Crypto Currency Prediction Model using ARIMA

G. Vidyulatha[1], M. Mounika[1], N. Arpitha[1]

[1]Department of CSE, Sree Dattha Institute of Engineering and Science Hyderabad, India.

ABSTRACT

Over the past few years, Bitcoin has been a topic of interest of many, from academic researchers to trade investors. Bitcoin is the first as well as the most popular cryptocurrency till date. Since its launch in 2009, it has become widely popular amongst various kinds of people for its trading system without the need of a third party and due to high volatility of Bitcoin price. Thus, this article presents a suitable model that can predict the market price of Bitcoin best by applying a few statistical analyses. Our work is done on five and half year's bitcoin data from 2015 to 2020 based on time series analysis approachcalled autoregressive integrated moving average (ARIMA) model. Further, it is also compared to existing machine learning algorithm named linear regression (LR) model. Extensive prediction results shown that the proposed ARIMA model acquiredsuperior performance for deciding volatility in weighted costs of bitcoin in the short run as compared to LR model.

Index Terms: Bitcoin, machinelearning, time series analysis, ARIMA, linear regression.

I. INTRODUCTION

Bitcoin [1] is the worlds' most valuable cryptocurrency and is traded on over 40 exchanges worldwide accepting over 30 different currencies. It has a current market capitalization of 9 billion USD according to https://www.blockchain.info/ and sees over 250,000 transactions taking place per day. As a currency, Bitcoin offers a novel opportunity for price prediction due its relatively young age and resulting volatility, which is far greater than that of fiat currencies [2]. It is also unique in relation to traditional fiat currencies in terms of its open nature; no complete data exists regarding cash transactions or money in circulation for fiat currencies. Prediction of mature financial markets such as the stock market has been researched at length [3], [4]. Bitcoin presents an interesting parallel to this as it is a time series prediction problem in a market still in its transient stage.

Traditional time series prediction methods such as Holt-Winters exponential smoothing models rely on linear assumptions and require data that can be broken down into trend, seasonal and noise to be effective [5]. This type of methodology is more suitable for a task such as forecasting sales where seasonal effects are present. Regardless of the substantial vacillations of Bitcoin prices (particularly during 2015 and early 2020) and the massive growth in the capitalization of the related market, the condemnations about illicit uses and social concerns, it has still managed to draw the attention of many investors, such as China who is buying Bitcoin, seeing this as an opportunity of investments [6], as well as researchers in the scientific community to study and understand the market in order to predict the worth of Bitcoin. Most importantly, for the huge popularity of bitcoin, the end of the year 2017 has been the time when the price has increased most noticeably which was worth to 1600 US dollar for 1 bitcoin [7].

Therefore, the analysis of financial data for predicting the future bitcoin price has always been an important field of research with a direct and indirect effect on world economy.Due to the lack of seasonality in the Bitcoin market and its high volatility, these methods are not highly effective for this task. Given the complexity of the task, machine learning makes for an interesting technological solution based on its performance in similar areas. Hence, a time series analysis is utilized in this paper in order to find out the pattern of bitcoin price movement and forecasting the closing price of the next few days as well as analyzing the performance of the time series models i.e., ARIMA model.

