

# **PREDICTING PRICES ON ETHEREUM**

**AMIR IZZAT AZIEM BIN MOHD AZIZAN**

**FACULTY OF COMPUTING AND  
INFORMATICS**

**UNIVERSITI MALAYSIA SABAH**

**2022**



**UMS**  
UNIVERSITI MALAYSIA SABAH

# **PREDICTING PRICES ON ETHEREUM**

**AMIR IZZAT AZIEM BIN MOHD AZIZAN**

**THESIS SUBMITTED IN PARTIAL FULFILLMENT  
FOR DEGREE OF COMPUTER SCIENCE WITH  
HONOURS (NETWORK ENGINEERING)**

**FACULTY OF COMPUTING AND  
INFORMATICS**

**UNIVERSITI MALAYSIA SABAH**

**2022**



**UMS**  
UNIVERSITI MALAYSIA SABAH

**NAME** : AMIR IZZAT AZIEM BIN MOHD AZIZAN  
**MATRIC NUMBER** : BI18110268  
**TITLE** : PREDICTION PRICE ON ETHEREUM  
**BACHELOR** : BACHELOR OF COMPUTER SCIENCE (NETWORK ENGINEERING)  
**DATE** : 24 FEBRUARY 2022

**VERIFIED BY:**

**1. SUPERVISOR**

MADAM SALMAH BINTI FATTAH

**SIGNATURE**

  
SALMA BINTI FATTAH  
Lecturer  
Faculty of Computing and Informatics  
Universiti Malaysia Sabah

---



## DECLARATION

I declare that this research project has indeed been submitted to Universiti Malaysia Sabah as a partial fulfilment of the requirements of the Degree of Computer Science with Honours (Network Engineering) and it has not been presented to any university for any degree. I also confirm that this work presented here is exclusively my own work. Information derived from literature has been duly recognized in the reference section provided.

24 February 2022



---

AMIR IZZAT AZIEM BIN MOHD AZIZAN  
BI18110268

## ACKNOWLEDGEMENT

Above all, I am grateful to Allah for his favours. I finished the report for my research project effectively. I'd want to convey my heartfelt gratitude to my supervisor, Madam Salmah Fattah, for her invaluable support during the whole study process. It's a privilege to be able to learn and work under her guidance. I am eternally thankful to my family for their love, prayers, care, and commitment to educating and preparing me for the future. My heartfelt gratitude also goes to my friends and classmates, who have cheerfully assisted me and provided encouraging support in completing the thesis. With the caring and compassionate support of many people, this theory becomes a reality. I'd want to offer my heartfelt gratitude to each and every one of them.



---

AMIR IZZAT AZIEM BIN MOHD AZIZAN

BI18110268



## ABSTRACT

Cryptocurrencies are quickly becoming one of the world's most important investment focus areas. People could become wealthy than ever before by investing in cryptocurrencies. Individuals who lack basic financial knowledge, on the other hand, risk losing all of their money in the matter of seconds. It is the outcome of dramatic price graph changes in the bitcoin market. Technical innovation, internal competition, market pressure to deliver, economic obstacles, security concerns, and political issues all have an impact on cryptocurrency pricing. These are the present limitations of the cryptocurrency market. Prophet forecast and Long-Short Term Memory Neural Network (LSTM) are the project's candidates. The datasets that are utilised are those derived from any cryptocurrency historical data that are used to test the prediction of the model that are developed for this project. The model that are implemented into the web application for user testing are the best fit. All the testing and result are conducted by functionality test case and integration test case.

## **ABSTRAK**

### **RAMALAN HARGA ETHEREUM**

*Mata wang kripto dengan cepat menjadi salah satu bidang tumpuan pelaburan yang paling penting di dunia. Orang ramai boleh menjadi kaya berbanding sebelum ini dengan melabur dalam mata wang kripto. Individu yang kurang pengetahuan asas kewangan, sebaliknya, berisiko kehilangan semua wang mereka dalam masa beberapa saat. Ia adalah hasil daripada perubahan graf harga yang dramatik dalam pasaran bitcoin. Inovasi teknikal, persaingan dalaman, tekanan pasaran untuk menyampaikan, halangan ekonomi, kebimbangan keselamatan dan isu politik semuanya mempunyai kesan ke atas harga mata wang kripto. Ini adalah had semasa pasaran mata wang kripto. Prophet Forecast dan Rangkaian Neural Memori Jangka Pendek (LSTM) adalah calon projek. Set data yang digunakan adalah yang diperoleh daripada mana-mana data sejarah mata wang kripto yang digunakan untuk menguji ramalan model yang dibangunkan untuk projek ini. Model yang dilaksanakan ke dalam aplikasi web untuk ujian pengguna adalah yang paling sesuai. Semua ujian dan keputusan dijalankan oleh kes ujian kefungsian dan kes ujian integrasi.*

