcoin and Ripple, a significantly longer history of value (60 days, green and orange bars in 2) is gainful. Be that as it may, we additionally find that exclusively increment the length of verifiable information as information highlights not really initiate a superior model, the model execution is balanced by presenting increasingly model parameters (as increment of information highlights). For instance, for Ethereum, utilizing 60-day value history as info highlights performs more regrettable than utilizing 30-day value history. In any case, all ANN models for the most part catch the variety of the value elements, demonstrated by the high rela-



ANN Model price prediction graph

also observed that the prices in three days of Ethereum could be forecasted more accurate than its other prices in the future.

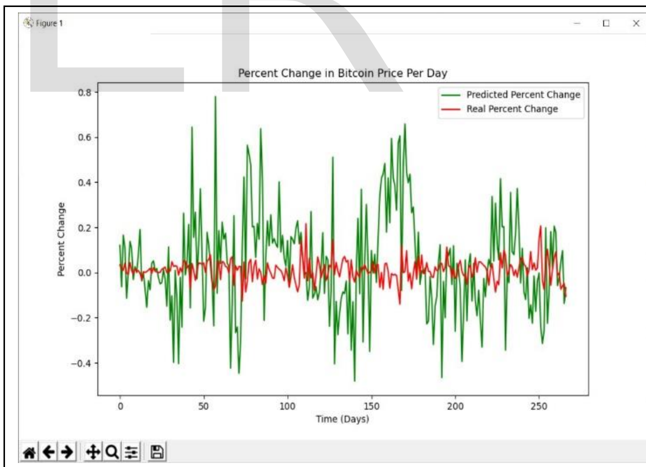
B. LSTM Price prediction model



LSTM Model Price prediction Graph

Concerning the LSTM model, it has a similar exhibition with the ANN model by and large, while foreseeing the one-day future costs of these cryptographic forms of money, in view of mean square blunder. It exhibits that in spite of the fact that ANN is absence of inward ability, it could successfully concentrate and utilize the valuable data covered up in the authentic value elements to anticipate a future cost. While LSTM is purposefully intended to demonstrate the inner memory stream and its effect on future expectation, subsequently, both ANN and LSTM are reasonable at the digital forms of money cost time arrangement forecast.

We likewise discover that LSTM required the length of value history is unique in relation to that of ANN. LSTM by and large incline toward short authentic memory. For instance, LSTM with seven days of chronicled memory for Ethereum and Ripple or 14 days of recorded memory for Bitcoin play out the best. The model-information relationship strongly decays as the length of verifiable memory increment (Fig. 4 right board). It shows that LSTM depends the model expectation more on the latest scarcely any days.



5. Performance of LSTM model, given 50 days of predictive length. Left and right panels represent model-data mean square error and Pearson correlation.

C. Limitation and Future Work

This study is limited from several perspectives, which will be improved in our future studies. The first one is that we only choose three of the most representative digital currencies to analyse their price dynamics so that the result may be not generalised enough and we should apply the experiment to other cryptocurrencies like Litecoin, Tether, and Stellar.

Furthermore, five different lengths of memory have been used to predict the price of the digital currencies in one day because of the limitation of computational resources, so the best predictive memory length maybe cannot be found exactly, and the 30-day historical price data are probably more useful for forecasting other lengths of future price.

Another limitation is the process of optimization and the parameters of our model are not being tested very well as the primary purpose of this experiment is to find out the reasonable historical memory length and predictive memory length, but it would be better if we could choose parameters based on grid search and try more different structural ANN and LSTM models.

IX. CONCLUSION

This investigation has distinguished and examined how extraordinary market elements can be utilized trying to make exact expectations on digital currency costs. The systems and advances that will permit these variables to be caught and controlled so as to make such forecasts has additionally been investigated. This arrangement expects to address the conduct of, and subsequently catch the estimation of, all cryptographic forms of money in the market as opposed to concentrating exclusively on a couple of the more mainstream advanced coins, while extricating from and considering data from the wide measure of information accessible.

ACKNOWLEDGMENT

My heartfelt thanks goes out to my supervisor and mentor Mrs. A.Meena Priyadharsini for all the constant guidance provided. I would also like to thank my family for the endless motivation and invaluable support provided.

REFERENCES

[1] 30 bitcoin hashtags popular on Twitter | RiteTag: Find the best hashtags: https://ritetag.com/best-hashtags-for/bitcoin. Accessed: 2017-12-20.

[2] Alexander D'Alfonso, Peter Langer, Z.V. 2016. The Future of Cryptocurrency. International Security. 7, 5 (2016), 7–45.

[3] Anantwar, S.G. and Shelke, R.R. 2012. Simplified Approach of ANN : Strengths and Weakness. International Journal of Engineering and Innovative Technology (IJEIT). 1, 4 (2012), 73–77.

[4] Byvatov, E. et al. 2003. Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification. Journal of Chemical Information and Modeling. 43, 6 (2003), 1882– 1889.

DOI:https://doi.org/10.1021/ci0341161.

[5] Ciaian, P. et al. 2016. Virtual Relationships: Short- and Long-run Evidence from BitCoin and Altcoin Markets. (2016).

[6] CryptoCurrency Market Capitalizations: https://coinmarketcap.com/. Accessed: 2017-07-22.

[7] Definitive portal for social media statistics globally | Socialbakers: https://www.socialbakers.com/statistics/. Accessed: 2017-08-28.

[8] Ethereum - Explore - Google Trends: https://trends.google.com/trends/explore?q=Ethereum. Accessed: 2017-12-17.

[9] Greaves, A. and Au, B. 2015. Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin. (2015).

[10] Joachims, T. 1998. Text categorization with support vector machines: Learning with many relevant features. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 1398, (1998), 137–142. DOI:https://doi.org/10.1007/s13928716.

[11] Kim, Y. Bin et al. 2016. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. PLoS ONE. 11, 8 (2016), 1–17.

DOI:https://doi.org/10.1371/journal.pone.0161197.

[12] Pang, B. and Lee, L. 2008. Opinion Mining and Sentiment Analysis.

[13] Patel, J. et al. 2015. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications. 42, 1 (2015), 259–268.

DOI:https://doi.org/10.1016/j.eswa.2014.07.040.

[14] Rowlands, J. 2014. Bitcoin Algorithmic Trading Bot. April (2014).

[15] Statnikov, A. et al. 2008. A comprehensive comparison of random forests and support vector machines for microarray-based cancer