Predicting Stock Prices with Neural Networks and other models

| 1st Do Huynh My Tam | 2nd Le Gia Bao | 3rd Trinh Duong Quoc Bao |
| --- | --- | --- |
| University of Information  Technology | University of Information  Technology | University of Information  Technology |
| [20520746@gm.uit.edu.vn](mailto:20520746@gm.uit.edu.vn) | [20521101@gm.uit.edu.vn](mailto:20521101@gm.uit.edu.vn) | [19521261@gm.uit.edu.vn](mailto:19521261@gm.uit.edu.vn) |

***Abstract* — Stocks are simply understood as valuable papers confirming ownership of stocks of business. Many companies decide to issue stocks to serve their growth plan.Stocks represent the ownership of the business by each shareholder, so the business is not obliged to return such capital contribution to the owners of its corporate stocks. The stock market has established itself as an effective capital mobilization channel for the economy. In this paper, we investigate the predictive stock price of some corporations in Vietnam with algorithms like LR, ARIMA, RNN, LSTM, Hot-Winters, SSA, RNN-Attention, NNAR to evaluate the volatility of stock price in the market. From there we can draw conclusions about the pros and cons of these models and predict the price of bitcoin**

***Keywords — Time-series analysis, stock price, LR, ARIMA, RNN, LSTM, Hot-Winters, SSA, RNN-Attention, NNAR***

1. **INTRODUCTION**

As the stock market is one of the most important components of modern economies, the ability to predict stock prices accurately is of great interest to investors and financial analysts alike. With the advent of machine learning and artificial intelligence, there has been a growing interest in developing predictive models for stock prices using various algorithms, including neural networks and other models.

Neural networks are a class of algorithms inspired by the structure and function of the human brain, and they have shown great promise in predicting stock prices due to their ability to capture complex patterns and relationships in the data. Other models such as regression, time series analysis and decision trees have also been used for this purpose.

In this article, we will explore the use of neural networks and other models for predicting stock prices. We will discuss the benefits and limitations of each approach and provide examples of successful applications. Additionally, we will examine the challenges and potential pitfalls associated with these models, and discuss strategies for overcoming them. Overall, this article aims to provide an overview of the current state of research in this field and to highlight the potential of these models for predicting stock prices in the future.

To predict the company's stock, we will use a time-series dataset with different models such as:

* RNN: is an artificial neural network specifically designed to process data that is sequential, like time series or text.
* LSTM: is a variant of recurrent neural network (RNN) designed to solve the vanishing gradient problem and improve the ability to understand and remember information from the past.
* Linear Regression: is a method in machine learning to find the linear relationship between independent and dependent variables.
* Holt-Winters: is a model of time series behavior. Forecasting always requires a model.
* SSA: is a time series analysis method that decomposes a series into components using eigenvectors, useful for identifying cyclical patterns and removing noise.
* RNN: attention model is a neural network that uses attention mechanism to selectively focus on important features of sequential data.
* ARIMA: is a statistical method for modeling and predicting time series of data. The algorithm is a model that combines autoregressive (AR), moving average (MA) and integration (I) components.
* Seq2Seq: is a deep learning model used to handle tasks related to string data, such as machine translation, text summarization, and chatbots
* The Neural Network Autoregression (NNAR) model is a type of neural network that uses historical values of a time series to predict future values.

From those algorithm models, we are going to evaluate the accuracy and relevance (Over-fitting) of the algorithm to the data set and make appropriate conclusions to decide to invest or trade in stock market to make a profit

1. **RELATED RESEARCH**

Realize the importance of stock price prediction for profitable investment and trading. Stock price prediction has become a hot topic among investors and academia.