II. RELATED WORK

Research on predicting the price of Bitcoin using machine learning algorithms specifically is lacking. [8] implemented a latent source model as developed by [9] to predict the price of Bitcoin noting 89% return in 50 days with a Sharpe ratio of 4.1. There has also been work using text data from social media platforms and other sources to predict Bitcoin prices. [10] investigated sentiment analysis using support vector machines coupled with the frequency of Wikipedia views, and the network hash rate. [11] investigated the relationship between Bitcoin price, tweets, and views for Bitcoin on Google Trends. [12] implemented a similar methodology except instead of predicting Bitcoin price they predicted trading volume using Google Trends views. However, one limitation of such studies is the often small sample size, and propensity for misinformation to spread through various (social) media channels such as Twitter or on message boards such as Reddit, which artificially inflate/deflate prices [13]. In the Bitcoin exchanges liquidity is considerably limited. As a result, the market suffers from a greater risk of manipulation. For this reason, sentiment from social media is not considered further. [14] analyzed the Bitcoin Blockchain to predict the price of Bitcoin using SVMand ANN reporting price direction accuracy of 55% with a regular ANN. They concluded that there was limited predictability in Blockchain data alone. [15] also used Blockchain data, implementing SVM, Random Forests and Binomial GLM (generalized linear model) noting prediction accuracy of over 97% however without cross validating their models limiting the generalizability of their results. Wavelets have also been utilized to predict Bitcoin prices, with [16], [17] noting positive correlations between search engine views, network hash rate and mining difficulty with Bitcoin price. Building on these findings, data from the Blockchain, namely hash rate and difficulty are included in the analysis along with data from the major exchanges provided by CoinDesk.Predicting the price of Bitcoin can be considered analogous to other financial time series prediction tasks such as forex and stock prediction. Several bodies of research have implemented the multi-layer perceptron (MLP) for stock price prediction [4] [18]. However, the MLP only analyses one observation at a time [19]. In contrast, the output from each layer in a recurrent neural network is stored in a context layer to be looped back in with the output from the next layer. In this sense, the network gains a memory of sorts as opposed to the MLP. The length of the network is known as the temporal window length[20] notes that the temporal relationship of the series is explicitly modelled by the internal states contributing significantly to model effectiveness

III. METHODOLOGY

3.1. LR model

The prediction of crypto currencyusing LR model based on the bitcoin datasets on the data and prices as the featurelist are inputs and target list are predicted values. This model is feasible to some extend for the prediction of the crypto currency.

Disadvantages

Using Linear Regression algorithm gives less approximate prediction compared to time series Algorithm in the proposed model in the project. As well the featurelist and targetlist fitted into the algorithm gives less predictions compared to the time series, Comparatively Linear regression performs poorly when there are non-linear relationships. They are not naturally flexible enough to capture more complex patterns and adding the right interaction terms or polynomials can be tricky and time-consuming.

3.2. ARIMA model

To fit ARIMA models to any time series data, the most important condition is that the dataset has to be consistent. In this paper, we have mainly focused on to create a consistent time series dataset and then predict the future Bitcoin closing price according to the nature of previous data. We collected the dataset from CoinDesk [10] which contains daily market capital, volume in transactions, opening and closing price of bitcoin in USD from July 2015 to February 2020.

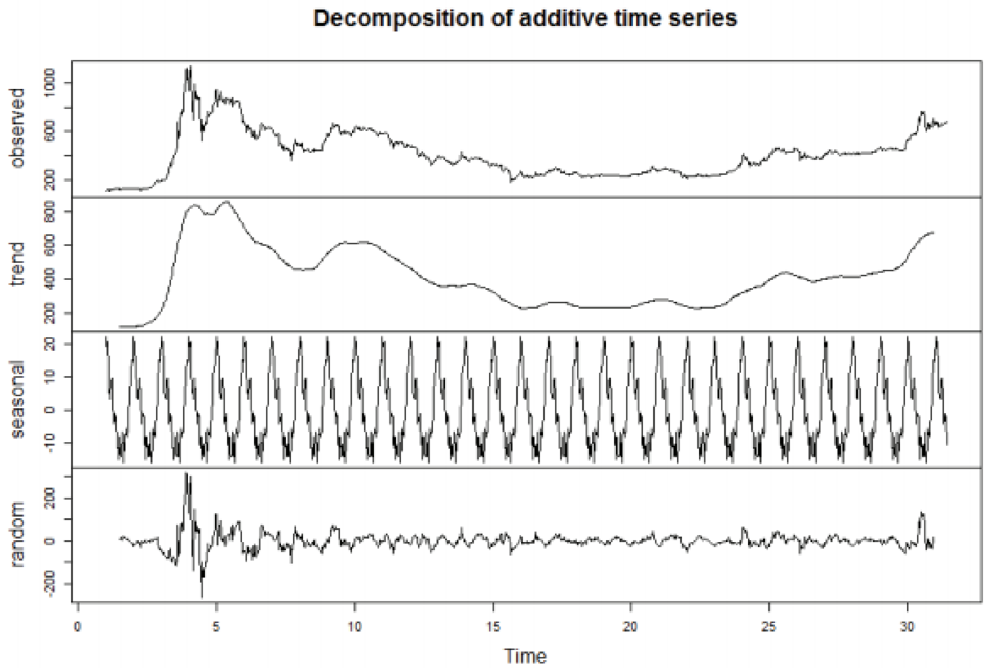

Figure 1: Decomposition of the Bitcoin time series data.

