Predictive Analysis of price of Ripple

Project Synopsis

of Major Project

Bachelor of Technology

(Information Technology)

Submitted By

DAMANPREET SINGH (1411247)

GURNOOR SINGH (1411254)

Guru Nanak Dev Engineering College

Ludhiana 141006

Abstract

This research is concerned with predicting the price of Ripple using machine learning. The goal is to ascertain with what accuracy can the direction of Ripple price in USD can be predicted.

The price data is sourced from the Ripple Price Index. The task is achieved with varying degrees of success through the implementation of a Bayesian optimised recurrent neural network (RNN) and Long Short Term Memory (LSTM) network. The LSTM achieves the highest classification accuracy of 52% and a RMSE of 8%.

The popular ARIMA model for time series forecasting is implemented as a comparison to the deep learning models. As expected, the non-linear deep learning methods outperform the ARIMA forecast which performs poorly. Wavelets are explored as part of the time series narrative but not implemented for prediction purposes. Finally, both deep learning models are benchmarked on both a GPU and a CPU with the training time on the GPU outperforming the CPU implementation by 67.7%.

Acknowledgement

We, students of Guru Nanak Dev Engineering College, Ludhiana, have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

The authors are highly grateful to Dr. Sehijpal Singh Director, Guru Nanak Dev Engineering College, Ludhiana for providing them with the opportunity to carry out their Major project.

The authors would like to wholeheartedly thank Sachin Bagga, Associate Professor, GNDEC who is a vast sea of knowledge and without whose constant and never ending support and motivation, it would never have been possible to complete the project and other assignments so efficiently and effectively.

Gurnoor Singh
Damanpreet Singh

LIST OF FIGURES

CONTENTS

CHAPTER 1 INTRODUCTION

1.1 Overview

Figure 1.1: Ripple logo

Ripple is real-time gross settlement system (RTGS), currency exchange by Ripple. It is also called the Ripple Transaction Protocol (RTXP) or Ripple protocol. It is built upon a distributed open source Internet protocol, ledger and cryptocurrency called XRP (ripples). Ripple enables, Secure, fast, instant and free global financial transactions of any size with no chargebacks. The digital currency XRP acts as a bridge currency to other currencies. It supports tokens that represents fiat currency, cryptocurrency. It is based around a shared, public database which uses a process that allows for payments, exchanges and remittance in a distributed process. It is fastest and most scalable Digital Asset in the world. Ripple is a very popular network as many banks across the world use it as the basis for their own settlement.

Ripple is a technology that acts as both a cryptocurrency and a digital payment network for financial transactions. It was first developed in 2004 by Ryan Fugger, a web developer in Vancouver, British Columbia. Fugger conceived of the idea after working on a local exchange trading system in Vancouver with his intent to create a decentralized monetary system that effectively allow individuals and communities to create their own money. This led to the conception of a new system by Jed McCaleb of eDonkey network, which was designed and built by Arthur Britto and David Schwartz In May 2011, the development of a digital currency system begins in which transactions were verified by consensus among members of the network, rather than by the mining process used by bitcoin, which relies on blockchain ledgers This new version of the Ripple system was designed to eliminate bitcoin's reliance on centralized exchanges, It use less electricity than bitcoin and perform transactions much more quickly than bitcoin. Chris Larsen founded the lending services companies E-Loan and prosper, joined the team in August 2012 and together McCaleb and Larsen approached Ryan Fugger with their digital currency idea. In September 2012 the team co-founded the corporation OpenCoin, or OpenCoin Inc. launched Ripple.

1.2 Project Category

This project is a research work which is concerned with predicting the price of Ripple using machine learning.

1.3 Problem analysis

There is currently no best way to predict real-valued quantity. LSTM is commonly used to predict values of time series data. This prediction can be used to gain financial benefits. This research can further be extended to predict country's growth rate or gross GDP.

1.4 Objectives

- Predicting price of Ripple cryptocurrency using Ripple's previous price.
- The Goal is to find out with what accuracy we can predict the prices.
- Implementing Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) Model.