In this section, we provide a summary of pertinent published works that are technically related to our research paper on time series prediction using mathematical attributes So now there are quite a few scientific papers that use different models to predict the final Price of stock, bitcoins, … other time series kind of datasets

Suhwan Ji, Jongmin Kim and Hyeonseung Im, they study and compare `state of the art` approaches of deep learning method such as a deep neural network (DNN), a long short-term memory (LSTM) model, a convolutional neural network, a deep residual network, and their combinations for Bitcoin price prediction. [[1]](https://www.zotero.org/google-docs/?ToqOIv)

Poongodi M, Vijayakumar V, Naveen Chilamkurti, they collected the dataset on bitcoin blockchain from April 28th, 2013 to July 31 st , 2017 which is publicly available on https://coinmarketcap.com and applied the ARIMA model for price prediction of bitcoin. [[2]](https://www.zotero.org/google-docs/?sfGPPT)

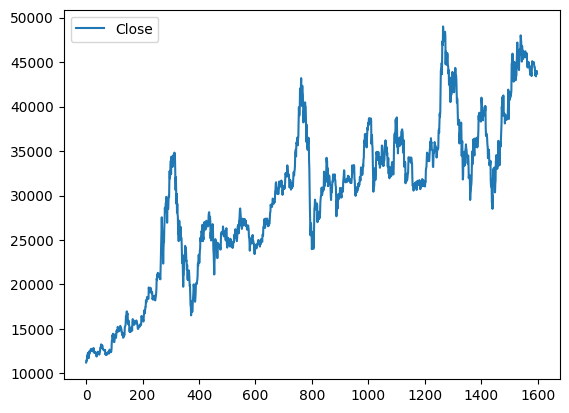
Aniruddha Dutta, Saket Kumar and Meheli Basu, they investigated a framework with a set of advanced machine learning forecasting methods with a fixed set of exogenous and endogenous factors to predict daily Bitcoin prices. Specifically, they studied and compared different approaches using the root mean squared error (RMSE). And the experimental results showed that the gated recurring unit (GRU) model with recurrent dropout performs better than popular existing models. [[3]](https://www.zotero.org/google-docs/?gvpy7C)

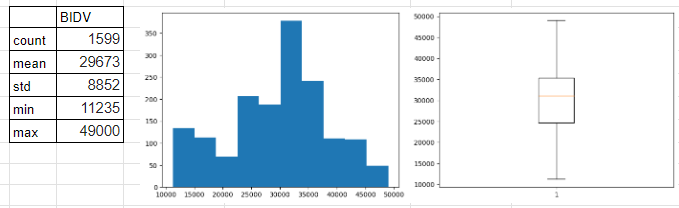
Hari Krishnan Andi normalized a particular dataset. Then, this dataset has been trained to deploy a more accurate forecast of the bitcoin price. Furthermore, this research work has evaluated different machine learning methods and found that the suggested work delivers better results [4]

The Holt-Winters Forecasting Procedure by Chatfield C. This paper points out that these empirical studies have used the automatic version of the method, whereas a non-automatic version is also possible in which subjective judgment is employed, for example, to choose the correct model for seasonality. [5]

