

Logo

Description automatically generated

**FINAL REPORT**

**Predicting Stock Prices with Neural Networks and other models**

Instructors: Assoc. Prof. Nguyen Dinh Thuan

Nguyen Minh Nhut

Pham Viet Huong

Class: IS403.N21.HTCL

Group: 13

Members: Do Huynh My Tam – 20520746

Le Gia Bao – 20521101

Trinh Duong Quoc Bao - 19521261

**Ho Chi Minh, June 2023**

***Instructor's Comments***

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

……………………………………………………………………………………………………………………………………………………………………………………

# TABLE OF CONTENTS

[TABLE OF CONTENTS 3](#_gjdgxs)

[I. INTRODUCTION 5](#_30j0zll)

[1. Joint Stock Commercial Bank for Investment and Development of Vietnam (BID) 6](#_1fob9te)

[2. The Corporation for Financing Promoting Technology (FPT) 7](#_3znysh7)

[3. Vietnam Construction and Import Export JSC (VCG) 8](#_2et92p0)

[II. RELATED WORK 8](#_tyjcwt)

[III. MATERIALS AND METHODS 10](#_3dy6vkm)

[1. Data collection 10](#_1t3h5sf)

[2. Algorithm 12](#_4d34og8)

[2.1 Linear Regression 12](#_2s8eyo1)

[2.2 ARIMA 13](#_2jxsxqh)

[2.3 Recurrent Neural Networks (RNN) 14](#_z337ya)

[2.4 Long Short-Term Memory Networks (LSTM) 16](#_3j2qqm3)

[2.5 Hot-Winters 17](#_1y810tw)

[2.6 Singular Spectrum Analysis (SSA) 19](#_4i7ojhp)

[2.7 Sequence to sequence 21](#_2xcytpi)

[2.8 Neural Network AutoRegression (NNAR) 22](#_1ci93xb)

[IV. RESULT 23](#_3whwml4)

[V. CONCLUSION 27](#_2bn6wsx)

[VI. GROUP WORK DISTRIBUTION 28](#_qsh70q)

[VII. REFERENCE 29](#_3as4poj)

**I. 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: Recurrent Neural Network 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: Singular Spectrum Analysis is a time series analysis method that decomposes a series into components using eigenvectors, useful for identifying cyclical patterns and removing noise.
* 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: Sequence-to-Sequence is a deep learning model used to handle tasks related to string data, such as machine translation, text summarization, and chatbots
* RNN with Attention Mechanism: Recurrent Neural Network with Attention model is a recurrent neural network that uses attention mechanism to selectively focus on important features of sequential data.
* 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

The following section will introduce some businesses selected for research.They are one of the most successful and typical businesses in their field.

## 

## **1. Joint Stock Commercial Bank for Investment and Development of Vietnam (BID)**

Joint Stock Commercial Bank for Investment and Development of Vietnam (or BIDV) provides various banking products and services for individuals, corporate customers, and financial institutions in Vietnam. The company was formerly known as Bank for Investment and Construction of Vietnam and changed its name to Joint Stock Commercial Bank for Investment and Development of Vietnam in 1990.[[1]](https://www.zotero.org/google-docs/?NxLnIE)**A picture containing text, screenshot, plot, font

Description automatically generated**

*Figure 1: BID’s stock price from 2014 to 2023*

## **2. The Corporation for Financing Promoting Technology (FPT)**

FPT Corporation provides IT and telecommunication products and services in Vietnam and internationally. The company operates through Information Technology and Telecommunication, and Investment, Education and Others segments. FPT Telecom is a member of FPT Corporation - one of the largest information technology service companies in Vietnam with the main business of providing information technology products and services.[[2]](https://www.zotero.org/google-docs/?u1JNFn)

*A picture containing text, screenshot, diagram, plot

Description automatically generated*

*Figure 2: FPT's stock price from 2007 to 2023*

## **3. Vietnam Construction and Import Export JSC (VCG)**

Viet Nam Construction and Import Export Joint Stock Corp (Vinaconex) is a Vietnam-based construction and engineering company. It constructs various types of commercial, institutional, industrial and residential structures, as well as provides architectural, design and engineering services. The Company is also engaged in real estate operations with investments in new urban areas, industrial zones, office and apartment buildings, hotels and resorts and commercial complexes. [[3]](https://www.zotero.org/google-docs/?AEvQJ3)A picture containing text, diagram, plot, screenshot

Description automatically generated

*Figure 3: VCG’s stock price from 2008 to 2023*

# II. RELATED WORK

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

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. [[5]](https://www.zotero.org/google-docs/?k7PL0d)

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 recurrent unit (GRU) model with recurrent dropout performs better than popular existing models. [[6]](https://www.zotero.org/google-docs/?hiJQRM)

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

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. The paper re-analyses seven series from the Newbold-Granger study for which Box-Jenkins forecasts were reported to be much superior to the (automatic) Holt-Winters forecasts. The series do not appear to have any common properties, but it is shown that the automatic Holt-Winters forecasts can often be improved by subjective modifications [[8]](https://www.zotero.org/google-docs/?nricy4)

Applied Attention-based LSTM neural networks in stock prediction. This study tried to predict stock price movements using deep learning models. Although the attention mechanism has gained popularity recently in neural machine translation, little focus has been devoted to attention-based deep learning models for stock prediction. This paper proposes an attention-based long short-term memory model to predict stock price movement and make trading strategies [[9]](https://www.zotero.org/google-docs/?gvaLlK)

Next-Day Bitcoin Price Forecast by Ziaul Haque Munim, Mohammad Hassan Shakil and Ilan Alon.This study analyzes forecasts of Bitcoin price using the autoregressive integrated moving average (ARIMA) and neural network autoregression (NNAR) models. Employing the static forecast approach, we forecast next-day Bitcoin price both with and without re-estimation of the forecast model for each step. [[10]](https://www.zotero.org/google-docs/?xqjHtT)