CHAPTER 2

REQUIREMENT ANALYSIS

2.1 Feasibility Study

Feasibility study aims to uncover the strengths and weaknesses of a project. These are some feasibility factors by which we can used to determine that the project is feasible or not:

2.1.0.1 Technical Feasibility

Technical feasibility is one of the first studies that must be conducted after the project has been identified. In large engineering projects consulting agencies that have large staffs of engineers and technicians conduct technical studies dealing with the projects. In individual agricultural projects financed by local agricultural credit corporations, the technical staff composed of specialized agricultural engineers, irrigation and construction engineers, and other technicians are responsible for conducting such feasibility studies.
The Technical feasibility assessment is focused on gaining an understanding of the present technical resources of the organization and their applicability to the expected needs of the proposed system. It is an evaluation of the hardware and software and how it meets the need of the proposed system. This assessment is based on an outline design of system requirements, to determine whether the company has the technical expertise to handle completion of the project. When writing a feasibility report, the following should be taken to consideration:

- A brief description of the business to assess more possible factors which could affect the study
- The part of the business being examined
- The human and economic factor
- The possible solutions to the problem

2.1.0.2 Economic Feasibility

The purpose of the economic feasibility assessment is to determine the positive economic benefits to the organization that the proposed system will provide. It includes quantification and identification of all the benefits expected. This assessment typically involves a cost/ benefits analysis.

Economic feasibility is the cost and logistical outlook for a business project or endeavor. Prior to embarking on a new venture, most businesses conduct an economic feasibility study, which is a study that analyzes data to determine whether the cost of the prospective new venture will ultimately be profitable to the company. Economic feasibility is sometimes determined within an organization, while other times companies hire an external company that specializes in conducting economic feasibility studies for them.

The purpose of business in a capitalist society is to turn a profit, or to earn positive income. While some ideas seem excellent when they are first presented, they are not always economically feasible. That is, that they are not always profitable or even possible within a company's budget. Since companies often determine their budget's several months in advance, it is necessary to know how much of the budget needs to be set aside for future projects. Economic feasibility helps companies determine what that dollar amount is before a project is ultimately approved. This allows companies to carefully manage their money to insure the most profitable projects are undertaken. Economic feasibility also helps companies determine whether or not revisions to a project that at first seems unfeasible will make it feasible.

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require. Economic feasibility determines whether the required software is capable of generating financial gains for an organization. It involves the cost incurred on the software development team, estimated cost of hardware and software, cost of performing feasibility study, and so on. For this, it is essential to consider expenses made on purchases (such as hardware purchase) and activities required to carry out software development. In addition, it is necessary to consider the benefits that can be achieved by developing the software. Software is said to be economically feasible if it focuses on the issues listed below.

- Cost incurred on software development to produce long-term gains for an organization.
- Cost required to conduct full software investigation (such as requirements elicitation and requirements analysis).
- Cost of hardware, software, development team, and training.

The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

2.2 Software Requirements

- **Python 2.7** - Python is a programming language. Its used for many different applications. Its used in some high schools and colleges as an introductory programming language because Python is easy to learn, but its also used by professional software developers at places such as Google, NASA, and Lucasfilm Ltd. It is used for Research Work in the field of Machine Learning.
- **Python pip** - pip is a package management system used to install and manage software packages written in Python. Many packages can be found in the default source for packages and their dependencies Python Package Index. We will use it to install required python packages.
- **Python packages** -
  - **Pandas** - The Pandas module is a high performance, highly efficient, and high level data analysis library. We will use it to analyse our Dataset.
  - **Sklearn** - Open Sourced simple and efficient tool for data mining and data analysis.
  - **Keras** - Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano.
  - **Lxml** - The most feature-rich and easy-to-use library for processing XML and HTML in the Python language.

2.3 Predictive Analytic Requirements

Building a decision requirements model to specify business understanding at the very beginning of a predictive analytic project allows the creation of predictive analytic requirements that:

1. Describe a clear target for the project. The decisions that the predictive analytic will influence are specified. The decision requirements diagram links these decisions to the ultimate business metrics or objectives that will be impacted by the analytic.
2. Identifythe analytics to be developed. Each piece of analytic knowledge can be described along with the information to be analyzed to produce it.
3. Is specific about which decisions are being influenced. The decision requirements diagram shows w hich part of the decision making is influenced, exactly, by each predictive analytic model being developed and what other factors influence that decision - making.
4. Is specific about deployment. The links for the decisions involved show w hich organizations will be involved , w hich business processes will be impacted and w hich systems will have to be altered.