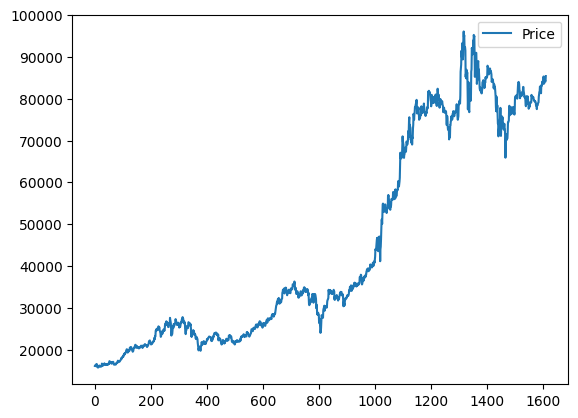
1. **DATASET**

**Preprocessing data**

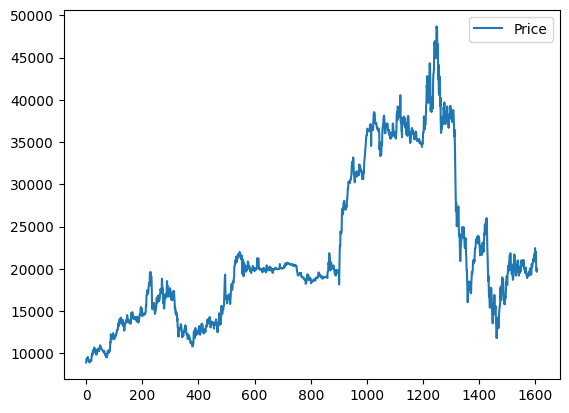
****

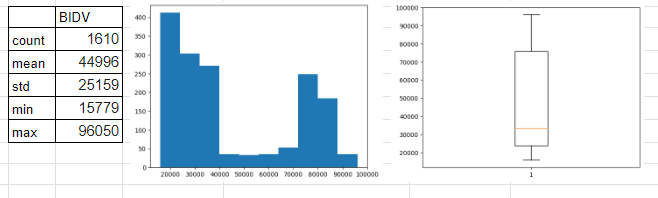
****

**The figure of the closing price of BIDV**

****

**The figure of the closing price of CVG**

****

****

**The figure of the closing price of FPT**

1. **METHODOLOGY**

In this paper, we use 10 models.

1. Linear Regression

Linear regression is a simple but popular prediction algorithm. It predicts the value based on the linear relationship between the independent input variables and the dependent variable to be predicted. If in the Linear Regression model, we use one dependent and one independent variable, it is called simple linear regression. If we use more than one dependent and independent variable, it is called multi-linear regression. Formula of Linear Regression

f(x) = a + bx

**With:**

* f(x) is the output (dependent variable)
* x: is the output (independent variable)
* a: is a constant
* b: is the coefficient of linear equation

1. ARIMA

ARIMA (Autoregressive Integrated Moving Average) is a popular time series forecasting model that has found extensive use in economics, finance, and many other disciplines. ARIMA combines three major factors to forecast time series data: autoregression (AR), moving average (MA), and integration (I).

* Autoregression (AR) is a forecasting model based on the values of the time series in the past.
* Moving average (MA) is a forecasting model based on the difference between the value of the time series and its predicted value.
* Integration (I) is a forecasting model based on the difference between the value of the time series and its mean value.

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

* p: the number of lag observations in the model, also known as the lag order.
* also known as the degree of differencing.
* q: the size of the moving average window, also known as the order of the moving average.·

1. RNN [[4]](https://www.zotero.org/google-docs/?NzMjR3)

A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (nlp), speech recognition, and image captioning;

Popular type of RNN:

Bidirectional recurrent neural networks (BRNN)

Long short-term memory (LSTM)

Gated recurrent units (GRUs)

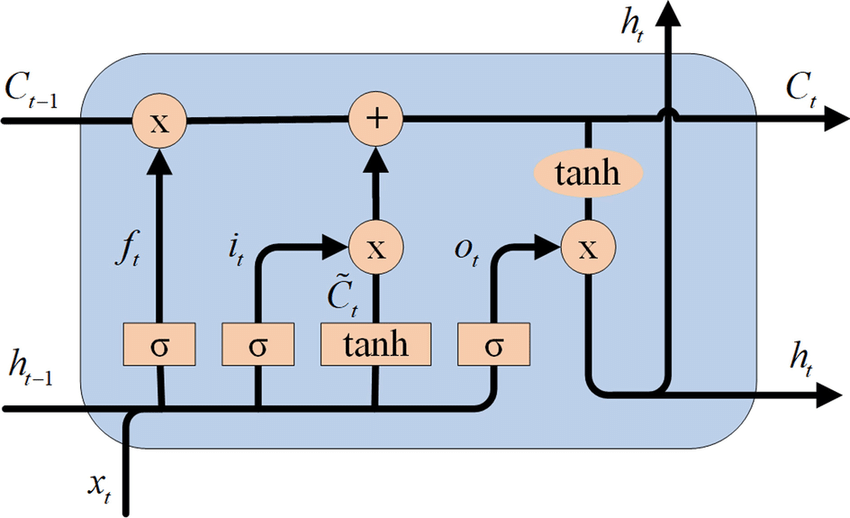
1. LSTM

Long Short-Term Memory Networks is a deep learning, sequential neural net that allows information to persist. It is a special type of Recurrent Neural Network which is capable of handling the vanishing gradient problem faced by traditional RNN.[[5]](https://www.zotero.org/google-docs/?iotvD4)