3.3. Feature Engineering and Feature Evaluation

Feature engineering is the art of extracting useful patterns from data to make it easier for machine learning models to perform their predictions. It can be considered one of the most important parts of the data mining process to achieve good results in prediction tasks. Several papers in recent years have included indicators including the simple moving average (SMA) for machine learning classification tasks. An example of an appropriate technical indicator is a SMA recording the average price over the previous x days and is correspondingly included.To evaluate which features to include, Boruta (a wrapper built around the random forest classification algorithm) was used. This is an ensemble method in which classification is performed by voting of multiple classifiers. The algorithm works on a similar principle as the random forest classifier. It adds randomness to the model and collects results from the ensemble of randomized samples to evaluate attributes and provides a clear view on which attributes are important. All features were deemed important to the model based on the random forest, with 5 day and 10 days (via SMA) the highest importance among the tested averages. The de-noised closing price was one of the most important variables also.

IV. IMPLEMENTATION

4.1. Autocorrelation

Autocorrelation is a measurement of the inter connection inside a time series. It is a method for estimating and clarifying interior relationship between perceptions in a time series analysis [13]. According to the concept of autocorrelation, if the first element is closely related to the second, and the second to the third, then the first element must also be somewhat related to the third one. Autocorrelation function (ACF) helps to determine the order of moving average (MA) model in the dataset. Starting from 0, the lag after which the ACF stops crossing the significance bound (red dashed line), is the order of the MA model. If the ACF does not cross the significance bound in the first lag, but does so in case of later lags, then we assume that order of MA is 0.In our paper, autocorrelation has been used for checking randomness in the data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. Calculating the order of parameters, autocorrelation helps to determine the optimal solution for a dataset.

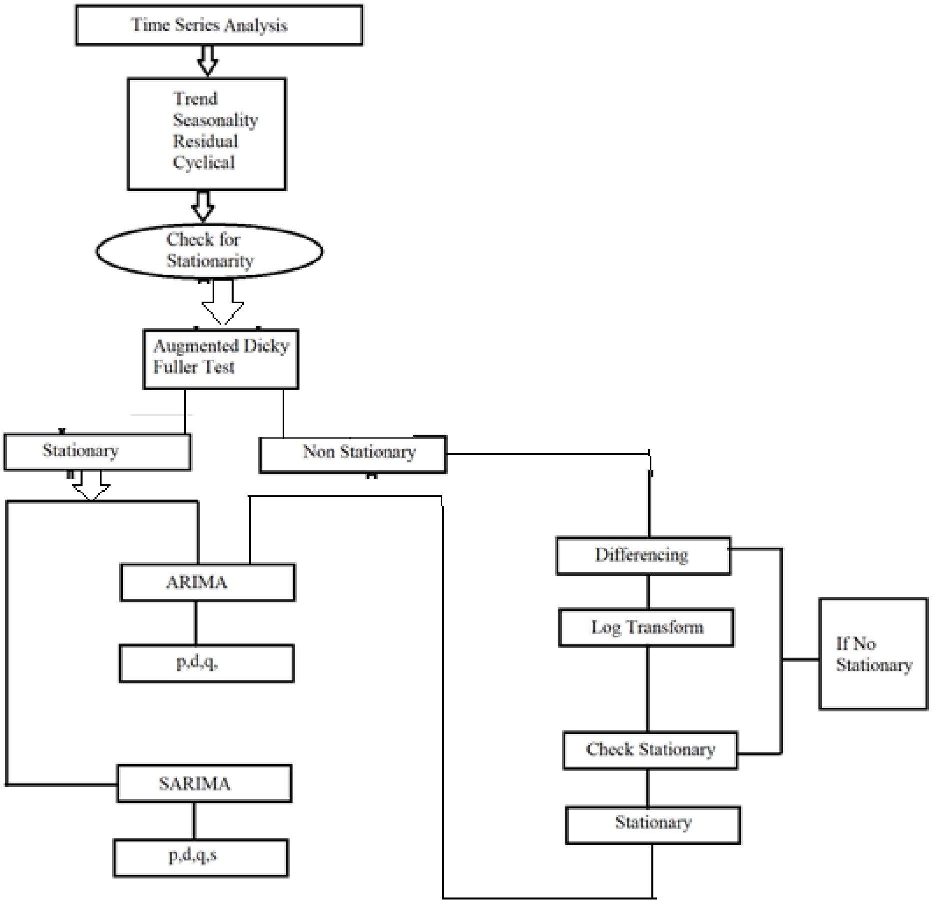

Figure 2. Data flow of proposed time series analysis model.