A complete set of requirements for a predictive analytic project should include other project details such as executive sponsor, timeline, resources, planned analytic approach , etc. The d ecision r equirements model allows the business problem being addressed by the project to be described more precisely, but it does not replace these other elements.

2.4 Software Requirements:

1. Python 2.7
2. Python pip
3. Anaconda

2.5 Python Packages Dependencies

1. Pandas
2. Sklearn

CHAPTER 3 SYSTEM DESIGN

3.1 Design Approach

Function oriented design approach is comprised of many smaller sub-systems known as functions. These functions are capable of performing significant task in the system. The system is considered as top view of all functions. Function oriented design inherits some properties of structured design where divide and conquer methodology is used.
This design mechanism divides the whole system into smaller functions, which provides means of abstraction by concealing the information and their operation. These functional modules can share information among themselves by means of information passing and using information available globally.
For this project, we are following a Function Oriented Design Approach. We have defined various functions depending upon our need. This way, we have complete control over different modules of our project. The main focus is on data.

3.2 System Design

- **Acquiring Data**:- This part is concerned with acquiring data required for research work. The sources of this data can be data-set from Kaggle or from API.
- **Training Model**:- The model used in this research is trained using the data acquired in previous step.
- **Long Short Team Memory Model**:- A deep learning model which uses historical data to train data set for predictive analysis i.e. prediction of time series data.

3.3 Workflow of Project

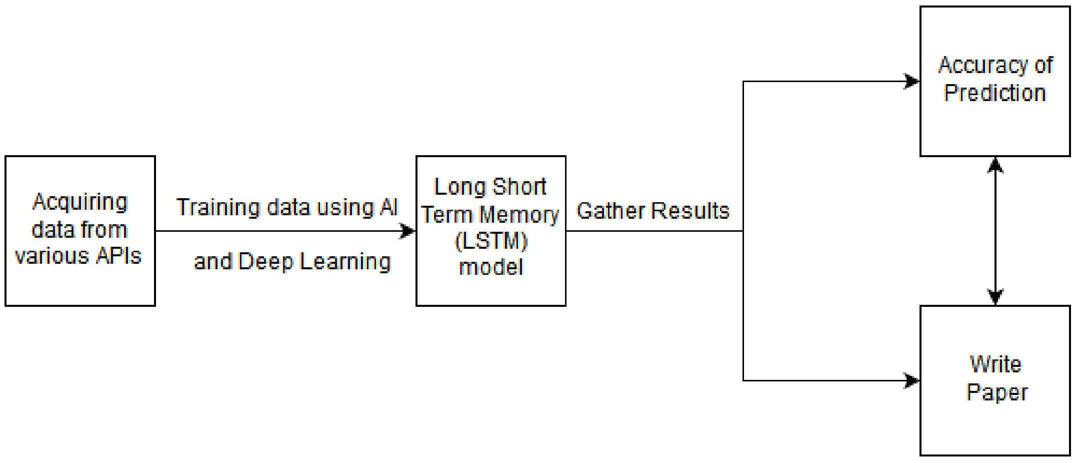

Acquiring data from various APIs
Training data using AI and Deep Learning
Long Short Term Memory (LSTM) model
Gather Results
Accuracy of Prediction
Write Paper

Figure 3.1: Workflow of Project

3.4 Methodology

1. **Acquiring Data**:- This part is concerned with acquiring data required for research work from various APIs. This data will be used for training model.
2. **Long Short Team Memory Model**:- A deep learning model which will use acquired data and will generate the predicted values.
3. **Accuracy of Prediction**:- Once the results are gathered, we need to check the accuracy of the gatherd results.
4. **Writing Paper**:- Write and publish a Research Paper using the work done during the course of this project and the results obtained.

CHAPTER 4 IMPLEMENTATION

4.1 Python

Figure 4.1: Python logo

Python is a dynamic language, as in python coding is very easy and also it require less coding and about its interpreted nature it is just exellent. Python is a high level programming language and Django which is a web development framework is written in python language.

Python is an easy to learn, powerful programming language.Python runs on Windows, Linux/Unix, Mac OS X. Python is free to use, even for commercial products. Python can also be used as an extension language for existing modules and applications that need a programmable interface. Python is free to use, even for commercial products, because of its OSI-approved open source license.