A common problem in deep networks is the “vanishing gradient” problem, where the gradient gets smaller and smaller with each layer until it is too small to affect the deepest layers. With the memory cell in LSTMs, we have continuous gradient flow (errors maintain their value) which thus eliminates the vanishing gradient problem and enables learning from sequences which are hundreds of time steps long.

LSTMs deal with both Long Term Memory (LTM) and Short Term Memory (STM) and for making the calculations simple and effective it uses the concept of gates.

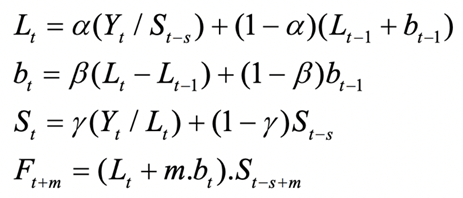
* Forget Gate: LTM goes to forget gate and it forgets information that is not useful.
* Learn Gate: Event (current input) and STM are combined together so that necessary information that we have recently learned from STM can be applied to the current input.
* Remember Gate: LTM information that we haven’t forgotten and STM and Event are combined together in Remember gate which works as updated LTM.
* Use Gate: This gate also uses LTM, STM, and Event to predict the output of the current event which works as an updated STM.



1. Hot-Winters[[6]](https://www.zotero.org/google-docs/?BMmjrh)

Holt-Winters is a model of time series behavior. Forecasting always requires a model, and Holt-Winters is a way to model three aspects of the time series: a typical value (average), a slope (trend) over time, and a cyclical repeating pattern (seasonality).

Time series anomaly detection is a complicated problem with plenty of practical methods. It’s easy to get lost in all of the topics it encompasses. Learning them is certainly an issue, but implementing them is often more complicated. A key element of anomaly detection is forecasting—taking what you know about a time series, either based on a model or its history, and making decisions about values that arrive later.



1. SSA[[7]](https://www.zotero.org/google-docs/?mpFNlE)

Singular Spectrum Analysis (SSA) is a powerful technique in time series analysis with broad application areas. SSA consists of two stages: decomposition and reconstruction. The main concept in studying the properties of SSA is separability. SSA does not require statistical assumptions such as stationarity or normality. SSA can be used for finding trends, smoothing, extracting seasonality components, and more. SSA also has several essential extensions including multivariate version, forecasting procedures, and change-point detection. In nonlinear time series analysis, SSA has been considered as a filtering method and used in biomedical diagnostic tests and noise reduction for longitudinal measurements.

SSA is a nonparametric technique that works well with arbitrary statistical processes, whether linear or nonlinear, stationary or non-stationary, Gaussian or non-Gaussian. Unlike traditional methods of time series forecasting, SSA method is non-parametric and makes no prior assumptions about the data. The real time series usually has a complex structure, and SSA method decomposes a series into its component parts, leaving the random (noise) component behind. SSA also works well even for small sample sizes. Although some probabilistic and statistical concepts are employed in SSA-based methods, no statistical assumptions such as stationarity or normality are required, and bootstrapping is used to obtain confidence intervals for the forecasts.

1. Sequence to sequence[[8]](https://www.zotero.org/google-docs/?rCEGH5)

Sequence to sequence or Seq2seq is a family of machine learning approaches used for natural language processing. Applications include language translation, image captioning, conversational models, and text summarization.

Seq2seq turns one sequence into another sequence (sequence transformation). It does so by use of a recurrent neural network (RNN) or more often LSTM or GRU to avoid the problem of vanishing gradient. The context for each item is the output from the previous step. The primary components are one encoder and one decoder network. The encoder turns each item into a corresponding hidden vector containing the item and its context. The decoder reverses the process, turning the vector into an output item, using the previous output as the input context.