4.2. Model Selection

We have applied time series models Autoregressive integrated moving average model (ARIMA), Autoregressive model (AR) and Moving average model (MA) in our processed dataset and plotted the resultant graph. Based on the accuracy of the models, we have chosen ARIMA to predict Bitcoin price as our data fitted well in it. Figure2 indicates the data flow of our time series model. Historical data is collected and stationarized. According to autocorrelation and partial auto-correlation graphs, randomness due to time lags is determined and the dataset has been fit in ARIMA/AR/MA model with all available features. Then our model has correlated the day wise closing price with other features such as market capital and volume in transactions which are in the dataset and has found out the pattern of forecasting method that fit more precisely. Then the prediction models make a prediction of next consecutive 4 months bitcoin price and user evaluate the result with the actual price of bitcoin that has been previously stored in CSV file. After calculating the accuracy user can find the best model for price prediction of Bitcoin for the given dataset.

V. EVALUATION

Evaluation enables us to test the model against the information that has never been utilized for the training. We have tried to use several different models and compare their results in this paper. These results were obtained using the following hardware: 4-core CPU, 16 GB RAM and by fitting each model ten times with different random states.We have analyzed the price of Bitcoin with respect to the US Dollar using some of the popular time-series models ARIMA, and LR model and then forecasted the Bitcoin price in USD for the next consecutive 4months. As we have dataset till January 2020 and bitcoin price has been quite unstable at that time, we have chosen to predict the 4 monthsbitcoin price and have kept the remained unstable data to fit our model for better prediction with an appropriate accuracy.