# TABLE OF CONTENT

	Page
<b>TITLE</b>	i
<b>DECLARATION</b>	iii
<b>ACKNOWLEDGEMENT</b>	iv
<b>ABSTRACT</b>	v
<i>ABSTRAK</i>	vi
<b>TABLE OF CONTENTS</b>	vii
<b>LIST OF FIGURES</b>	xi
<b>LIST OF TABLES</b>	xiv
<b>CHAPTER 1: INTRODUCTION</b>	1
1.1 PROBLEM BACKGROUND	1
1.3 PROBLEM STATEMENTS	2
1.4 PROJECT OBJECTIVES	2
1.5 PROJECT SCOPE	3
<b>CHAPTER 2: LITERATURE REVIEW</b>	4
2.1 INTRODUCTION	4
2.2 CRYPTOCURRENCY	4
2.3 CONCEPT OF CRYPTOCURRENCIES	5
2.4 BLOCKCHAINS	6
2.5 TIME SERIES FORECASTING	7
2.6 RELATED WORKS IN FORECASTING CRYPTOCURRENCIES PRICES.	7
2.6.1 ARTIFICIAL NEURAL NETWORK	7





2.6.2 RANDOM SAMPLING METHOD	9
2.6.3 REALISE VOTATILITY	9
2.6.4 LONG SHOT TERM MEMORY NEURAL NETWORK	10
2.7 EXISTING RESEARCH ON CRYPTOCUURRENCY PREDICTION	12
2.8 EXISTING CRYPTOCURRENCY FORECASTING	15
2.8.1 CYRPTOPREDICTIONS.COM	15
2.8.2 TRADINGBEASTS.COM	17
2.9 CONCLUSION	21
<b>CHAPTER 3: METHODOLOGY</b>	22
3.1 PROTOTYPE MODEL APPROACH	22
3.1.1 PHASE 1: REQUIREMENT	22
3.1.2 PHASE 2: DESIGN	23
3.1.3 PHASE 3: DEVELOPMENT	23
3.1.4 PHASE 4: TESTING	23
3.1.5 PHASE 5: DEVELOPMENT	23
3.1.6 PHASE 6: REVIEW	24
3.2 PROJECT FRAMEWORK	24
3.3 SOFTWARE AND HARDWARE REQUIREMENT	26
3.4 DATASETS	26
3.5 PROPHET FORECAST MODEL	27
3.6 LONG SHORT-TERM MEMORY NEURAL NETWORK	28
3.7 CONCLUSION	29
<b>CHAPTER 4: SYSTEM ANALYSIS AND DESIGN</b>	30
4.1 INTRODUCTION	30



4.2 SYSTEM ANALYSIS	30
4.3 SYSTEM DESIGN	31
4.4 ENTITY-RELATIONSHIP DIAGRAM (ERD)	32
4.4.1 DATA DICTIONARY	33
4.5 DATA FLOW DIAGRAM	34
4.6 CONTEXT DIAGRAM	35
4.7 DFD LEVEL-1	36
4.8 DFD LEVEL-2	36
4.9 INTERFACE DESIGN	37
4.10 CONCLUSION	39
<b>CHAPTER 5: IMPLEMENTATION</b>	40
5.1 INTRODUCTION	40
5.2 DEVELOPMENT TOOLS	40
5.2.1 GOOGLE COLABORATORY	41
5.2.2 JUPYTER NOTEBOOK	42
5.2.3 ANACONDA	42
5.3 PROJECT IMPLEMENTATION	43
5.4 PHASE 1 COMPARISON BETWEEN THE TWO SELECTED MODELS	43
5.4.1 DATASET GATHERING	43
5.4.2 BUILDING MODEL	46
5.5 PHASE 2: SYSTEM IMPLEMENTATION	53
5.6 CONCLUSION	60
<b>CHAPTER 6: TESTING AND RESULT</b>	61
6.1 INTRODUCTION	61