# 

# III. MATERIALS AND METHODS

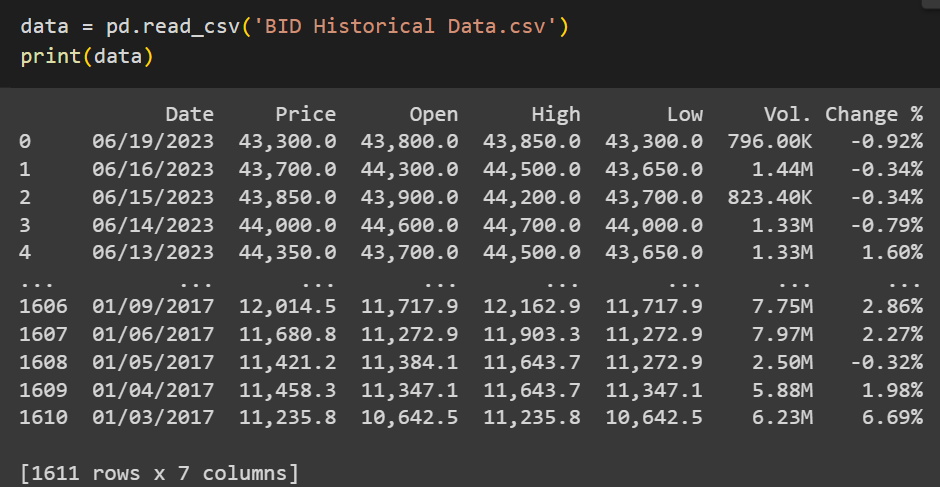
## **1. Data collection**

We retrieved a dataset regarding the stock market price of BID, FPT and VCG from website [investing.com](http://investing.com), which provides all stock trading information for all firms in the globe, including Vietnam. Each data collection contains:

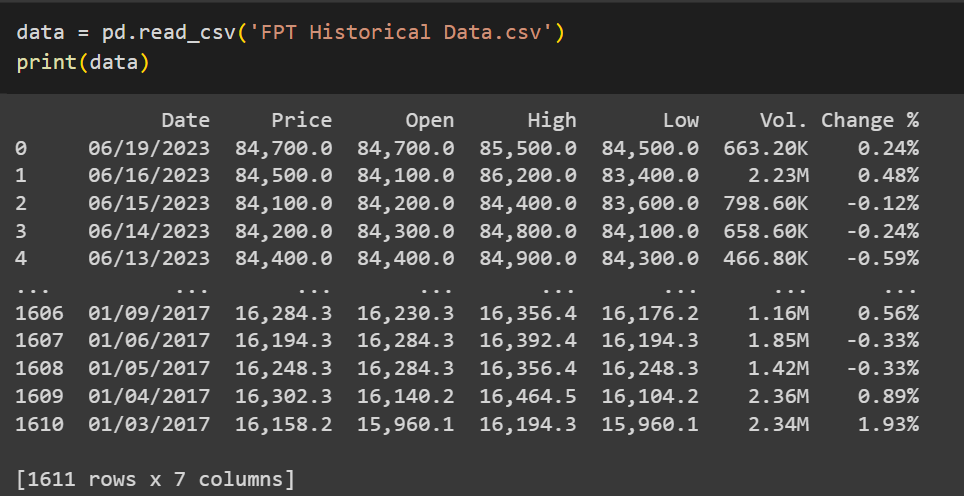
* **Date**: stock trading opening day.
* **Price** (also known as Close Price): the last price at which a stock trades upon the end of the exchange.
* **Open**: the first price at which a stock opens for trading.
* **High**: highest stock price of the day.
* **Low**: lowest stock price of the day.
* **Volume**: the number of shares that the trader buys and sells.
* **Change**: today's change in stock price from the previous day expressed as a percentage.

## **2. Data Statistics**

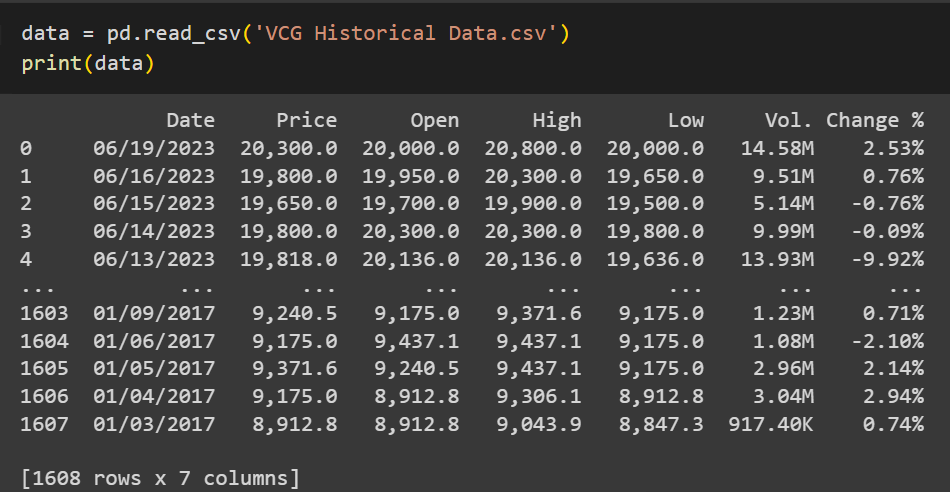
***2.1. Draw Excel files***

******

***BIDV start day is 3/1/2017 and last day in dataset is 19/6/2023***

******

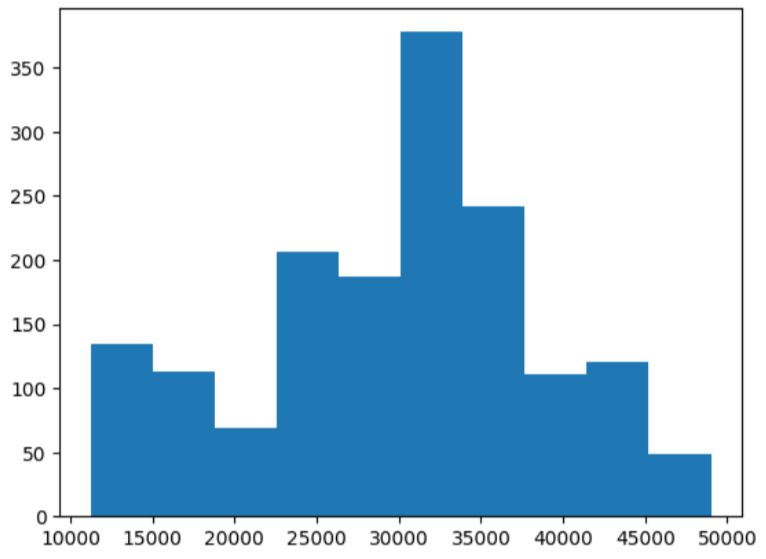
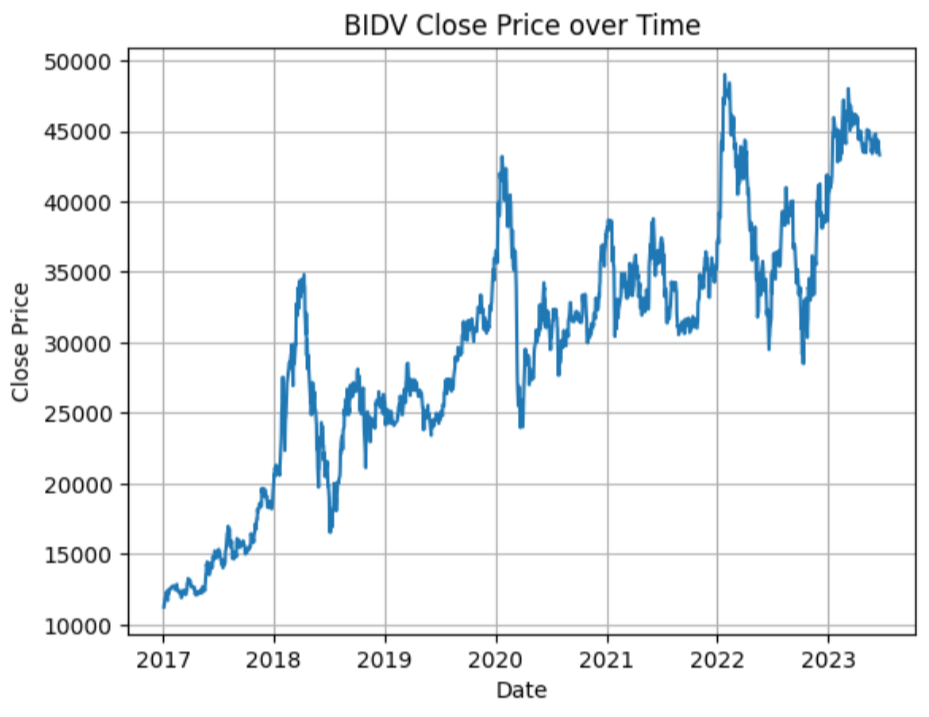
***FPT start day is 3/1/2017 and last day in dataset is 19/6/2023***