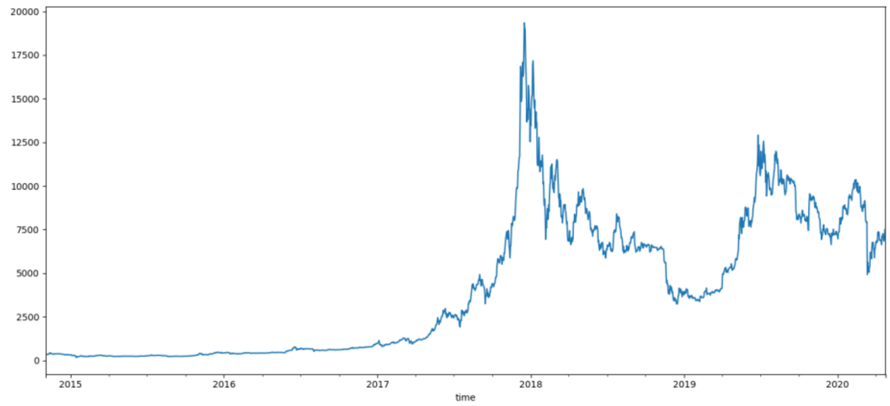

Fig 2: bit coin prediction over years

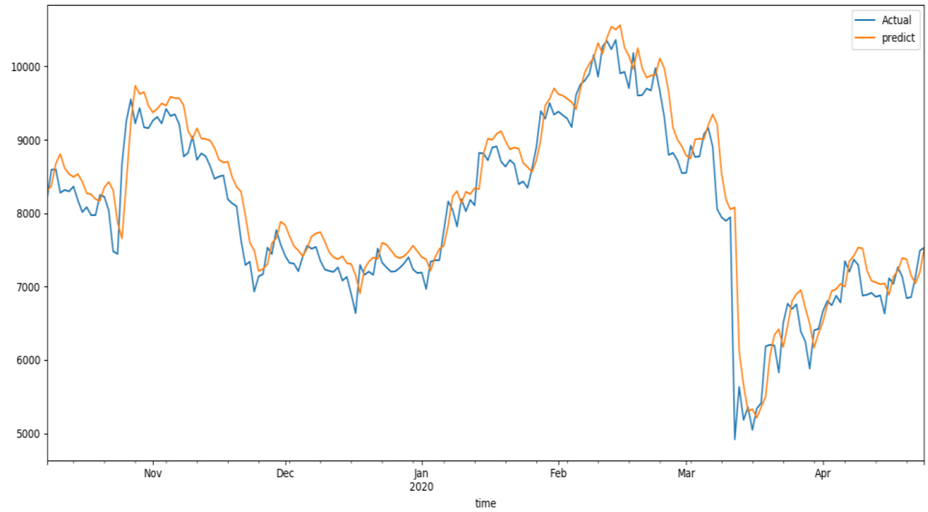

Fig. 3: Bit coin actual and prediction analysis over years.

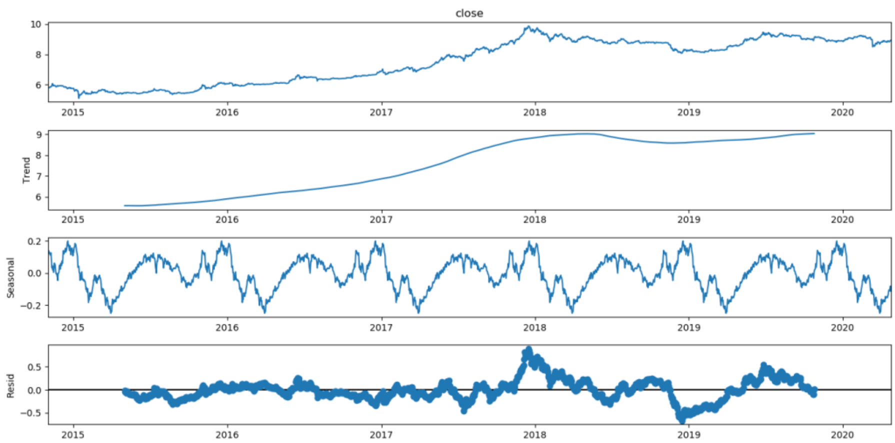

Fig. 4: Time series analysis of bit coin.

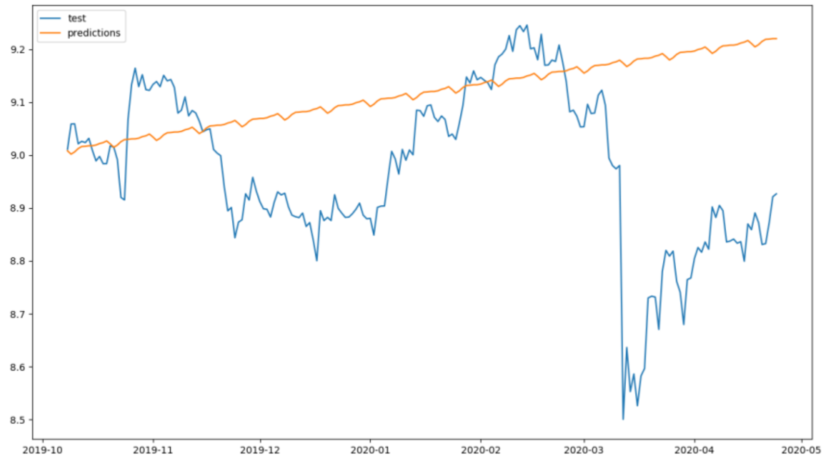

Fig. 5: Bit coin test and prediction analysis over years.

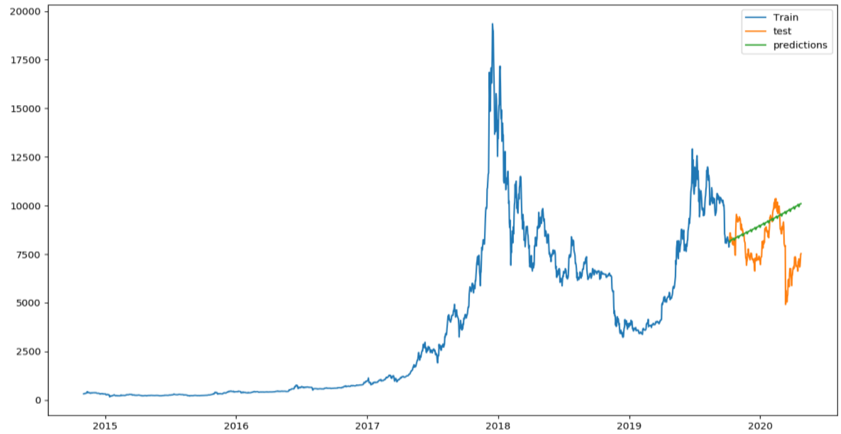

Fig. 6: Bit coin train, test, and prediction analysis over years.