6.2 USABILITY TESTING	61
6.3 FUNCTIONALITY TEST	68
6.3.1 REGISTER	68
6.3.2 LOGIN	69
6.3.3 UPLOAD DATA INPUT	70
6.3.4 LOGOUT	70
6.4 INTEGRATION TEST	71
6.5 CONCLUSION	72
<b>CHAPTER 7: CONCLUSION</b>	73
7.1 INTRODUCTION	73
7.2 SUMMARY OF PROJECT	73
7.3 FUTURE WORKS	74
7.4 PROJECT CONSTRAINT	75
7.5 CONCLUSION	75
<b>REFERENCES</b>	76

## LIST OF FIGURES

	Page
Figure 2.1: Example of a blockchain	6
Figure 2.2: Main page of the website	15
Figure 2.3: The price prediction interface	16
Figure 2.4: The interface of the prediction table by year	17
Figure 2.5: The main page of the system	18
Figure 2.6: Example of a blockchain	19
Figure 2.7: Example of a blockchain	20
Figure 3.1: Agile SDLC Methodology.	24
Figure 3.2: Project Framework.	25
Figure 3.3: RNN of LSTM	28
Figure 4.1: Entity Relationship Diagram.	32
Figure 4.2: Context Diagram.	35
Figure 4.3: Data Flow Diagram (DFD) level 1.	36
Figure 4.4: Data Flow Diagram (DFD) level 2.	37
Figure 4.5: Web App Main Interface Design.	38
Figure 5.1: Google Colaboratory	41
Figure 5.2: Jupyter Notebook	42
Figure 5.3: Raw Dataset	44
Figure 5.4: Cleaning Dataset	44
Figure 5.5: Bitcoin Graph Over the Years	45
Figure 5.6: Plotting Data to Past 200 Days	45



Figure 5.7: Resample Dataset and Storing Resample Dataset	46
Figure 5.8: Calling Libraries, Data Fitting and Model Fitting	47
Figure 5.9: Adjusting Data, Storing Prediction and Returning the RSME	47
Figure 5.10: Plotting the Forecast Data	48
Figure 5.11: Plotting Forecast Data and Test Data.	49
Figure 5.12: Building LSTM Model	50
Figure 5.13: Plotting Forecast Data and Real Price	51
Figure 5.14: RMSE Output	51
Figure 5.15: Side by Side Graph Comparison Between Prophet and LSTM	52
Figure 5.16: Predict Model and Forecast Graph Code of the Web App	54
Figure 5.17: Code snippet, Cryptocurrency Dataset and Time of Prediction	54
Figure 5.18: Result of Title, Cryptocurrency Dataset and Time of Prediction	55
Figure 5.19: Code snippet of Raw Data Table and Graph	55
Figure 5.20: Result Raw Data	56
Figure 5.21: My sidebar for insert value of water quality data	57
Figure 5.22: Forecast Data Table, Graph and Additional Component	57
Figure 5.23: Login page of the LSTM web application.	58
Figure 5.24: Login successfully using registered account.	59
Figure 5.25: Prediction page of the LSTM web application	60
Figure 6.1: Question 1 of questionnaire.	62
Figure 6.2: Question 2 of questionnaire.	62
Figure 6.3: Question 3 of questionnaire.	63
Figure 6.4: Question 4 of questionnaire.	63
Figure 6.5: Question 5 of questionnaire.	64



Figure 6.6: Question 6 of questionnaire.	64
Figure 6.7: Question 7 of questionnaire.	65
Figure 6.8: Question 8 of questionnaire.	65
Figure 6.9: Question 9 of questionnaire.	66
Figure 6.10: Question 10 of questionnaire.	66
Figure 6.11: The SUS Scoring Instrument	67

## LIST OF TABLES

	Page
Table 2.1: Existing research on cryptocurrency price prediction	12
Table 2.2: The comparison between the existing web app and purposed system	20
Table 4.1: Data Dictionary	23
Table 5.1: Preliminary Experiment Result using RMSE	51
Table 6.1: Register Module Positive Case	68
Table 6.2: Register Module Negative Case	68
Table 6.3: Login Module Positive Case	69
Table 6.4: Login Module Negative Case	69
Table 6.5: Upload Data Input Module Positive Case	70
Table 6.6: Logout Module Positive Case	71
Table 6.7: Integration Test Case	