******

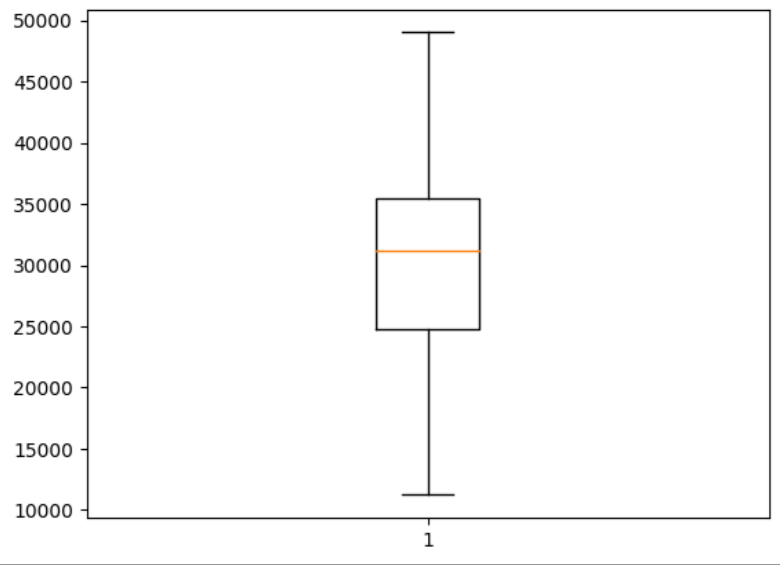
***Same as VCG, start day is 3/1/2017 and last day in dataset is 19/6/2023***

***2.2. Data Statistic***

* **For BID or BIDV closing price**

****

*Figure 4: BID Histogram*

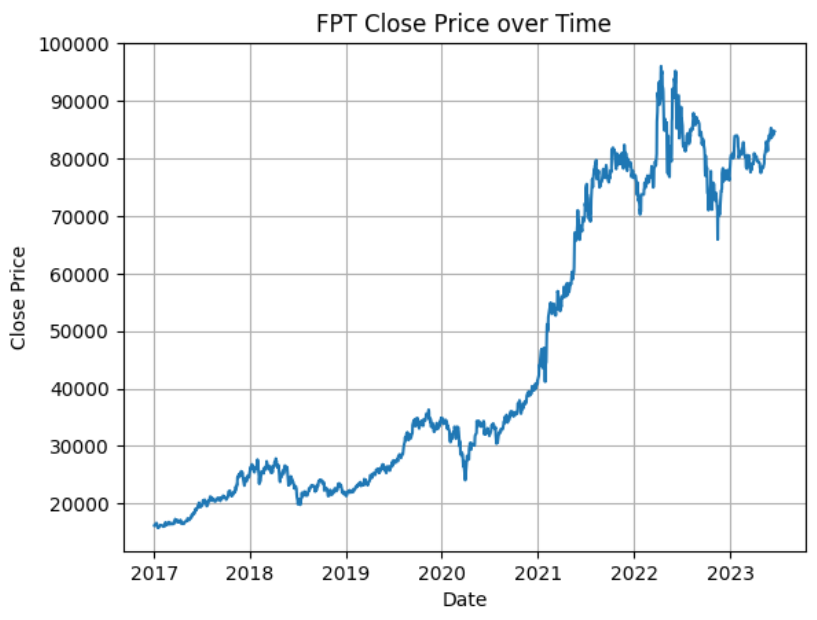
****

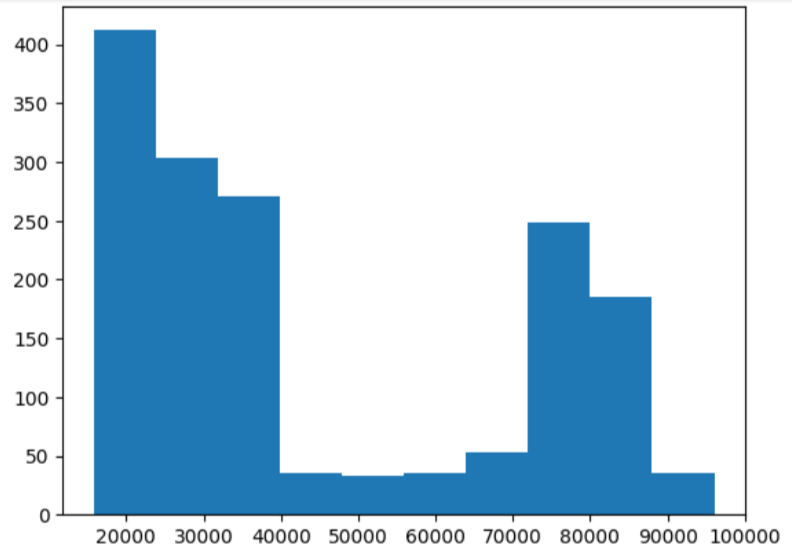
*Figure 5: BID Box Plot*

| ***Name*** | ***Value*** |
| --- | --- |
| **Count** | **1611** |
| **Mean** | **29780.6** |
| **Std** | **8905.51** |
| **Min** | **11235.8** |
| **25%** | **24720.6** |
| **50%** | **31130.1** |
| **75%** | **35436.25** |
| **Max** | **49000** |