Attention: The input to the decoder is a single vector which stores the entire context. Attention allows the decoder to look at the input sequence selectively.

Beam Search: Instead of picking the single output (word) as the output, multiple highly probable choices are retained, structured as a tree (using a Softmax on the set of attention scores). Average the encoder states weighted by the attention distribution.[8]

Bucketing: Variable-length sequences are possible because of padding with 0s, which may be done to both input and output. However, if the sequence length is 100 and the input is just 3 items long, expensive space is wasted. Buckets can be of varying sizes and specify both input and output lengths.

Training typically uses a cross-entropy loss function, whereby one output is penalized to the extent that the probability of the succeeding output is less than 1.[8]

1. Neural Network AutoRegression (NNAR)[[9]](https://www.zotero.org/google-docs/?uv7osc)

The NNAR model is an Artificial Neural Network

NNAR-Neural Network Autoregression Model- has two components,

denotes the number of lagged values that are used as inputs. k denotes the number of hidden nodes that are present.

Output is denoted by NNAR(p,k) If the dataset is seasonal then also the notation is pretty similar, i.e., NNAR(p,P,k)

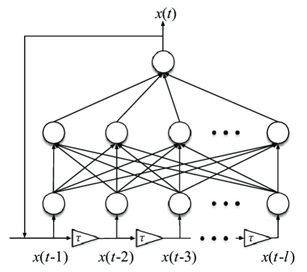
where

P denotes the number of seasonal lags.

p is choosen based on the information criterion, like AIC.

Neural nets has inherent random component. Therefore, it is suggested that the neural net model is run several times, 20 is the minimum requirement. Final result is then presented as mean or median. Also neural nets are known to not work well with the trend data. We should therefore, de-trend or difference the data before running neural net model

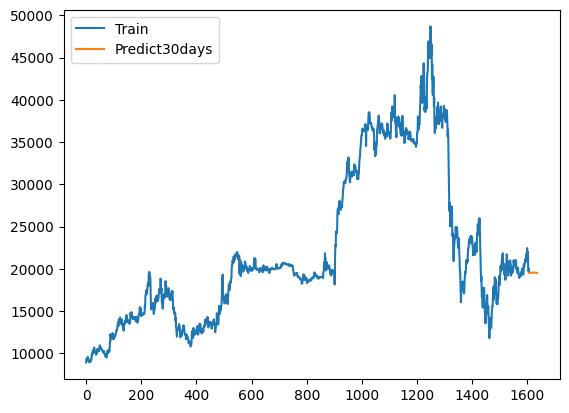
Also neural nets are known to not work well with the trend data. We should therefore, de-trend or difference the data before running neural net model