# CHAPTER 1

## INTRODUCTION

### 1.1 PROBLEM BACKGROUND/MOTIVATION

Predicting the cryptocurrency market is a specific challenge since reading the graph of the crypto market requires a decent deal of financial expertise. It also takes a long time to analyse the graph since the market constantly moves up and down (Bonneau et al., 2015). As a result, failing to pay attention to the pricing market might result in a person's largest financial loss. As a result, having a price forecast on Ethereum might stimulate the market and make Ethereum a viable alternative for ordinary people with necessary financial expertise and assist small business owners in building their empires. There has been some research done to show that utilising deep learning to anticipate pricing is capable of doing so. According to the study conducted for this project, most projects are based on paper. There is no proof that this research has been translated into development projects such as a website utilising the prophet forecasting model and a long short-term memory neural network. As a result, this research aims to compare the most often used deep learning model, the long short-term memory neural network, with the contemporary deep learning model, the prophet forecasting model.



## 1.2 PROBLEM STATEMENT

The primary motivation for this research endeavour is that individuals continue to rely on manual calculations and expensive machine processing. Investing in the bitcoin market is a dangerous choice since a single blunder in decision-making may lead to massive losses. Furthermore, with the announcement that Ethereum will be updated to Ethereum 2.0, it will become safer and more advanced than it was before. As a result, it is an ideal moment to convince people that Ethereum is the finest long-term investment since its blockchain can be converted into real-world assets.

More individuals will be able to invest more cautiously in bitcoin with artificial intelligence and machine learning since they will have a safe guideline to show them when to participate in the cryptocurrency market. Many studies have been conducted about this topic, as is well known. Even though there have been many studies, most of them are based on paper theory and laboratory experiments. The primary goal of this project is to compare the most commonly used deep learning model, the long short term memory neural network, with a more recent deep learning model, the prophet forecasting model, and develop a consumer function web base.

## 1.3 PROJECT OBJECTIVES

- i. To compare the Prophet model and Long Short-Term Memory Neural Network using data from reputable sources.
- ii. To design and develop best fit model into a working web-based that can produce more accessible data to be read by a normal user.
- iii. To implement and test the usability on the web base model by providing real time user testing.

#### **1.4 PROJECT SCOPE**

To carry out the project, certain processes will be preceded, such as collecting real-time cryptocurrency data from the cryptocurrency market website, preparing the data collected for training, testing and comparing, predicting cryptocurrency prices using the preferred method related to artificial intelligence and machine learning, and displaying data the prediction result to ensure the project runs smoothly. To have an artificial intelligence-based prediction of the price of the cryptocurrency market, real-time datasets are required in this project to ensure that the project runs smoothly and without errors. The dataset for this project will be provided by Yahoo!Finance because the majority of their data on cryptocurrency is accurate and reputable.

# CHAPTER 2

## LITERATURE REVIEW

### 2.1 INTRODUCTION

This chapter summarizes the extensive research conducted and a review of previous related research. This chapter explains the study's relationship to previous research and the impact of selected research on the proposed system. An examination of the literature is a thorough examination of the sources relevant to the pattern and includes both review and creation. This literature review provides a solid foundation for the project of a researcher paper. Finding relevant publications, such as books and journal articles, critically analysing them, and explaining the results are all part of writing a literature review. Most research papers benefit from a large amount of knowledge in the literature review.

### 2.2 CRYPTOCURRENCY

World Wide Web is introduced in 30 April 1993. It made it possible for us to send copies of the "originals" of information so that everyone could get the same information. As the time goes by, the most popular technology that is available on the World Wide Web is blockchain technology. The internet before only



provides us with information and entertainment but now the internet has become the internet with information to the internet with value. Before going deeper into cryptocurrency, one must know the basics of cryptocurrency, which is blockchain. Based on Dylan Yaga, blockchains are tamper-evident and tamper-resistant digital ledgers that are distributed and typically decentralized (Yaga et al., 2019). In the earlier stage of blockchain, blockchains allow a community of users to record the transaction in a shared ledger within that community so that no transaction can be altered after publishing during the operation of the blockchain network is carried out (Yaga et al., 2019). During 2008, the concept of blockchain is integrated with a number of other technologies and computer principles to produce what we know now as cryptocurrencies (Yaga et al., 2019).