4.1.1 Features of Python

- Very clear, readable syntax.
- Strong introspection capabilities.
- Intuitive object orientation.
- Natural expression of procedural code.
- Full modularity, supporting hierarchical packages.
- Exception-based error handling.
- Very high level dynamic data types.
- Extensive standard libraries and third party modules for virtually every task.
- Extensions and modules easily written in C, C++ (or Java for Jython, or .NET languages for IronPython).
- Embeddable within applications as a scripting interface.

4.1.2 Installation of Python

Installation of python is a very easy proccess. The current python versions are: Python 2.7.1 and Python 3.2. Type the commands in the terminal:

```
$ wget http://www.python.org/ftp/python/2.7/Python-2.7.tgz

$ tar xzf Python-2.7.tgz
```

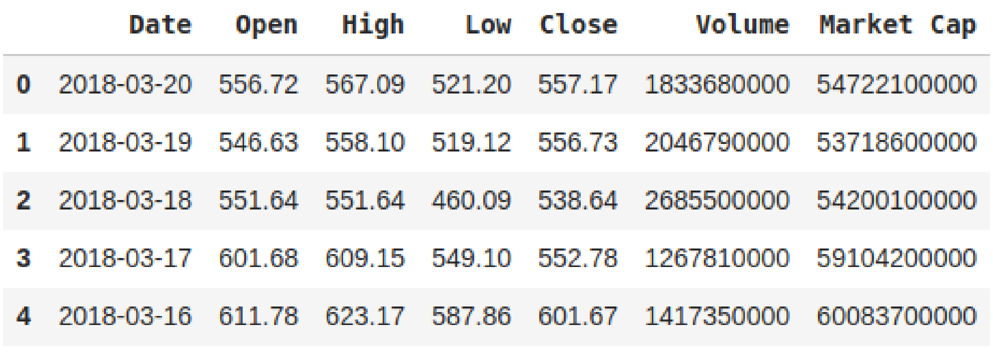
This will install the python on your pc/laptop.

4.2 Steps of Implementation

Were going to employ a Long Short Term Memory (LSTM) model; its a particular type of deep learning model that is well suited to time series data (or any data with temporal/spatial/structural order e.g. movies, sentences, etc.).

4.2.1 Data

Before we build the model, we need to obtain some data for it. Theres a dataset on Kaggle that details minute by minute Bitcoin prices (plus some other factors) for the last few years (featured on that other blog post). Over this timescale, noise could overwhelm the signal, so well opt for daily prices. The issue here is that we may have not sufficient data (well have hundreds of rows rather than thousands or millions). In deep learning, no model can overcome a severe lack of data. I also dont want to rely on static files, as thatll complicate the process of updating the model in the future with new data. Instead, well aim to pull data from websites and APIs.



	Date	Open	High	Low	Close	Volume	Market Cap
0	2018-03-20	556.72	567.09	521.20	557.17	1833680000	54722100000
1	2018-03-19	546.63	558.10	519.12	556.73	2046790000	53718600000
2	2018-03-18	551.64	551.64	460.09	538.64	2685500000	54200100000
3	2018-03-17	601.68	609.15	549.10	552.78	1267810000	59104200000
4	2018-03-16	611.78	623.17	587.86	601.67	1417350000	60083700000

Figure 4.2: Data used

As well be combining multiple cryptos in one model, its probably a good idea to pull the data from one source. Well use coinmarketcap.com. For now, well only consider Bitcoin and

Ether, but it wouldnt be hard to add the latest overhyped altcoin using this approach. Before we import the data, we must load some python packages that will make our lives so much easier.

4.2.2 Training

We have some data, so now we need to build a model. In deep learning, the data is typically split into training and test sets. The model is built on the training set and subsequently evaluated on the unseen test set. In time series models, we generally train on one period of time and then test on another separate period. As such, the training data may not be representative of the test data, undermining the models ability to generalise to unseen data.The most basic model is to set tomorrows price equal to todays price (which well crudely call a lag model).

4.3 Long Short Term Memory (LSTM)

Long short-term memory (LSTM) units (or blocks) are a building unit for layers of a recurrent neural network (RNN). A RNN composed of LSTM units is often called an LSTM network. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell is responsible for "remembering" values over arbitrary time intervals; hence the word "memory" in LSTM. Each of the three gates can be thought of as a "conventional" artificial neuron, as in a multi-layer (or feedforward) neural network: that is, they compute an activation (using an activation function) of a weighted sum. Intuitively, they can be thought as regulators of the flow of values that goes through the connections of the LSTM; hence the denotation "gate". There are connections between these gates and the cell.