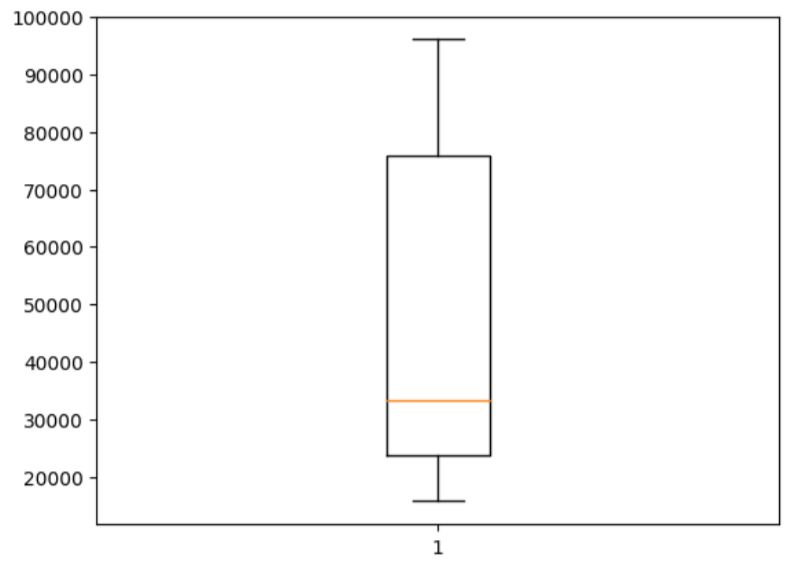
**-> We can see this dataset doesn't have much variance. Data points are also distributed fairly evenly around the mean, and there are no outliers.**

* **For FPT closing price**

****

****

*Figure 5: FPT Histogram*

****

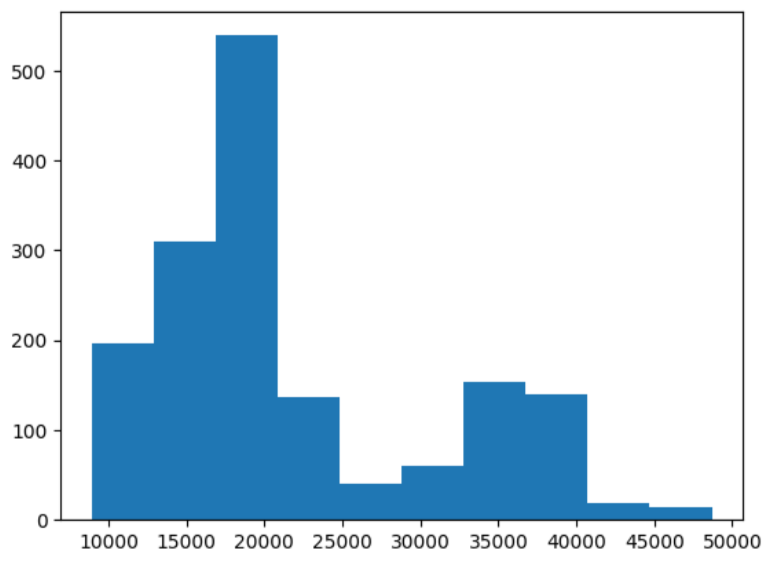
*Figure 6: FPT Box Plot*

| ***Name*** | ***Value*** |
| --- | --- |
| **Count** | **1611** |
| **Mean** | **45020.79** |
| **Std** | **25169.95** |
| **Min** | **15779.9** |
| **25%** | **23691.4** |
| **50%** | **33397.1** |
| **75%** | **75792.5** |
| **Max** | **96050** |

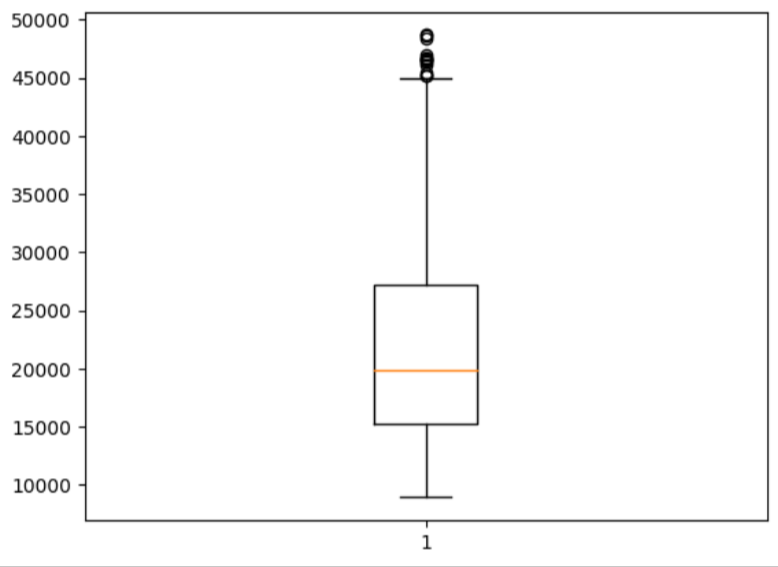
**-> With this dataset, there is no doubt to see the difference in variance compared to the previous dataset. It’s much higher (25169.95 compared to 8905.51). The distribution of the data points is also skewed above the mean quite a lot, and there are no outliers**

* **For VCG closing price**

****

****

*Figure 7: VCG Histogram*

****

*Figure 8: VCG Box Plot*

| ***Name*** | ***Value*** |
| --- | --- |
| **Count** | **1608** |
| **Mean** | **22018.89** |
| **Std** | **9066.18** |
| **Min** | **8912.8** |
| **25%** | **15218.7** |
| **50%** | **19806.7** |
| **75%** | **27215** |
| **Max** | **48688** |

**-> VCG is the Last dataset, could say it has an acceptable Variance. Data points are also distributed fairly evenly around the mean. But this one have outliers**

## 2. Algorithm

### **2.1 Linear Regression** [**[11]**](https://www.zotero.org/google-docs/?4XchHa)

Linear Regression is a supervised machine learning algorithm used to model the linear relationship between an independent variable (also known as the explanatory variable) and a dependent variable (also known as the predicted variable).

### Linear regression 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

### **2.2 ARIMA [8]**

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.

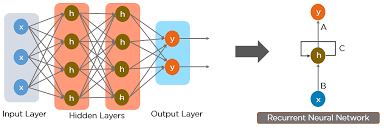
Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

* 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.·

### **2.3 Recurrent Neural Networks (RNN) [9]**

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

RNN’s structure:

****

*Figure 4: Model of RNN architectural*

* **Input Layer:** Networks have only one input layer.
* **Hidden Layer**: Network will have one or more hidden layers.
* **Output Layer:** Networks have only one output layer.