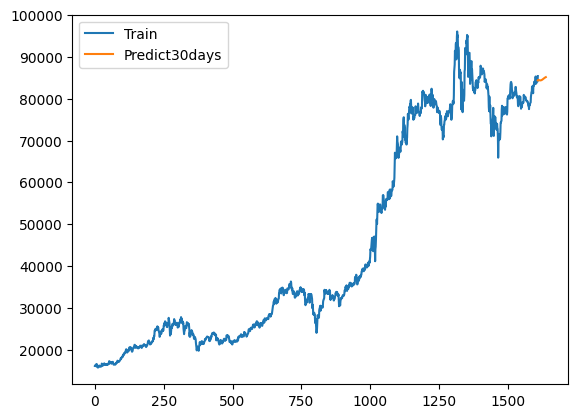
1. **RESULT**

| **Dataset** | **Model** | **Train:Test** | **RMSE** | **MAPE(%)** |
| --- | --- | --- | --- | --- |
| BIDV | LR | 7:3 | 6089 | 14.75 |
| 8:2 | 5003 | 11.16 |
| 9:1 | 4109 | 8.5 |
| ARIMA | 7:3 | 6306 | 18.143 |
| 8:2 | 10356 | 25.916 |
| 9:1 | 12027 | 26.196 |
| RNN | 7:3 | 39570 | 53.12 |
| 8:2 | 1752 | 3.11 |
| 9:1 | 1757 | 3.89 |
| LSTM | 7:3 | 7918 | 20.101 |
| 8:2 | 6960 | 19.706 |
| 9:1 | 9530 | 3.6766 |
| Holt-Winters | 7:3 | 7513 | 28.75 |
| 8:2 | 6351 | 25.16 |
| 9:1 | 5551 | 15.5 |
| SSA | 7:3 | 9314 | 25.36 |
| 8:2 | 7655 | 22.12 |
| 9:1 | 6351 | 18.25 |
| Seq2seq | 7:3 | 1011 | 2.11 |
| 8:2 | 919.3 | 2.06 |
| 7.5:2.5 | 1078.6 | 2.67 |
| NNAR | 7:3 | 42274 | 15.542 |
| 8:2 | 38826 | 19.884 |
| 9:1 | 38181 | 22.58 |
| VCG | LR | 7:3 | 6157 | 15.75 |
| 8:2 | 5035 | 12.18 |
| 9:1 | 4237 | 9.16 |
| ARIMA | 7:3 | 15306 | 60.823 |
| 8:2 | 18499 | 90.573 |
| 9:1 | 3580 | 17.181 |
| LSTM | 7:3 | 9945 | 72.005 |
| 8:2 | 6988 | 36.751 |
| 9:1 | 4195 | 8.2231 |
| Holt-Winters | 7:3 | 6070 | 20.89 |
| 8:2 | 5009 | 18.23 |
| 9:1 | 4589 | 10.15 |
| SSA | 7:3 | 8947 | 25.32 |
| 8:2 | 6555 | 20.15 |
| 9:1 | 5545 | 19.24 |
| Seq2seq | 8:2 | 1062.5 | 2.23 |
| 7:3 | 1222.7 | 2.11 |
| 7.5:2.5 | 968.3 | 1.98 |
| NNAR | 7:3 | 28735 | 75.497 |
| 8:2 | 22320 | 49.055 |
| 9:1 | 19059 | 30.738 |
| FPT | LR | 7:3 | 9515 | 30.21 |
| 8:2 | 8565 | 25.16 |
| 9:1 | 6165 | 19.24 |
| ARIMA | 7:3 | 72801 | 77.06 |
| 8:2 | 8622 | 7.697 |
| 9:1 | 3101 | 4.651 |
| RNN | 7:3 | 77649 | 96.12 |
| 8:2 | 7184 | 16.06 |
| 7.5:2.5 | 9050 | 2.19 |
| LSTM | 7:3 | 8099 | 8.974 |
| 8:2 | 8097 | 6.886 |
| 9:1 | 6972 | 4.009 |
| Holt-Winters | 7:3 | 166293 | 493.74 |
| 8:2 | 57776 | 230.6 |
| 9:1 | 29643 | 118.05 |
| SSA | 7:3 | 30.79 | 21.05 |
| 8:2 | 135.18 | 131.17 |
| 9:1 | 22671 | 22.42 |
| Seq2seq | 7:3 | 1874.4 | 4.06 |
| 8:2 | 2504.1 | 4.67 |
| 7.5:2.5 | 2036.5 | 4.07 |
| NNAR | 7:3 | 80705 | 8.364 |
| 8:2 | 81629 | 8.737 |
| 9:1 | 78566 | 6.148 |

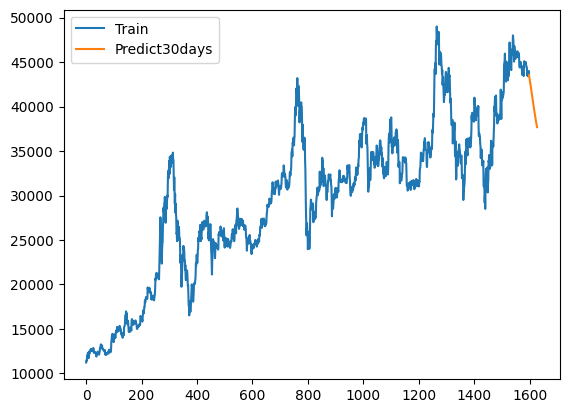
The predicted results of CVG company's stock for the next 30 days using the LSTM model with a train-test split of 9:1.