### **2.3 CONCEPT OF CRYPTOCURRENCIES**

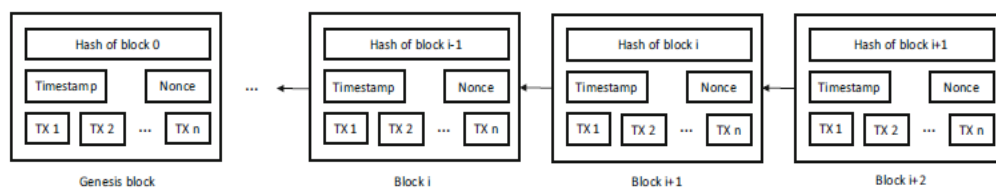
In the research paper entitled "A Short Introduction to the World of Cryptocurrencies" published by Aleksander Berentsen and Fabian Schär, the first cryptocurrency is created in 2008 by an alias named "Satoshi Nakamoto" which is Bitcoin (Berentsen & Schär, 2018). Based on Aleksander Berentsen and Fabian Schär, Bitcoin is developed to create a payment in the form of a cash system that allows electronic transactions with a bonus of advantages similar to physical cash (Berentsen & Schär, 2018).

To gain more understanding of Bitcoin, first must need to step back one step to fully understand Bitcoin. Bitcoin is only one of many types of cryptocurrencies that are available nowadays. Cryptocurrencies are actually similar to the physical type of payment which is cash or paper money that we use nowadays, but cryptocurrencies are entirely digital types of payment. Cryptocurrencies are basically just a collection of numbers and letters. Although cryptocurrencies are a collection of numbers and letters, it is very similar to physical money bill that is currently used up to date. It is because every physical bill has a unique serial number that corresponds to when the money is printed and where the money is printed. All the unique serial numbers of these physical bills are stored in a record

that is kept by a central bank and shared throughout smaller banks and governments (Giudici et al., 2020).

## 2.4 BLOCKCHAINS

Blockchain is a chain of blocks that contains information. The technique was invented back in 1991 by a group of researchers and the purpose is to timestamp documents to prevent any individual to backdate or tamper them. This technique seldom being implement until the creator of digital cryptocurrencies, Bitcoin which is known widely as "Satoshi Nakamoto". The reason of the usage of blockchain is when any data is stored in a blockchain, it is very difficult to make changes to it. By using Bitcoin block as an example, the data stored in the blockchain consists of the sender, receiver and number of coins. The block in blockchains also has a hash that recognizes the block and all of its data inside which is always unique of its own. Hash is calculated when a block is created. Therefore, changing any of the data inside the blockchain will change the calculated hash and it will not be the same block anymore. Hence, this will improve the security of the blockchain but it is not fully secured. To make sure that blockchain is fully secured, proof-of-work are added in the blockchain to avoid anyone from tampering the data.



**Figure 2.1 Example of a blockchain** (Nofer et al., 2017; Xie et al., 2018)

## **2.5 TIME SERIES FORECASTING**

Time series is a sample of data collected on the values that a changing at different point in time. Data are gathered on a recurring basis or yearly. Time series can be utilised in many sections in the real-world fields such as statistics, weather forecasting, finance, stock market and fields that have data and recorded time interval. Time series forecasting is used in this project because the datasets that will be used in the project will require this type of information and understanding. There are two type of time series forecasting that will be used frequently in this project. First, univariate time series is one of the important aspects to understand as the future project will require some understanding on these properties to handle the datasets that will be used in future project. Univariate time series is time series that include a single observation recorded over regular time intervals. Cross sectional data also is an important aspect that need to be understand as it is type of data that is collected by observing many subjects at the same point of time. Therefore, time series forecasting is needed to run the project.