Another difference is that in other neural networks, each input is independent from each other, while in RNNs the inputs are related to each other. RNN establishes relationships between inputs in order to follow the next step. It realizes education based on these relationships.

Recurrent Neural Networks are awesome, they usually thought of as a stepping stone to understanding fancier things like LSMT and Transformers.

“x” is the input layer, “h” is the hidden layer, and “y” is the output layer. A, B, and C are the network parameters used to improve the output of the model. At any given time t, the current input is a combination of input at x(t) and x(t-1). The output at any given time is fetched back to the network to improve on the output.[[12]](https://www.zotero.org/google-docs/?W0GWt0)

A picture containing diagram, circle, line, screenshot

Description automatically generated

*Figure 5: RNN*

*Where:*

* : new state
* : function with parameter c
* : old state
* : input vector at time step t

**RNN model with sigmoid nonlinearity and softmax output:**

)

**Loss Function at a step:**

**Popular type of RNNs:**

Bidirectional recurrent neural networks (BRNN)

Long short-term memory (LSTM)

Gated recurrent units (GRUs)

### **2.4 Long Short-Term Memory Networks (LSTM)**

Long Short-Term Memory (LSTM) is an improvement from the RNNs, which able to solve the Gradient Problem. The LSTM models essentially extend the RNN's memory to enable them to keep and learn long-term dependencies of inputs. The figure below will show the LSTM Architecture.[[13]](https://www.zotero.org/google-docs/?Av87rX)

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.

A diagram of a flowchart

Description automatically generated with medium confidence

*Figure 6: Model of LSTM architectural*

**Forget gate**: f\_t= σ(W\_f.[h\_(t-1),x\_t ]+b\_f)

**Input gate:** i\_t= σ(W\_i . [h\_(t-1),x\_t ]+b\_i)

**Temporary cell state:** (C\_t ) ̃=tanh(W\_c . [h\_(t-1),x\_t ]+b\_c)

**Current cell state:** C\_t=f\_t\*C\_(t-1)+i\_t\*(C\_t ) ̃

**Output gate:** o\_t= σ(W\_i . [h\_(t-1),x\_t ]+b\_o)

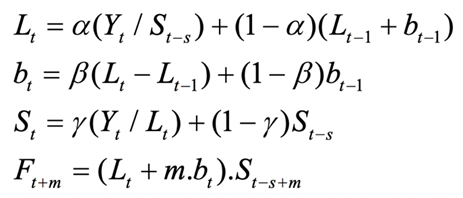
### **2.5 Hot-Winters** [**[14]**](https://www.zotero.org/google-docs/?0puOJs)

The Holt-Winters model is a time series forecasting model used to simulate and predict seasonal data patterns. It was named after the mathematicians Charles Holt, Peter Winters, and Kenneth Brown, who introduced it in 1960.

**Structure of the Holt-Winters model:**

The Holt-Winters model consists of three main components: the level component, the trend component, and the seasonal component. These components are combined to predict future values.

**Formula:**

****

s represents the number of seasons in a year (for quarterly data, s=4; for monthly data, s=12).

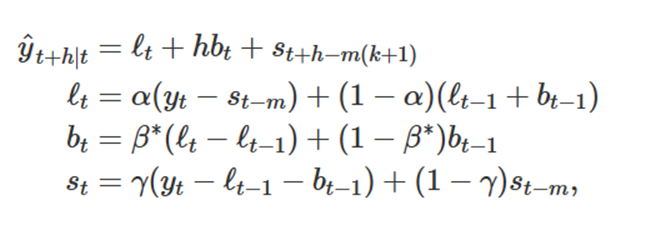
L\_t represents the level component of the time series.

b\_t represents the trend component.

S\_t represents the seasonal component.

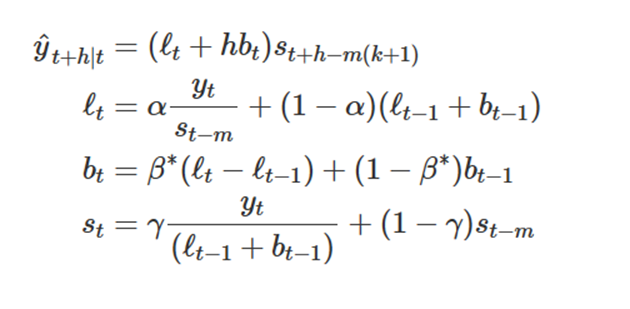
F\_(t+m) represents the forecasted value for m periods ahead.

**Holt-Winters’ additive method**

****

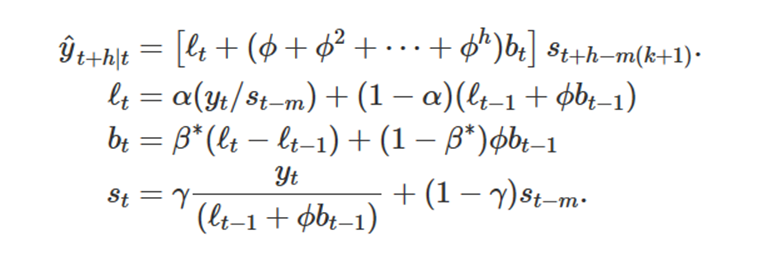
**Holt-Winters’ multiplicative method**

The component form for the multiplicative method is:

****

**Holt-Winters’ damped method**

Damping is possible with both additive and multiplicative Holt-Winters’ methods. A method that often provides accurate and robust forecasts for seasonal data is the Holt-Winters method with a damped trend and multiplicative seasonality:



**How the Holt-Winters model works:**

* Level Component: Represents the average price level of the data in a cycle. This component is updated by assigning a weight (alpha) to the current value and a different weight (1-alpha) to the previous level value. This process estimates the current average price level based on the past data.
* Trend Component: Represents the increasing or decreasing trend of the data over a certain period of time. This component is updated by estimating the current trend value based on the past data and assigning a weight (beta) to the trend value.
* Seasonal Component: Represents the seasonal or cyclical factor in the data. This component helps the model predict the cyclic variations. Different methods can be used to estimate the seasonal component, such as increasing, decreasing, or nonlinear methods.
* Combining the Components: The Holt-Winters model combines the components to predict future values. The level component, trend component, and seasonal component are updated based on past data, and the parameters (alpha, beta, gamma) are adjusted to optimize the model.

**There are some common methods to preprocess input data before applying the Holt-Winters model. Below are some methods for preprocessing input data:**

1. Handling missing values:

• Fill missing values: If the data contains missing values, methods such as mean imputation, median imputation, or nearest value imputation can be used to fill in the missing values.