The predicted results of FPT company's stock for the next 30 days using the LSTM model with a train-test split of 9:1.



The predicted results of BIDV company's stock for the next 30 days using the LSTM model with a train-test split of 9:1

.

1. **CONCLUSION**

Studies have shown that, compared to other prediction models, the LSTM model often produces the most accurate predictions for complex time series. This suggests that the LSTM model has the potential to be widely applied in fields such as finance, e-commerce, and healthcare to predict events such as sales figures, stock prices, and important health variables. Compared to traditional RNN algorithms, LSTM has the ability to learn long-term dependencies in the data, making it widely used in time series forecasting applications.

However, to effectively apply the LSTM model in practice, high-quality and complete input data is needed, the model parameters must be properly determined, and the model must be trained with enough diverse and large data. This requires investment and effort from experts and developers to apply the LSTM model to real-world prediction problems.

**REFERENCES**

[**[1] S. Ji, J. Kim, and H. Im, “A Comparative Study of Bitcoin Price Prediction Using Deep Learning,” *Mathematics*, vol. 7, p. 898, Sep. 2019, doi: 10.3390/math7100898.**](https://www.zotero.org/google-docs/?kYgtjk)

[**[2] “(PDF) Bitcoin price prediction using ARIMA model.” https://www.researchgate.net/publication/342694270\_Bitcoin\_price\_prediction\_using\_ARIMA\_model (accessed Jun. 17, 2023).**](https://www.zotero.org/google-docs/?kYgtjk)

[**[3] “JRFM | Free Full-Text | A Gated Recurrent Unit Approach to Bitcoin Price Prediction.” https://www.mdpi.com/1911-8074/13/2/23 (accessed Jun. 17, 2023).**](https://www.zotero.org/google-docs/?kYgtjk)

[**[4] “What are Recurrent Neural Networks? | IBM.” https://www.ibm.com/topics/recurrent-neural-networks (accessed Jun. 18, 2023).**](https://www.zotero.org/google-docs/?kYgtjk)

**[5] C. Chatfield, “The Holt-Winters Forecasting Procedure,” J. R. Stat. Soc. Ser. C Appl. Stat., vol. 27, no. 3, pp. 264–279, 1978, doi: 10.2307/2347162.**

[**[6] “LSTM | Introduction to LSTM | Long Short Term Memory Algorithms.” https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/ (accessed Jun. 17, 2023).**](https://www.zotero.org/google-docs/?kYgtjk)

[**[7] Snehal\_bm, “Holt Winter’s Method for Time Series Analysis,” *Analytics Vidhya*, Aug. 03, 2021. https://www.analyticsvidhya.com/blog/2021/08/holt-winters-method-for-time-series-analysis/ (accessed Jun. 17, 2023).**](https://www.zotero.org/google-docs/?kYgtjk)

[**[8] “(PDF) Singular Spectrum Analysis for Time Series.” https://www.researchgate.net/publication/260124592\_Singular\_Spectrum\_Analysis\_for\_Time\_Series (accessed Jun. 17, 2023).**](https://www.zotero.org/google-docs/?kYgtjk)

[**[9] “Seq2seq,” *Wikipedia*. Jun. 16, 2023. Accessed: Jun. 18, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Seq2seq&oldid=1160506318**](https://www.zotero.org/google-docs/?kYgtjk)

[**[10] *12.4 Neural network models | Forecasting: Principles and Practice (3rd ed)*. Accessed: Jun. 17, 2023. [Online]. Available: https://otexts.com/fpp3/nnetar.html**](https://www.zotero.org/google-docs/?kYgtjk)