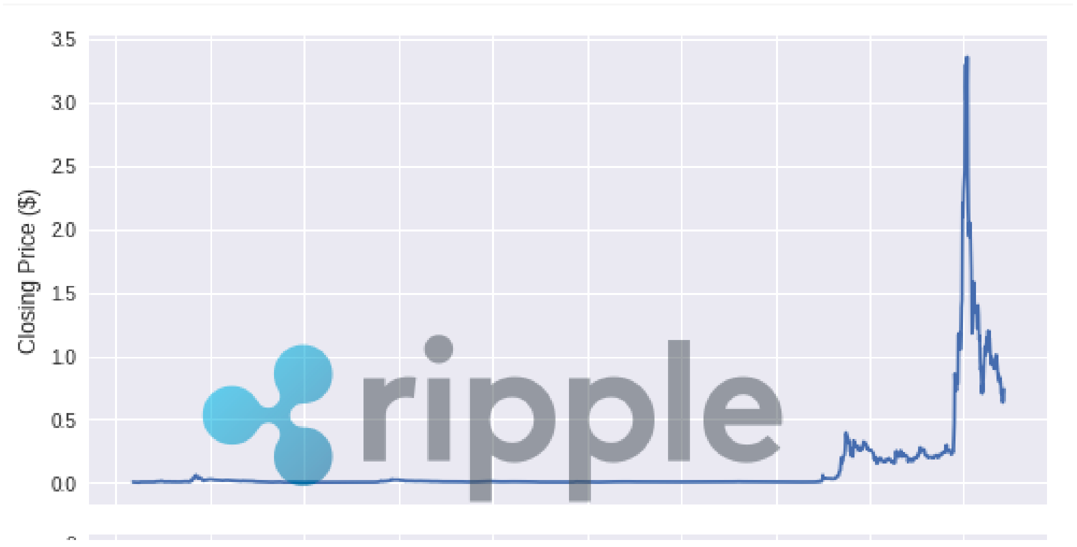


Figure 4.3: Expected Results

4.4 Training and Random Walk

We have some data, so now we need to build a model. In deep learning, the data is typically split into training and test sets. The model is built on the training set and subsequently evaluated on the unseen test set. In time series models, we generally train on one period of time and then test on another separate period. Rather arbitrarily, we'll set the cut-off date to June 1st 2017 (i.e. model will be trained on data before that date and assessed on data after it).

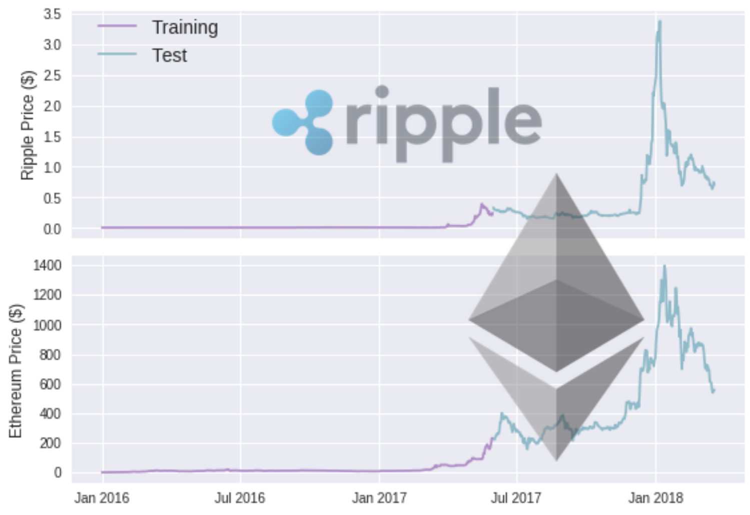


Figure 4.4: Splitting data into Test and Train

The training period mostly consists of periods when cryptos were relatively cheaper. As such, the training data may not be representative of the test data, undermining the model's ability to generalise to unseen data. The most basic model is to set tomorrow's price equal to today's price (which we'll crudely call a lag model). This is how we'd define such a model in graphical terms:

First, we may want to make sure the daily change in price follows a normal distribution. We'll plot the histogram of values. This will give us a graphical tool to check for daily changes in values over time.