• Remove missing values: If the proportion of missing values is not too large, the rows or columns containing missing values can be removed from the data set.

2. Handling outliers:

• Detect outliers: Methods such as distribution checking, standard deviation, or model-based methods can be used to detect outliers in the data.

• Handle outliers: Outliers can be removed, replaced with the mean value, or other outlier handling methods such as capping or using a different distribution can be applied.

3. Normalize data:

• Min-max normalization: Transform the variables' values into a standardized value range like [0, 1] by subtracting the minimum value and dividing by the value range.

• Z-score normalization: Transform the variables' values into z-score values by subtracting the mean value and dividing by the standard deviation.

• Logarithmic normalization: Apply logarithmic transformation to variables with uneven distribution or high variability to make the distribution more normal.

4. Handling seasonality:

• Check and remove seasonality: If the data contains seasonality, seasonality can be removed to make the data no longer cyclic.

• Apply seasonality model: If seasonality is important, methods such as time series analysis or the Holt-Winters model can be applied to estimate and handle seasonality.

The data preprocessing process depends on the characteristics of the data and the goals of the model. Choosing the appropriate preprocessing method can improve the accuracy and performance of the Holt-Winters model.

### **2.6 Singular Spectrum Analysis (SSA)**[**[15]**](https://www.zotero.org/google-docs/?d4Rh7T)

Singular Spectrum Analysis (SSA) is a data analysis method used to analyze time series. SSA is often used to solve problems such as time series prediction, time series analysis with noise, or decomposition of time series into components.

SSA involves dividing the initial time series into components, called singular vectors, and corresponding eigenvalues. Then the components are analyzed and combined to form an estimation of the original time series.

SSA is applied in many fields, including statistics, economics, environmental science, and medicine.

The method involves the following steps:

Divide the original time series into parts by constructing Hankel matrices. Each Hankel matrix contains a part of the original data.

Perform eigenvalue decomposition on the Hankel matrices . This eigenvalue decomposition generates the singular vectors and corresponding eigenvalues.

Choose a number of singular vectors and eigenvalues to use. This number can be determined using statistical methods or based on experience.

Calculate the estimates of the elements of the original time series by analyzing the selected singular vectors and eigenvalues from the previous step. These estimates are combined to form an estimation of the original time series.

The formula for this method is:

Divide the original time series into Hankel matrices:



Where n is the length of the original time series and L is the length of the sub-data. The sub-data is constructed by taking consecutive samples of the original time series.

Perform eigenvalue decomposition on the Hankel matrices: 

In there U is the matrix containing the singular vectors of X, Sigma is the matrix containing the corresponding eigenvalues, and V is the transpose matrix of U.

Choose a number of singular vectors and eigenvalues to use.

Calculate the estimates of the elements of the original time series by using the selected singular vectors and eigenvalues from the previous step: 

Where {X^} is the estimation of the original time series, k is the number of selected singular vectors and eigenvalues, sigma\_i is the i-th eigenvalue, u\_i is the i-th singular vector in the matrix U, and v\_i is the i-th singular vector in the matrix V. Then, the estimates of the elements in the original time series can be calculated by taking the corresponding elements from the estimation matrix {X^}.

In summary, the SSA method uses Hankel matrices to decompose a time series into components. The singular vectors and eigenvalues of the Hankel matrix can be used to generate estimates for the original time series. These estimates can be used for time series prediction or time series analysis with noise.

### **2.7 Sequence to sequence** [**[16]**](https://www.zotero.org/google-docs/?tsJaz9)

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.

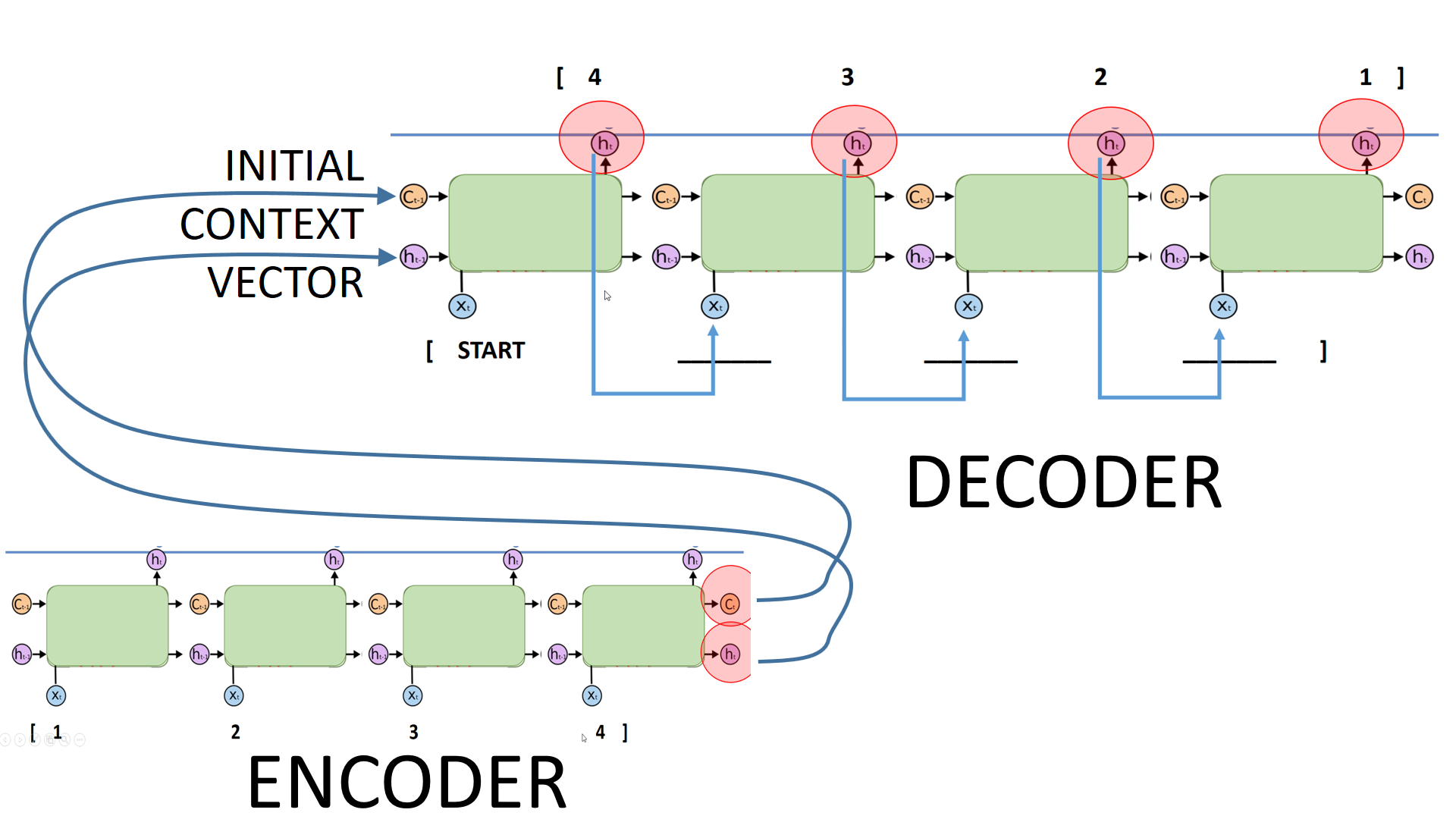


Figure 7: Structure of encoder-decoder seq2seq

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.A black text on a white background

Description automatically generated with medium confidence

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.