VI. CONCLUSION AND FUTURE WORK

In this work, we have contributed in financial market area by enabling the investors to figure out how to dissect Bitcoin information and furthermore to utilize that learning to predict the future bitcoin price movement. This article represented consecutive next 4 months bitcoin price forecasting method using time series analysis model named ARIMA. After the analysis, finally we have found that ARIMA model performance issuperior as compared to LR model. In future, prediction of cryptocurrency can be done by implementing more effective deep learning frameworks like LST, RNN and CNN instead of machine learning and time series analysis models.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] M. Briere, K. Oosterlinck, and A. Szafarz, "Virtual currency, tangible ` return: Portfolio diversification with bitcoins," Tangible Return: Portfolio Diversification with Bitcoins (September 12, 2013), 2013.

[3] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10, no. 3, pp. 215–236, 1996.

[4] H. White, "Economic prediction using neural networks: The case of ibm daily stock returns," in Neural Networks, 1988., IEEE International Conference on. IEEE, 1988, pp. 451–458.

[5] C. Chatfield and M. Yar, “Holt-winters forecasting: some practical issues,” The Statistician, pp. 129–140, 1988.

[6] B. Scott, “Bitcoin academic paper database,” suitpossum blog, 2016.

[7] M. D. Rechenthin, “Machine-learning classification techniques for the analysis and prediction of high-frequency stock direction,” 2014.

[8] D. Shah and K. Zhang, “Bayesian regression and bitcoin,” in Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014, pp. 409–414.

[9] G. H. Chen, S. Nikolov, and D. Shah, “A latent source model for nonparametric time series classification,” in Advances in Neural Information Processing Systems, 2013, pp. 1088–1096. [10] I. Georgoula, D. Pournarakis, C. Bilanakos, D. N. Sotiropoulos, and G. M. Giaglis, “Using time-series and sentiment analysis to detect the determinants of bitcoin prices,” Available at SSRN 2607167, 2015.

[11] M. Matta, I. Lunesu, and M. Marchesi, “Bitcoin spread prediction using social and web search media,” Proceedings of DeCAT, 2015.

[12] ——, “The predictor impact of web search media on bitcoin trading volumes.”

[13] B. Gu, P. Konana, A. Liu, B. Rajagopalan, and J. Ghosh, “Identifying information in stock message boards and its implications for stock market efficiency,” in Workshop on Information Systems and Economics, Los Angeles, CA, 2006.

[14] A. Greaves and B. Au, “Using the bitcoin transaction graph to predict the price of bitcoin,” 2015.

[15] I. Madan, S. Saluja, and A. Zhao, “Automated bitcoin trading via machine learning algorithms,” 2015.

[16] R. Delfin Vidal, “The fractal nature of bitcoin: Evidence from wavelet power spectra,” The Fractal Nature of Bitcoin: Evidence from Wavelet Power Spectra (December 4, 2014), 2014.

[17] L. Kristoufek, “What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis,” PloS one, vol. 10, no. 4, p. e0123923, 2015.

[18] Y. Yoon and G. Swales, “Predicting stock price performance: A neural network approach,” in System Sciences, 1991. Proceedings of the Twenty-Fourth Annual Hawaii International Conference on, vol. 4. IEEE, 1991, pp. 156–162.

[19] T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski, “Time series prediction with multilayer perceptron, fir and elman neural networks,” in Proceedings of the World Congress on Neural Networks. Citeseer, 1996, pp. 491–496.

[20] C. L. Giles, S. Lawrence, and A. C. Tsoi, “Noisy time series prediction using recurrent neural networks and grammatical inference,” Machine learning, vol. 44, no. 1-2, pp. 161–183, 2001.