4.5 Long Short Term Memory (LSTM)

We don't need to build the network from scratch. There exists packages that include standard implementations of various deep learning algorithme - TensorFlow, Keras, PyTorch etc. We are using Keras here because it is the most intuitive for non-experts.

Our LSTM model will use previous data to predict the next day's closing price of a specific coin. We must decide how many previous days it will have access to. We build little data frames consisting of 10 consecutive days of data (called windows), so the first window will consist of the 0-9th rows of the training set, the second will be the rows 1-10, etc. Picking a small window

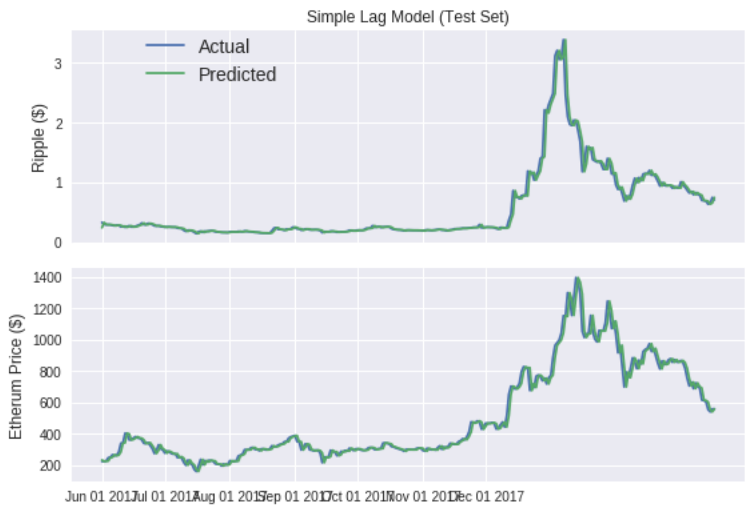


Figure 4.5: Simple Lag Model

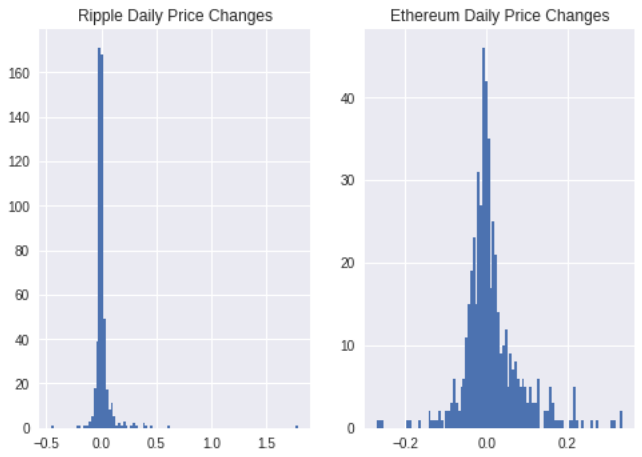


Figure 4.6: Daily Price Change Histogram

size means we can feed more windows into our model; the downside is that the model may not have sufficient information to detect complex long term behaviours.

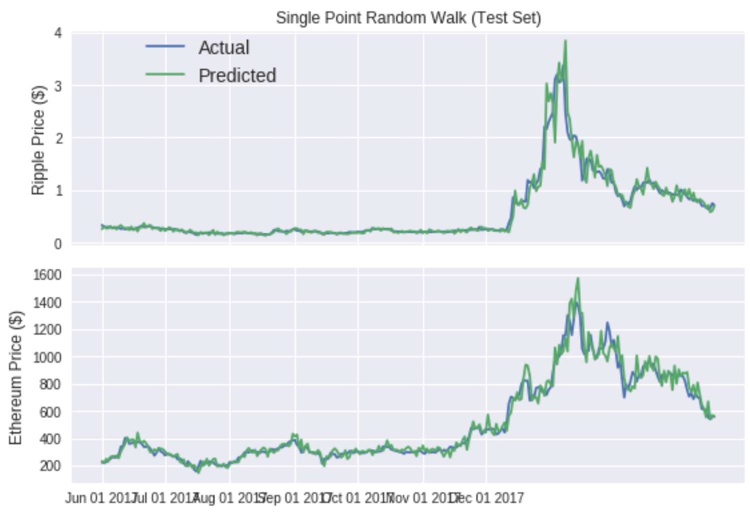
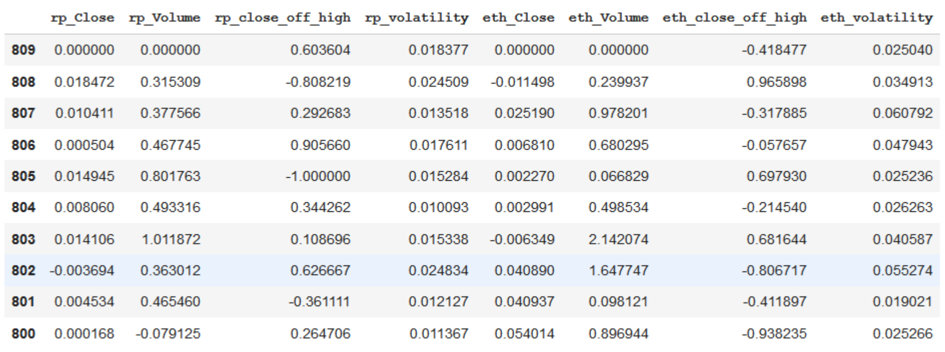