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.

### **2.8 Neural Network AutoRegression (NNAR)** [**[17]**](https://www.zotero.org/google-docs/?2dszzu)

The NNAR model is an Artificial Neural Network

A picture containing sketch, diagram, drawing, line

Description automatically generated

*Neural Network Autoregressive structure*

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 chosen based on the information criterion, like AIC.
* and k denotes the number of hidden nodes that are present.

Neural nets have an 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

### **2.9 Recurrent Neural Network with Attention Mechanism (RNN-Attention)**

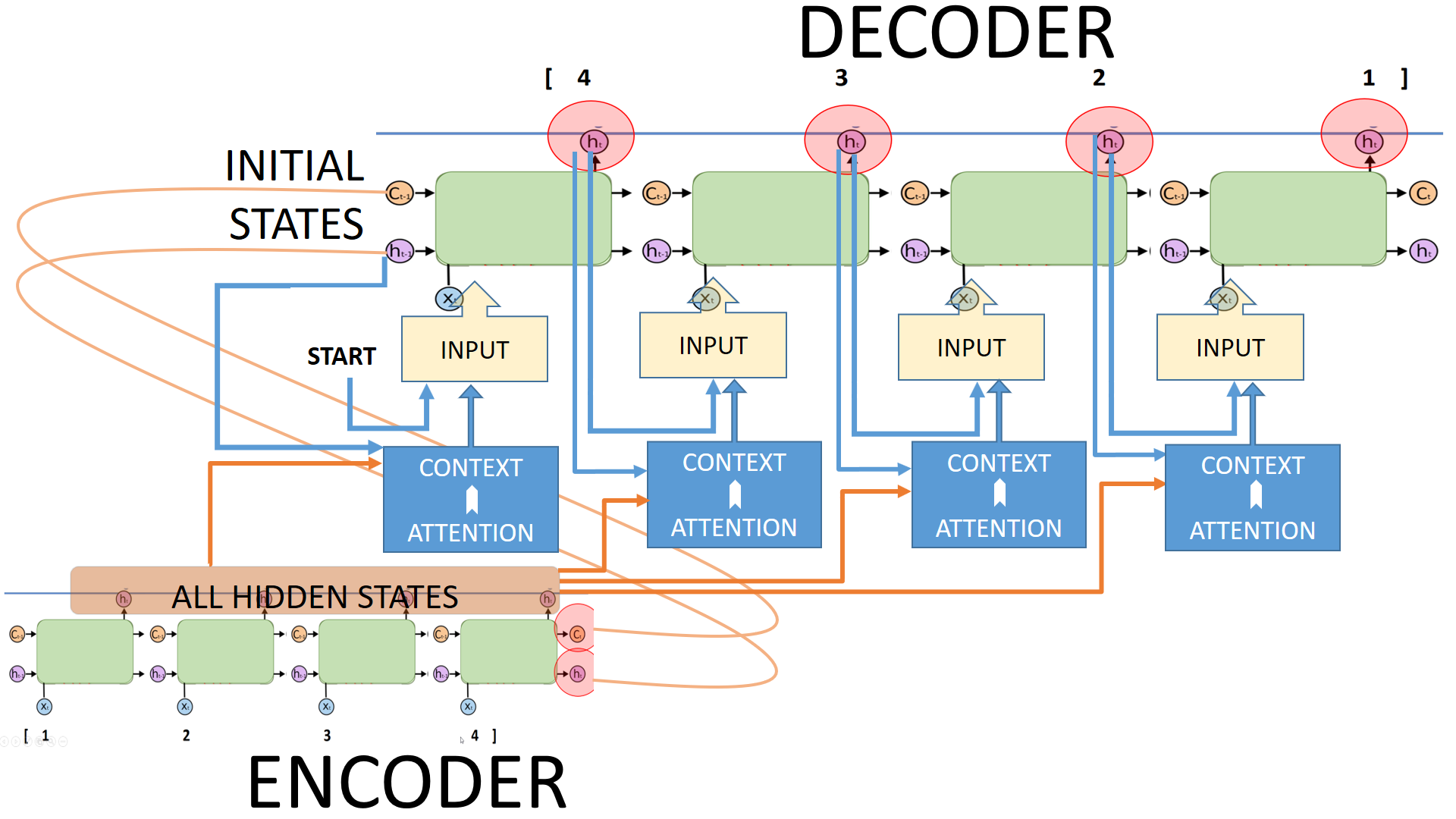
Recurrent Neural Networks (RNNs) have been used successfully for many tasks involving sequential data such as machine translation, sentiment analysis, image captioning, time-series prediction etc. Improved RNN models such as Long Short-Term Memory networks (LSTMs) enable training on long sequences overcoming problems like vanishing gradients.

Attention is a mechanism combined in the RNN allowing it to focus on certain parts of the input sequence when predicting a certain part of the output sequence, enabling easier learning and of higher quality. Combination of attention mechanisms enabled improved performance in many tasks making it an integral part of modern RNN networks. [[18]](https://www.zotero.org/google-docs/?GJNckL)

**How is our RNN-Attention Works** [[19]](https://www.zotero.org/google-docs/?j6v2qf)**:**

We use the Seq2seq base to build RNN-Attention. Let’s call this RNN-Attention is “LSTM Sequence to Sequence Encoder Decoder with Bahdanau Attention Mechanism”.

* First, we will observe that the Basic Encoder Decoder model, developed in Seq2Seq, will fail to handle longer input sequences.
* Then, we will generate each output considering all the inputs using the global attention mechanism.
* We will implement Bahdanau attention mechanism as a Keras custom layer using subclassing.
* And, we will integrate the attention layer to the Encoder-Decoder model.
* After observing the effect of attention layer on performance, we will depict the "created attention" between inputs and outputs.