## **2.6 RELATED WORKS IN FORECASTING CRYPTOCURRENCIES PRICES**

### **2.6.1 ARTIFICIAL NEURAL NETWORK**

Cryptocurrency has established itself as the decentralisation leader. Many individuals keep them as a kind of speculation because to the enormous unpredictability in their value. As a result, it is crucial to comprehend the underlying characteristics and predictability of those cryptocurrencies. On the early research of the price dynamics of Bitcoin, Ethereum, and Ripple using sophisticated artificial intelligence frameworks of fully linked Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) Recurrent Neural Network (Yiyang & Yeze, 2019). Yinyang and Yeze state that despite their differences in underlying architectures, the ANN and LSTM models are similar and perform relatively well in price prediction(Yiyang & Yeze, 2019). After the research have been carried out, Wang Yiyang and Zang Yeze conclude that ANN relies more on long-term history, while LSTM relies more on short-term dynamics, both of them

stated that LSTM is more efficient than ANN in bringing up relevant information concealed in historical memory(Yiying & Yeze, 2019).ANN could achieve an accuracy which is on par to LSTM if the model was given enough data. However, Yinying and Yeze stated that the rationale for predictability may differ clean depends on the nature of the underlying machine-learning model(Yiying & Yeze, 2019). This project have a major weakness where without the utilization of multiple technical indicators, the non-stationary data will overfit the prediction model data.

Another research is about Bitcoin intraday technical trading using artificial neural networks to anticipate returns. Masafumi Nakano develop an artificial neural network (ANN) model for classification in order to extract relevant trading signals from input technical indicators generated from time-series return data at 15-minute intervals. (Nakano et al., 2018). Based to the research, Nakano's strategy outperformed the buy-and-hold and unsophisticated technical trading techniques (Nakano et al., 2018). Nakano discovered significant implications on the relationship of ANN-based model performance with execution costs, the number of layers, the determination of activation functions, the number of classes in the output layer, and the input data by implementing a wide range of sensitivity analysis(Nakano et al., 2018). Nakano's findings suggest that the inclusion of numerous technical indicators may reduce over-fitting in the categorization of non-stationary financial time-series data, hence improving trading success. Although Nakano have concentrated on Bitcoin as a single asset, it is proven advantageous to build a multi-asset portfolio from multiple cryptocurrencies, which is one of their future study subjects (Nakano et al., 2018). The model performs particularly effectively during a difficult time from December 2017 to January 2018, it is when the Bitcoin suffers from significant negative returns. Furthermore, several sensitivity analyses for changes in the number of layers, activation functions, input data, and output classification are done to prove the robustness of the technique (Nakano et al., 2018). From these findings, the data utilize from Bitcoin are ranged from open, high, low and close of Bitcoin prices and the block difficulty and hash rate will be the weakness of the project.

## 2.6.2 RANDOM SAMPLING METHOD

Another related research is about random sampling method (RSM) for cryptocurrency times series. Shintate, says that it is hard to predict unstable Bitcoin values in the OkCoin market, hence, he compared the performance of their approach to two standard baseline techniques and demonstrate that the baseline approaches are readily skewed by class imbalance, but the model mitigates this issue (Shintate & Pichl, 2019a). Shintate state that the classification performance of their method significantly outperforms the odds of a uniform random process with three outcomes, demonstrating that deterministic pattern extraction for trend classification, and market prediction, is possible to some extent(Shintate & Pichl, 2019b). The profit rates based on RSM exceeded those based on LSTM, but the profit rates of random sampling methods were lower than those obtain compare to the classical buy-and-hold strategy.

In the report, Shintate create a new trend prediction classification learning method and proof that his research performs well in domains where the non-stationarity assumption which are acceptable.(Shintate & Pichl, 2019a). Shintate also perform trials with extremely tiny scaled models to differentiate the impact of his technique and validate its superiority over the MLP and LSTM baselines (Shintate & Pichl, 2019a).

## 2.6.3 REALISE VOLATILITY

Realised volatility (RV) is the sum of the squares of logarithmic returns on a high-frequency sampling grid aggregated over a certain time period, which is a trading day (Miura et al., 2019).Miura says the technique used is to aggregate RV values across 3-hour periods using minute-sampled Bitcoin returns(Miura et al., 2019). By using samples from variety of machine learning methods, the RV time series is to calculate future values and compare it to the Heterogeneous Auto-Regressive Realized Volatility (HARRV) model with optimum lag settings. The best supporting auto regressive dynamics predicted by the HARRV model is the