Figure 4.7: Single Point Random Walk



	rp_Close	rp_Volume	rp_close_off_high	rp_volatility	eth_Close	eth_Volume	eth_close_off_high	eth_volatility
809	0.000000	0.000000	0.603604	0.018377	0.000000	0.000000	-0.418477	0.025040
808	0.018472	0.315309	-0.808219	0.024509	-0.011498	0.239937	0.965898	0.034913
807	0.010411	0.377566	0.292683	0.013518	0.025190	0.978201	-0.317885	0.060792
806	0.000504	0.467745	0.905660	0.017611	0.006810	0.680295	-0.057657	0.047943
805	0.014945	0.801763	-1.000000	0.015284	0.002270	0.066829	0.697930	0.025236
804	0.008060	0.493316	0.344262	0.010093	0.002991	0.498534	-0.214540	0.026263
803	0.014106	1.011872	0.108696	0.015338	-0.006349	2.142074	0.681644	0.040587
802	-0.003694	0.363012	0.626667	0.024834	0.040890	1.647747	-0.806717	0.055274
801	0.004534	0.465460	-0.361111	0.012127	0.040937	0.098121	-0.411897	0.019021
800	0.000168	-0.079125	0.264706	0.011367	0.054014	0.896944	-0.938235	0.025266

Figure 4.8: LSTM Training table

CHAPTER 5 RESULTS

5.1 Mean Absolute Error (MAE)

We are using Mean Absolute Error to check the absolute losses in Epouchs. We are plotting a graph between Mean Absolute Error and number of Epouch. This graph will give us more in-sight into the standard error which occures during Training LSTM model.

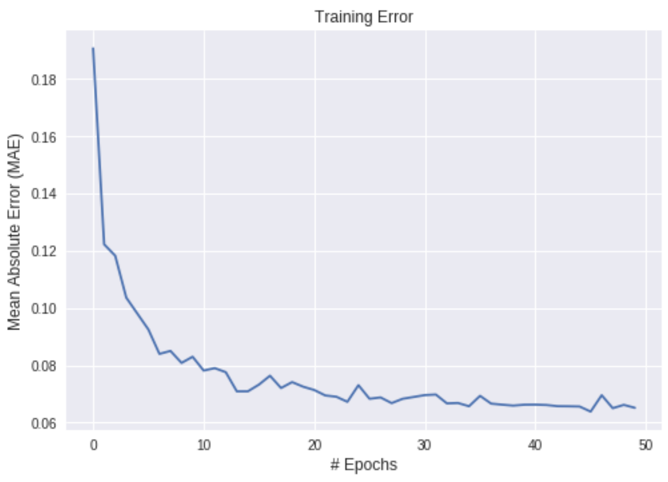


Figure 5.1: Mean Absolute Error

If everything went to plan, then we'd expect the training error to have gradually decreased over time. In other words, the accuracy of model increase as the model trains.

5.2 Single Point Prediction

Once we have trained our model, it is time to test our model. Initially, we'll test for one particular point. This point will be from our Testing data-set. A good strategy is to keep 80% of data for training and 20% for testing.