RNN-Attention Structure

# IV. 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 |
| RNN-ATT | 7:3 | 18590 | 13.12 |
| 8:2 | 983 | 6.11 |
| 7.5:2.5 | 901 | 3.89 |

| **Dataset** | **Model** | **Train:Test** | **RMSE** | **MAPE(%)** |
| --- | --- | --- | --- | --- |
| 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 | 52.005 |
| 8:2 | 6988 | 26.751 |
| 9:1 | 4195 | 5.2231 |
| 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 |
| RNN-ATT | 7:3 | 2096 | 7.45 |
| 8:2 | 21089.98 | 26.12 |
| 7.5:2.5 | 1045.6 | 6.19 |

| **Dataset** | **Model** | **Train:Test** | **RMSE** | **MAPE(%)** |
| --- | --- | --- | --- | --- |
|  | LR | 7:3 | 8157 | 53.21 |
|  | 8:2 | 7035 | 38.18 |
|  | 9:1 | 7237 | 41.16 |
|  | 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 | 135.184 | 131.17 |
| 9:1 | 22671 | 22.42 |
| Seq2seq | 7:3 | 1874.4 | 2.06 |
| 8:2 | 2504.1 | 2.67 |
| 7.5:2.5 | 2036.5 | 2.07 |
| NNAR | 7:3 | 80705 | 8.364 |
| 8:2 | 81629 | 8.737 |
| 9:1 | 78566 | 6.148 |
| RNN-ATT | 7:3 | 2370.7 | 13.1 |
| 8:2 | 32590 | 23.12 |
| 7.5:2.5 | 1901 | 9.89 |

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.

A picture containing text, screenshot, plot, diagram

Description automatically generated

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.

A picture containing text, screenshot, plot, line

Description automatically generated

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

A picture containing text, screenshot, plot, font

Description automatically generated.

# V. 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.

**VI. FUTURE IMPROVEMENT IDEAS**

Holt-Winters is a simple algorithm in statistics and forecasting. However, it is necessary to try all variations of this algorithm on the dataset to find the optimal and most suitable solution.

SSA is an algorithm that helps improve the results of analysis and prediction. Therefore, it is recommended to use SSA in combination with other algorithms to optimize the forecasting results and achieve more accurate predictions. For example, combining SSA with Holt-Winters can improve the forecast results for the next 30 days.

Try on other varieties of attention mechanisms for our RNN-Attention.

# VII. GROUP WORK DISTRIBUTION

| WORK | My Tam | Gia Bao | Quoc Bao |
| --- | --- | --- | --- |
| Problem Statement | X | X | X |
| Data collection | X |  |  |
| Visualizing data |  | X |  |
| Data Analysis | X |  |  |
| Data preprocessing | X |  |  |
| Models research | X | X | X |
| Arima Model |  | X |  |
| RNN Model | X |  |  |
| Holt-Winter Model | X |  | X |
| Linear Regression |  |  | X |
| LSTM Model |  | X |  |
| SSA |  |  | X |
| Seq2Seq | X |  |  |
| NNAR |  | X |  |
| RNN-ATT | X |  |  |
| Compiling codes |  | X |  |
| Visualizing results |  | X |  |
| Applying goodness-of-fit measures | X | X | X |
| Writing report | X | X | X |

# VIII. REFERENCE

[***[1]*** *“Joint Stock Commercial Bank for Investment and Development of Vietnam Stock Price Today” Investing.com, Jun. 16, 2023. https://www.investing.com/equities/commercial-bank-investment-develop (accessed Jun. 19, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[2]*** *“FPT Corp Stock Price Today” Investing.com, Jun. 21, 2023. https://www.investing.com/equities/fpt-corp (accessed Jun. 21, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[3]*** *“Vietnam Construction and Import Export JSC Stock Price Today” Investing.com, Jun. 21, 2023. https://www.investing.com/equities/vinaconex.,jsc (accessed Jun. 21, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[4]*** *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/?Obsz7o)

[***[5]*** *“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/?Obsz7o)

[***[6]*** *“Bitcoin Price Prediction Using Machine Learning Techniques.” https://www.researchgate.net/publication/363018597\_Bitcoin\_Price\_Prediction\_Using\_Machine\_Learning\_Techniques (accessed Jun. 17, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[7]*** *H. Andi, “An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model,” J. Soft Comput. Paradigm, vol. 3, pp. 205–217, Sep. 2021, doi: 10.36548/jscp.2021.3.006.*](https://www.zotero.org/google-docs/?Obsz7o)

[***[8]*** *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.*](https://www.zotero.org/google-docs/?Obsz7o)

[***[9]*** *L.-C. Cheng, Y.-H. Huang, and M.-E. Wu, “Applied attention-based LSTM neural networks in stock prediction,” in 2018 IEEE International Conference on Big Data (Big Data), Dec. 2018, pp. 4716–4718. doi: 10.1109/BigData.2018.8622541.*](https://www.zotero.org/google-docs/?Obsz7o)

[***[10]*** *Z. Munim, M. Shakil, and I. Alon, “Next-Day Bitcoin Price Forecast,” J. Risk Financ. Manag., vol. 12, Jun. 2019, doi: 10.3390/jrfm12020103.*](https://www.zotero.org/google-docs/?Obsz7o)

[***[11]*** *K. MALI, “Everything you need to Know about Linear Regression!,” Analytics Vidhya, Oct. 04, 2021. https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/ (accessed Jun. 19, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[12]*** *“Autoregressive Integrated Moving Average (ARIMA) - Applications.” https://corporatefinanceinstitute.com/resources/data-science/autoregressive-integrated-moving-average-arima/ (accessed May 14, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[13]*** *“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/?Obsz7o)

[***[14]*** *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/?Obsz7o)

[***[15]*** *“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/?Obsz7o)

[***[16]*** *“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/?Obsz7o)

[***[17]*** *T. P. Singh, Chapter 8 Neural Networks in Time Series Analysis. Accessed: Jun. 19, 2023. https://bookdown.org/singh\_pratap\_tejendra/intro\_time\_series\_r/neural-networks-in-time-series-analysis.html*](https://www.zotero.org/google-docs/?Obsz7o)

[***[18]*** *N. Arbel, “Attention in RNNs,” Medium, Mar. 28, 2019. https://medium.datadriveninvestor.com/attention-in-rnns-321fbcd64f05 (accessed Jun. 22, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)

[***[19]*** *“Seq2Seq Learning PART F: Encoder-Decoder with Bahdanau & Luong Attention Mechanism - Murat Karakaya Akademi.” https://www.muratkarakaya.net/2022/11/seq2seq-learning-part-f-encoder-decoder.html (accessed Jun. 22, 2023).*](https://www.zotero.org/google-docs/?Obsz7o)