5.3 Calculating Error

Moving back to the single point predictions, our deep machine artificial neural model looks okay, but so did that boring random walk model. Like the random walk model, LSTM models can be sensitive to the choice of random seed (the model weights are initially randomly assigned). So, if we want to compare the two models, we'll run each one multiple (say, 25) times to get an estimate for the model error. The error will be calculated as the absolute difference between

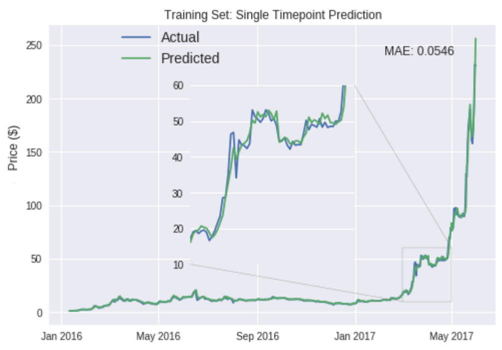


Figure 5.2: Single TimePoint Prediction

the actual and predicted closing prices changes in the test set. Those graphs show the error on

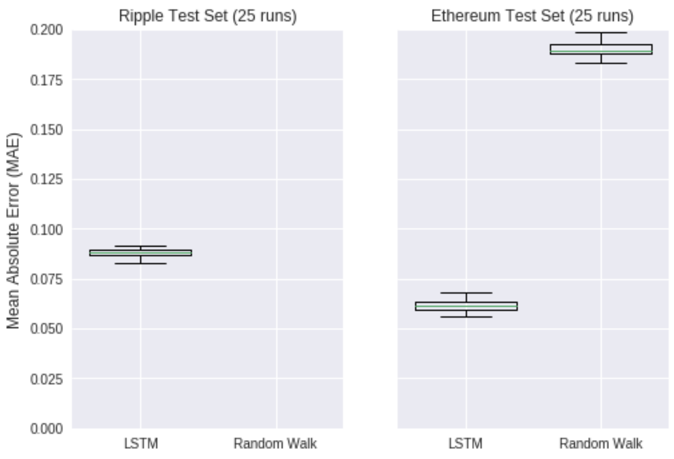


Figure 5.3: Calculating Error

the test set after 25 different initialisations of each model. The LSTM model returns an average error of about 0.0125 on ripple prices, crushing the corresponding random walk models.

CHAPTER 6 CONCLUSION

6.1 Summary

We've collected some crypto data and fed it into a deeply intelligent machine learning LSTM model. Unfortunately, its predictions were not that different from just spitting out the previous value. We can make the model learn more sophisticated behaviours by -

- **Change Loss Function:** MAE doesn't really encourage risk taking. For example, under mean squared error (MSE), the LSTM model would be forced to place more importance on detecting spikes/troughs. More bespoke trading focused loss functions could also move the model towards less conservative behaviours.
- **Penalise conservative AR-type models:** This would incentivise the deep learning algorithm to explore more risky/interesting models. Easier said than done!
- **Get more and/or better data:** If past prices alone are sufficient to decently forecast future prices, we need to include other features that provide comparable predictive power. That way, the LSTM model wouldn't be so reliant on past prices, potentially unlocking more complex behaviours. This is probably the best and hardest solution.

It's entirely possible that there is no detectable pattern to changes in crypto prices; that no model (however deep) can separate the signal from the noise (similar to the merits of using deep learning to predict earthquakes). And any pattern that does appear can disappear as quickly. All of this suggests you might as well save yourself some time and stick to autoregression.

The LSTM model returns an average error of about 0.04 on the ripple prices, crushing the corresponding random walk models.

6.2 Future Scope

- Write and publish a Research Paper using the work-done and finding of this project.
- Further train model to get more accurate results.
- Obtain and refine Dataset from services like - coinbase.
- Improving the performance of deep learning model by using GPUs.

BIBLIOGRAPHY

[1] G. H. Chen, S. Nikolov, and D. Shah, A latent source model for nonparametric time series classification, in Advances in Neural Information Processing Systems , pp. 10881096, 2013.

[2] K. Fukunaga, Introduction to statistical pattern recogni- tion . Academic press, 1990.

[3] A. W. Lo and A. C. MacKinlay, Stock market prices do not fol low random walks: Evidence from a simple specification test . Princeton, NJ: Princeton University Press, 1999.

[4] G. Caginalp and H. Laurent, The predictive power of price patterns, Applied Mathematical Finance , vol. 5, pp. 181206, 1988..

[5] C.-H. Park and S. Irwin, The profitability of technical analysis: A review, AgMAS Project Research Report No. 2004-04 , 2004.